Statistical Modelling in Non-clinical Pharmaceutical Research

Leacky Muchene

Promotor: Prof. dr. Ziv Shkedy
Co-Promotor: Dr. Tom Jacobs

Acknowledgements

My sincere gratitude to my promoter Ziv Shkedy for sporting my potential at an early stage and nurturing me throughout my master thesis and the PhD. This dissertation is a culmination of meetings, formal and informal discussions we had. The journey has been eventful considering the many train rides we ran to catch, the interesting meetings with industry partners, the discussions over conferences/workshops and many more opportunities that we interacted. You made each day worth looking forward to, and it was always a pleasure working with you. My co-promoter Tom Jacobs, I appreciate the effort you made to not only explain the scientific basis for the datasets used in this dissertation, but also the challenge you always posed to me of tailoring the statistical solutions to the needs of the pharmaceutical industry. The Thursday meetings at Janssen pharmaceutica were always thought provoking and fulfilling.

This dissertation would not have been a success without the input of various partners. Marleen Verhoye, your gentle introduction to the concept of magnetic resonance imaging with regards to this project formed a good foundation for my research. Moreover, the contributions from a biological perspective from your team in Antwerp university-Bioimaging lab completed the picture or even challenged the statistics we presented, which made the work wholesome. The partnership with Janssen pharmaceutica cannot be overemphasized. In particular, the inputs of all the partners I collaborated with for the MRI project is appreciated. My gratitude to Luc Bijnens for extending an opportunity for me to collaborate with his team beyond the MRI project.

As they say, home is where the heart is. The biggest sacrifice in making this dissertation a success has been being over 6000km away from home. I acknowledge the support, patience and emotional sacrifices mum (Joyce) and dad (George) have had to bear with to make us all happy. To Elvis, thank you for running all the errands in my absence. Sisters, Banice, Eunice and Charity, *asanteni kwa kunichangamsha* and keeping me updated. It has been emotionally draining to miss all the happy moments and sad times but the journey has been worth the sacrifice. To my dear wife Nancy, you have ran the race with me and the end is finally here. My extended family and family friends (especially Karanja's, family and Mugi's family), thank you always for the encouragement. Gideon, Richu, Agnes, Judy and Lawrence, thank you for being great

and supportive friends.

Finally, my gratitude to the Uhasselt student community I have interacted with since 2010, notably, Edmund (*baba yao*), Elasma, Sammy, Agogo, Rosette, Nolen, Theophile, Leandro and the African student community at large. You all made life in Belgium fun. To the larger Uhasselt fraternity, thank you for coming to our aid whenever we desperately needed it. You gave us a shoulder to lean on whenever we needed one. May blessings be upon you.

Leacky Muchene,
Diepenbeek, 22 September, 2016

Publications

Part of the materials presented in this thesis, are based on the following publications:

Muchene, L., Shkedy, Z., Fantahun Aregay, M., Molenberghs, G., Jacobs, T. and Otava, M. (2016). Hierarchical Binomial-Poisson Overdispersed Models for Correlated Dose-Response Data: A Joint Modelling Approach. *Journal of Biopharmaceutical Statistics. (Submitted)*.

Muchene, L., Shkedy, Z., Bijnens, L., Manyakov, N., Van De Casteele, T., Jacobs, T., Verhoye, M., Praet, J., Van der Linden, A., Bottelbergs, A., Mark Schmidt, M., and Pemberton, D. (2016). Hierarchical Bayesian Model for the Evaluation of Magnetic Resonance Imaging as a Biomarker of Histology in Alzheimers' Disease. *Communications in statistics. Simulation and computation. (Submitted)*.

Muchene, L., Shkedy, Z., Bijnens, L., Manyakov, N., Van De Casteele, T., Jacobs, T., Verhoye, M., Praet, J., Van der Linden, A., Bottelbergs, A., Schmidt, M. and Pemberton, Darrel (2016). Evaluation of Magnetic Resonance Imaging as a Biomarker in Alzheimer's Disease. In *Applied Surrogate Endpoint Evaluation Methods with SAS and R*. Chapman & Hall/CRC Biostatistics Series.

Muchene, L., Shkedy, Z., Jacobs, T. and Otava, M. (2016). Joint Modelling of Correlated Dose-Response Data: Estimation and Model Selection Under Model Uncertainty Using Bayesian Variable Selection (BVS) Models. *Statistics in Biopharmaceutical Research. (To be submitted)*.

Praet, J. , Manyakov, N., **Muchene, L.**, Mai, Z., Terzopoulos, V., De Backer, S., Torremans, A., Guns, P., Van De Casteele, T., Bottelbergs, A., Van Broeck, B., Sijbers, J., Smeets, D., Shkedy, Z., Bijnens, L., Pemberton, D., Schmidt, M., Verhoye, M., van der Linden, A. (2016). *NEuroImage. (To be submitted).*

Long Tuan Ho, Duy Tan Pham, Wout Van Echelpoel, **Leacky Muchene**, Ziv Shkedy, Andres Alvarado, Olivier Thas, Peter L.M. Goethals (2016). Mixed Model Reveals Spatiotemporal Effects on Oxygen Level in Waste Stabilisation Pond. (*To be submitted*)

Contents

Chapter 1

General Introduction

The research presented in this dissertation focuses on the analysis of non-clinical studies in drug development. Two case studies will be discussed. The first is a non-clinical experiment that was conducted to detect early stage biomarkers for Alzheimer's disease. The second case study is a dose-response screening behavioural study (the Differential Reinforcement of Low-rate, DRL-72 study) that was conducted to evaluate the efficacy of a new antidepressant compound.

Alzheimer's Disease (AD) research has gained much interest in recent past (Winblad *et al.*, 2016, O'Bryant *et al.*, 2015). Being an irreversible condition, current therapeutic treatments aim at mainly managing the condition in order to improve the quality of life for the patients. The challenge related with AD is that early diagnosis is hampered by lack of appropriate diagnostic tools, which would pick out the disease in its earliest stages and distinguish it from other forms of dementia. Currently, pathological examination of AD can only be performed post-mortem, which from a patient management point of view, is too late. Several attempts at identifying and validating potential biomarkers for AD, for instance Cerebral Spinal Fluid (CSF) and blood-based biomarkers, which would not only allow for early diagnosis, but also ease the follow-up during treatment, have been made. Sensitivity of blood-based biomarkers to early AD has not been proven, while CSF acquisition is too intrusive to be viable for disease follow-up. Non-invasive imaging-based biomarkers hold the most premise, if validated, for clinical follow-up of AD (Fletcher *et al.*, 2013, López-de-Ipiña *et al.*, 2013). In the first part of this thesis, we examine Magnetic Resonance Imaging (MRI) markers for AD pathology. In particular, while Diffusion Tensor Imaging (DTI) has been shown to reasonably quantify disease pathology in the brain and is currently mostly included in AD imaging, Diffusion Kurtosis Imaging (DKI) may be superior (Cheung *et al.*, 2009, y Palacios *et al.*, 2011).

The first part of this thesis focuses on the MRI DTI and DKI parameters and their application in AD research. Chapter 2 presents a description of an MRI experiment for AD in mice. Moreover, the acquisition of diffusion metrics and the pathological histology is discussed. The resulting longitudinal MRI dataset allows for the evaluation of disease

pathology development and therefore, can be used to discriminate between diseased and healthy mice. In Chapter 3, a longitudinal analysis of the dataset is performed for several brain regions using both DTI and DKI parameters, while Chapter 4 presents the linear discriminant analysis methodology used to discriminate between the healthy and diseased mice.

The last three chapters in the first part of the thesis are focused on the establishment of a linkage between the MRI data and the true pathological histology, in order to validate several MRI parameters as biomarkers for histology. Chapter 5 presents a short overview of the methodology for surrogate endpoints based on the meta-analytic approach of Buyse *et al.* (2000), in the context of AD studies, from a graphical point of view. An application of the methodology to several regions in the brain is discussed in Chapter 6 and a "surrogacy map of the brain" is presented. Finally, Bayesian approach to the validation of MRI markers as a surrogate for pathological histology is presented in Chapter 7.

We shift gears in the second part of the thesis whereby, an application of Bayesian methods to address diverse challenges in dose-response modelling is discussed. In Chapter 8, an introduction to model selection and uncertainty resulting from model selection in dose-response modelling is provided and the datasets, both Gaussian and non-Gaussian, that are used for illustration are presented. Count and data often exhibits extra variability than that which would be expected for their respective distribution. This overdispersion, if not accounted for, may results in invalid inference. Chapter 9 is devoted to the methodology for accounting for overdispersion in correlated dose-response joint models (with Poisson and binomial outcomes). In order to account for uncertainty resulting from model selection, Bayesian Variable Selection (BVS) is proposed, which results in model-averaged estimates of the dose effect. The BVS methodology and an application in Bayesian hierarchical dose-response models with and without overdispersion is presented in Chapter 10.

In modelling dose-response data, monotonicity constraints may be of interest especially in testing directional hypotheses such as an increasing or decreasing 0 trend. Chapter 11 discusses the methodology for the analysis of dose-response data where the hypothesis of ordered dose effects is of interest. These models may suffer from uncertainty in determining the correct shape of the monotone dose-response profile and therefore, in Chapter 11, we extend the order-restricted model and incorporate Bayesian variable selection procedure. In Chapter 12, we propose a finite mixture model for analysing the DRL-72 dose-response data. The resulting analysis not only allows inference for the dose-effect, but also enables the quantification of potential side effects of the drug. Finally, in Chapter 13, we provide a brief discussion of the main results, limitations and proposal for future analyses.

Part I

Diffusion Kurtosis Magnetic Resonance Imaging in Neurodevelopment and Neurodegeneration

Chapter 2

Alzheimer's Disease: An Introduction

2.1 Alzheimer's Disease

Alzheimer's disease, the most common cause of dementia, is an irreversible age-related condition resulting in an increase in dependency on care providers for basic functioning. Clinical symptoms of sporadic AD manifest mostly in the elderly population (at least 65 years) and include progressive deterioration of specific cognitive functions such as memory, speech, motor skills and perception (McKhann *et al.*, 1984). A proper diagnosis of AD suffers from the lack of diagnostic tools that can accurately distinguish AD from other causes of cognitive impairment especially at an early stage of the disease (Blennow, 2004, Chetelat and Baron, 2003, Galvin and Sadowsky, 2012). Moreover, AD results in multiple pathological changes in the brain, which do not manifest the same way in all patients. The most common AD related pathological changes in the brain include: amyloid-beta protein plaque deposition (Masters *et al.*, 1985, Hardy and Selkoe, 2002), neurofibrillary tangle (hyperphosphorylated tau) formation, and neuro-degeneration (Hol *et al.*, 2003, Serrano-Pozo *et al.*, 2011). How these changes influence the progression of AD is unfortunately not clearly understood since the onset of clinical symptoms of AD occurs much later than the onset of the pathological changes associated with the disease (Agronin, 2007). Considering the fact that there is no known cure for AD, an early diagnosis of the disease would therefore be preferable in order to allow for the introduction of treatments that may delay the progression of the disease such as a lifestyle intervention or novel therapeutic management of the patients.

From a practical point of view, although the pathological markers of AD are indicative of the disease progression, they can only be measured cross-sectionally (once per patient). This is due to the fact that pathological histology staining involves post-mortem examination whose acquisition comes too late from a diagnostic point of view (Perl, 2010). Thus, potential biomakers which can be easily acquired in clinical

follow-up of patients are of interest in early diagnosis of the disease (Hampel *et al.*, 2009).

One of the challenges in current AD research is the availability of a suitable animal model, fully representative for the disease pathology. The advantage of using an animal model exhibiting only one pathological indication of AD is that, it enables us to study the influence of one aspect of the disease, without the interaction of other pathological indications (Duff and Suleman, 2004). An animal model with co-expressed Amyloid Precursor Protein (APP) gene and Presenilin (PS) results in variants of mouse models such as the APP/PS1 mouse model that was used in this experiment (Götz and Götz, 2009, Radde *et al.*, 2006).

2.2 Magnetic Resonance Imaging and Histology Parameters

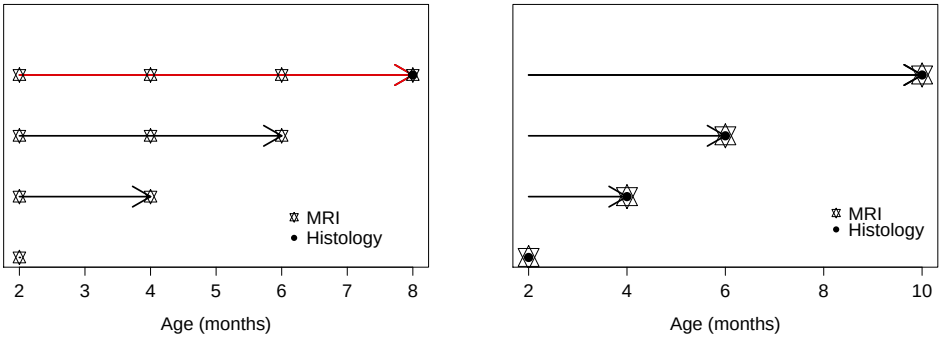
Non-invasive neuroimaging based technologies such as Positron Emission Tomography (PET) scan and Magnetic Resonance Imaging (MRI), if adequately validated, hold the most premise for adoption in both diagnosis and clinical follow-up of disease progression in AD (Dickerson and Sperling, 2005). Using neuroimaging, differences in brain anatomy, chemistry and physiology can be detected via the measured MRI parameters. Additionally, longitudinal MRI studies enable the assessment of neuro-anatomical changes as the animal ages. MRI technology is highly advanced with different scanning technologies resulting in different measures. Diffusion Tensor Imaging (DTI) has been shown to characterize AD progression in white matter (Alexander *et al.*, 2007, Klohs *et al.*, 2013). While DTI quantifies the diffusivity of water molecules in the brain microstructure, which is hypothesized to follow a Gaussian distribution, Diffusion Kurtosis Imaging (DKI) aims at simultaneously quantifying both the Gaussian (diffusion tensor) and non-Gaussian (diffusion kurtosis) behaviour of water. Note that the acquisition of DKI is time consuming. Several studies have reported the superiority of DKI over DTI in detecting AD pathology in both white and grey matter (Hui *et al.*, 2008, Cheung *et al.*, 2009, Veraart *et al.*, 2011).

The degree of neuronal myelination was determined using four histology stains: (1) Immunohistochemical visualization of myelin basic protein (MBP), which is the major protein component of myelin sheaths, (2) Glial Fibrillary Acidic Protein (GFAP) as a marker for astrocytes, (3) Ionizing Calcium-Binding Adaptor molecule 1 (IBA-1) as a marker for microglia and (4) 4G8 labelling was performed to detect amyloid-beta in the brains of APP/PS1 transgenic mice.

2.3 Design of the Experiment

For the case studies presented in Chapter 3-7, two sets of experiments were used as illustrated in Figure 2.1. Panel 2.1a illustrates acquisition of longitudinal MRI data for a given experimental unit (mice). MRI was performed at 2 months of age and every

two months thereafter, until 8 months. This resulted in a (possibly) correlated set of MRI parameters for a given animal. For a selected subset of animals in this longitudinal study, in addition to MRI, histology staining was performed at 8 months of age. In order to obtain paired MRI and histology data at other ages, data based on the same animal model was acquired from previous cross-sectional studies (see panel 2.1b). The final paired MRI and histology data comprised of cross-sectional datasets at 2, 4, 6, and 10 months and paired MRI and histology data from the longitudinal study at 8 months.



(a) Longitudinal study design. **(b)** cross-sectional study design.

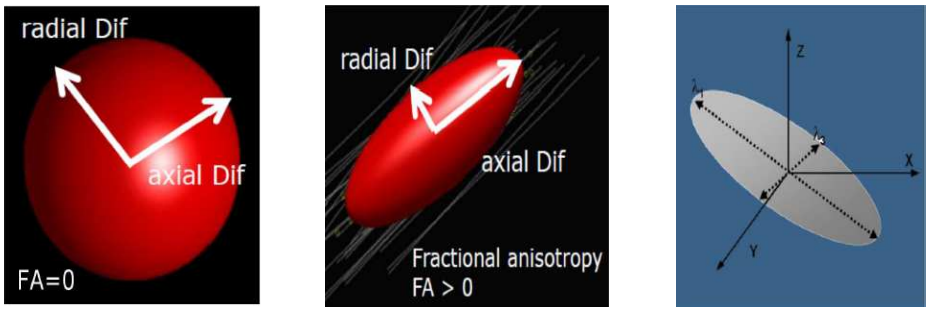
Figure 2.1: Illustration of the study design for MRI and histology data acquisition. The symbols denote the age at which a measurement was taken.

2.4 MRI and Histology Data Acquisition

The experiment comprised two mice genotypes: (1) transgenic APP/PS1 mice which over-expressed the KM670/671NL APP mutation and the L166P PS1 mutation and (2) wildtype mice which represents a healthy control group.

2.4.1 MRI Data Acquisition

Diffusion weighted data were acquired on a 7T Bruker Pharmascan system: 28 slices with resolution $(0.2136 \times 0.214 \times 0.2)\mathrm{mm}^3$; 3 x (20 diffusion gradient directions), $\delta = 5\mathrm{ms}, \Delta = 12\mathrm{ms}$, 7 b-values $(400 - 800 - 1200 - 1600 - 2000 - 2400 - 2800)\mathrm{s/mm}^2$. Seven diffusion parameter maps were estimated (Veraart *et al.*, 2013), including Axial Kurtosis (MRI-AK), Radial Kurtosis (MRI-RK), Mean Kurtosis (MRI-MK), Axial Diffusivity (MRI-AD), Radial Diffusivity (MRI-RD), Mean Diffusivity (MRI-MD) and Fractional Anisotropy (MRI-FA). A population based atlas was created from the 3D T2-weighted anatomical images of a random subset of animals, on which 23 regions of interest (ROIs) were delineated. From the transformation matrices which combine the coregistration of the individual DKI-data to its corresponding 3D data and the normalization of the 3D data to the atlas, the ROI's were back-transformed to each animals DKI data set. The



(a) Isotropic. **(b)** Anisotropic. **(c)** Eigenvectors.

Figure 2.2: Illustration of diffusion tensor metrics used to derive DTI parameters.

final step involved extracting for each subject and each ROI, the diffusion tensor and diffusion kurtosis metrics.

Diffusion Tensor and Diffusion Kurtosis Metrics

Figure 2.2 shows an illustration of the diffusion tensor metrics used to derive the DTI diffusion coefficents. Panel a shows the diffusion components of an unrestricted water molecule, which is spherical, hence fractional anisotropy is zero. Note that the mean diffusion in this case is equal to both the radial and axial diffusion metrics. Within the brain tissue, the flow of water molecules is restricted and assumes an elliptical shape as shown in Panel b. In this case, the relationship between the three diffusion tensor metrics is derived from the corresponding eigenvectors, λ_1, λ_2 and λ_3, shown in Panel c. Note that MRI-FA is the measure of anisotropy, with MRI-FA=0, implying perfectly isotropic diffusion and MRI-FA=1 denoting a hypothetical case of an infinite cylinder.

From the three eigenvectors of the diffusion tensor matrix, the corresponding DTI parameters are given by

$$\begin{aligned} \mathrm{MRI-AD}: &\ \lambda_1, \\ \mathrm{MRI-RD}: &\ (\lambda_2+\lambda_3)/2, \\ \mathrm{MRI-MD}: &\ (\lambda_1+\lambda_2+\lambda_3)/3. \end{aligned} \tag{2.1}$$

DKI is parametrized by 22 elements including the independent diffusion tensor matrices. Note that the model used to derive DKI parameters is an extension of the DTI model to incorporate kurtosis. For DTI, the relationship between the diffusion-weighted signal and the apparent diffusion coefficients si given by

$$\ln\left[S(b)/S_0\right] = -bD_{app}. \tag{2.2}$$

Here, $S(b)$ is the diffusion-weighted signal along a given direction with a certain b value and S_0 is the non-diffusion-weighted signal. The parameter b is specified during MRI acquisition while D_{app} is the apparent diffusion coefficient.

By extending (2.2) to incorporate kurtosis, the DKI model is given by

$$\ln\left[S\left(b\right)/S_0\right] = -bD_{app} + \frac{1}{6}b^2 D_{app} K_{app}. \tag{2.3}$$

The parameter K_{app} is the apparent kurtosis parameter along the given direction, while the other parameters are as defined for DTI.

2.4.2 Histology Data Acquisition

Following the magnetic resonance image acquisition, mice were sacrificed and the brains dissected. Paraffin slices were made from one half of the brain. To allow a 3D histological reconstruction of the brain, 10 paraffin slices of $4\mu m$ thickness were taken at $150\mu m$ intervals, starting from the midline of the brain all the way until the most lateral side of the brain. The first slice of each interval was stained using myelin basic protein (MBP). The latter gives a good idea on the myelination of the brain, but also shows very easily identifiable anatomical locations, which allowed to stack all these images of MBP stained slices. This resulted in a 3D histological stack of the MBP staining, which was co-registered to the 3D T2-weighted anatomical magnetic resonance scan. In addition, the 3 consecutive slices were stained using glial fibrillary acidic protein (GFAP) for astrocytes, ionizing calcium-binding adaptor molecule 1 (IBA-1) for microglia and 4G8 for amyloid beta. These slices were then co-registered to the corresponding MBP stained slide, which eventually allowed to also correlate these 3 markers to the voxel-based MRI.

2.5 Longitudinal and Cross-Sectional Datasets

2.5.1 Longitudinal MRI Data

MRI acquisition on 20 APP/PS1 (transgenic) and 19 wildtype (WT) mice was performed repeatedly in two month intervals from the age of 2 months till 8 months. At each age, three MRI Diffusion Kurtosis Imaging (DKI) and four Diffusion Tensor Imaging (DTI) parameters were computed from the MR scan in each of the 23 brain Regions of Interest (ROI) that had been delineated from the white and grey matter. At each age, the diffusion tensor and diffusion kurtosis tensor were estimated from which, the diffusion maps were calculated. The average diffusion tensor metrics (MD, RD, AD,FA) and diffusion kurtosis metrics (MK, RK, RK) were extracted for each of the 23 regions of the brain. This resulted in a longitudinal structure consisting of four measurements (corresponding to 2, 4, 6, and 8 months) for each MRI parameter. It is also expected that the multivariate set of seven MRI parameters may be correlated since they are measured on the same subject.

The analysis presented in Chapter 3 and 4 of this dissertation focuses on identifying the parameters and ROI for which significant differences of brain tissue structure between the two genotypes can be detected. This implies that $7 \times 23 = 161$ independent longitudinal analyses (for each parameter and ROI) ought to be performed.

2.5.2 Cross-Sectional Paired MRI and Histology Data

In total, five cohorts of mice aged 2, 4, 6, 8 and 10 months were scanned (sample sizes are given in Table 2.1). Data is available from 23 brain regions of interest (ROI) covering both the white and grey matter. For each region, the averages of each of the seven MRI parameters and four histology stains (each quantified by the percentage of area stained) was obtained per animal where possible.

Table 2.1: Summary of the data: Number of animals per age and genotype.

Age (months)	2	4	6	8	10
Transgenic	10	10	9	9	3
Wildtype	2	2	2	2	2

2.6 Analysis Plan

Figure 2.3 presents an overview of the analysis plan for the MRI experiment. The analysis approach is broadly categorized into two sets of analyses. The first set of analyses, presented in Chapter 3 and 4, involves univariate analysis of MRI data. The goal of the analyses presented in these two chapters is to determine whether MRI parameters can be used to distinguish between wildtype and APP/PS1 mice, at which age and with which accuracy. For the longitudinal MRI data (Chapter 3), we wish to identify the age at which significant separation between the evolution profiles for wildtype and transgenic mice is observed. In Chapter 4, a classifier is developed at each age group, with the aim to identify the age at which classification of the mice disease status can be performed with the least misclassification error. Furthermore, in building the classifier, we evaluate whether MRI-DKI parameters improve the classification results compared to a classifier based only on MRI-DTI parameters.

While the analysis presented in Chapter 3 and 4 may show differences between the two mice genotypes, a connection between the observed differences with the actual disease pathology is not established. Chapter 5-7 establishes this connection by jointly modelling the MRI data with the true disease pathology status as quantified by various histology stains. We seek for a linkage between MRI parameters and the disease pathology (histology staining) using the methodology for biomarker evaluation in clinical trials. In Chapter 5, a graphical overview of the implication of surrogacy is presented while the subsequent two chapters describe the implementation of a frequentist and Bayesian analysis for the surrogacy evaluation.

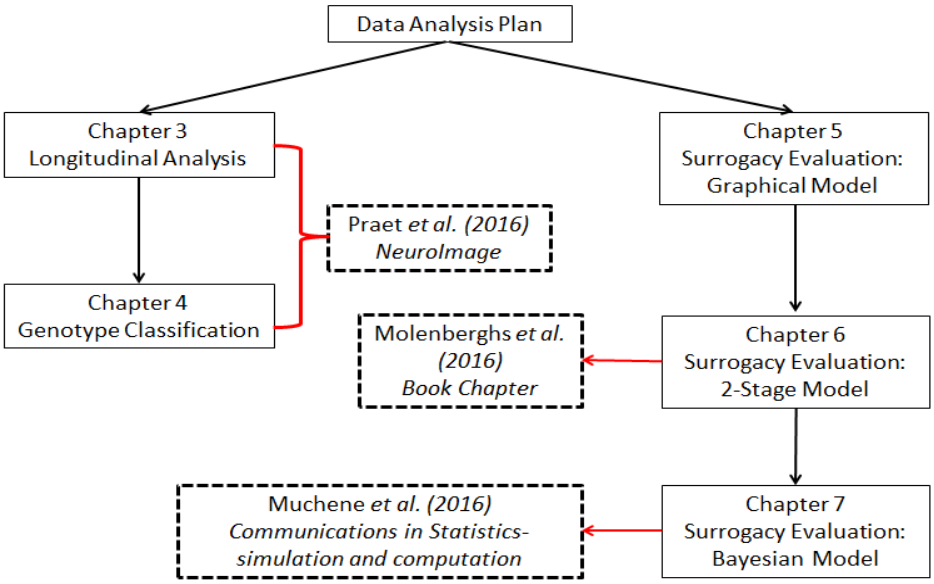


Figure 2.3: MRI experiment: Overview of the analysis plan and the publication strategy.

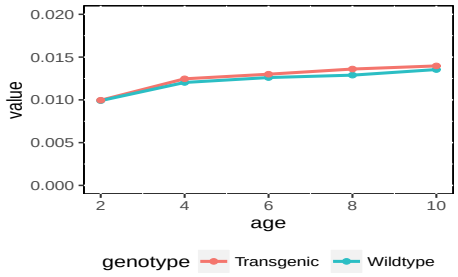
Chapter 3

Longitudinal analysis of Alzheimer's Disease Progression

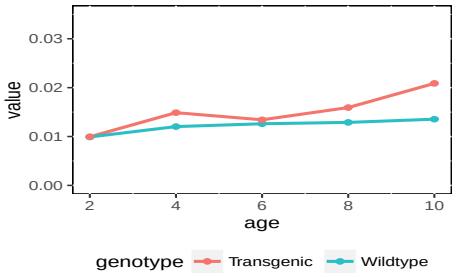
3.1 Introduction

Magnetic resonance imaging is a non-invasive technique that, if validated, can be used as a diagnostic tool for new patients as well as for clinical follow-up of disease progression, since MRI measurements can be acquired on the same patient repeatedly. To perform these roles, MRI parameters ought to: (1) discriminate between diseased and healthy patients and (2) quantify the evolution of underlying disease pathology accurately over time. In this chapter, focus is on the analysis of longitudinal MRI data, with the aim of identifying the difference in evolution profiles for wildtype and transgenic mice. Figure 3.1 presents a graphical illustration of the analysis problem. Panel 3.1a presents a case where the MRI parameter does not quantify a significant difference between the genotypes even though there is progression of disease pathology in the transgenic mice even for older ages. In this case, the accuracy of the MRI parameter in discriminating between transgenic and wildtype mice is expected to be relatively low as seen in Panel c, e and g. In Figure 3.1b, there is an age-dependent shift in the MRI parameter for transgenic mice, which grows with age. This leads to better separation between the transgenic and wildtype mice as age increases (see Panel d, f and h).

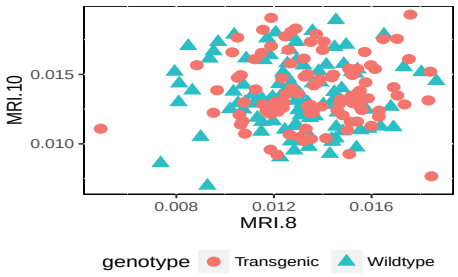
From the longitudinal magnetic resonance imaging data, evolution of the disease pathology was quantified by MRI parameters at 2, 4, 6, and 8 months of age for each rat. At each age, DKI parameters (MRI-AK, MRI-RK, MRI-MK) and DTI parameters (MRI-AD, MRI-RD, MRI-MD, MRI-FA) were obtained for each animal. The aim of the analysis is to identify the age from which there is evidence of a significant difference in MRI parameters between transgenic mice and wildtype mice. For the remaining of this chapter, we focus on longitudinal analysis of the data. Section 3.2 describes the methodology for longitudinal data analysis, while in Section 3.3, the motor cortex data used to illustrate the methodology is presented. Section 3.4 and 3.5 presents the results



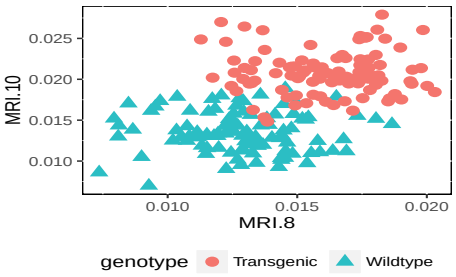
(a) Average profiles without separation.



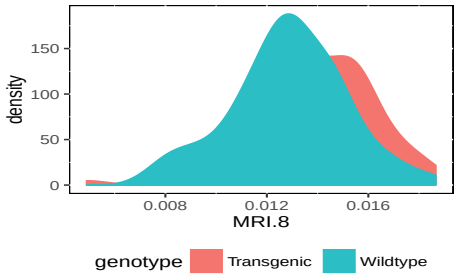
(b) Average profiles with separation.



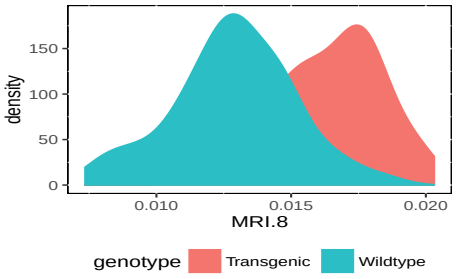
(c) Scatter plot showing no separation.



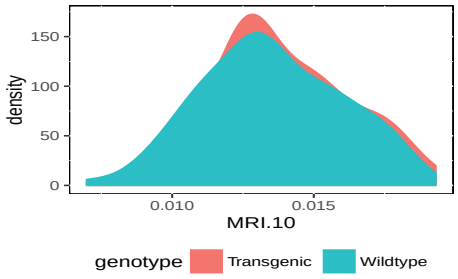
(d) Separation between genotype.



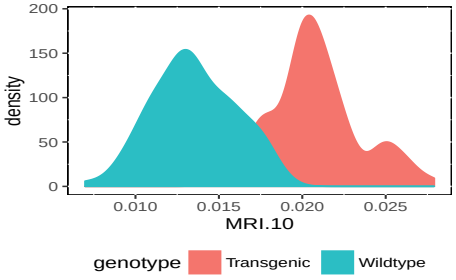
(e) Genotype overlap at 8 months.



(f) Age-dependent evolution (8 months).



(g) Genotype overlap at 10 months.



(h) Age-dependent evolution (10 months).

Figure 3.1: MRI parameter illustration of the age-dependent separation at 8 and 10 months for a given ROI. The scatterplot and density plots presents simulated MRI data (arbitrary units) at 8 and 10 months of age.

of applying the methodology in several brain regions and in Section 3.6, we discuss the results.

3.2 Model Formulation

For a given ROI_r, $r = 1, \ldots, 23$, and MRI parameter m, let Y_{ij} be the outcome of mouse i at age j, $j = \{1, 2, 3, 4\}$ for 2, 4, 6 and 8 months of age, respectively. In matrix notation, the linear mixed-effects model can be denoted as

$$\begin{cases} \mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \\ \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D}), \\ \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i). \end{cases} \tag{3.1}$$

Here, $\mathbf{Y}_i$ is a vector of outcomes of length n_i for subject i, $1 \leq i \leq N$, N is the total number of subjects. the matrix $\mathbf{X}_i$ is an $(n_i \times p)$ design matrix of covariates (genotype and age in this case), $\mathbf{Z}_i$ is an $(n_i \times q)$ dimensional matrix of random effects, $\boldsymbol{\beta}$ is a p-dimensional vector of fixed effects, $\mathbf{b}_i$ is a q-dimensional vector of random effects, and $\boldsymbol{\epsilon}_i$ is a vector of measurement errors. The matrices $\mathbf{D}$ and $\boldsymbol{\Sigma}_i$ are the covariance matrices for the random and fixed effects, respectively.

The expected value and the variance of $\mathbf{Y}_i$ in Equation 3.1 obtained by integrating out the random effects and considering that $E(\boldsymbol{\epsilon}_i) = 0$, $E(\boldsymbol{b}_i) = 0$ and $Cov(\boldsymbol{\epsilon}_i, \boldsymbol{b}_i) = 0$, is given by,

$$\begin{aligned} &E(\mathbf{Y}_i) = \mathbf{X}_i\boldsymbol{\beta}, \\ &Var(\mathbf{Y}_i) = \boldsymbol{\Sigma} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top = \mathbf{V}_i. \end{aligned} \tag{3.2}$$

The covariance matrix comprises of two components: a residual covariance matrix Σ and the random effects matrix component $\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$. The dependence between measurements of the same animal can be captured in either of these two components. For the marginal models presented in this chapter, $\mathbf{V}_i = \Sigma$ where Σ is a 4×4 matrix. The marginal model is given by,

$$\mathbf{Y}_i \sim MVN(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i). \tag{3.3}$$

The mean structure $\mathbf{X}_i\boldsymbol{\beta}$ is specified to be unstructured, thus obtaining an estimate for each genotype and age. In particular, for $j = 1, 2, 3, 4$ for age = 2,4,6,8 months, respectively, the mean structure for the ith animal is given by,

$$E(Y_{ij}) = \beta_{11}X_{1i} + \beta_{12}X_{2i} + \beta_{13}X_{3i} + \beta_{14}X_{4i} \\ +\beta_{21}X_{1i}W_i + \beta_{22}X_{2i}W_i + \beta_{23}X_{3i}W_i + \beta_{24}X_{4i}W_i,$$

while,

$$X_{ij} = \begin{cases} 1, \text{age} = j, \\ 0, \text{otherwise}, \end{cases} \tag{3.4}$$

and

$$W_i = \begin{cases} 1, \text{if transgenic mice}, \\ 0, \text{if wildtype mice}. \end{cases}$$

The regression coefficients $\{\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}\}$ are the age-specific mean MRI responses for wildtype mice and $\{\beta_{21}, \beta_{22}, \beta_{23}, \beta_{24}\}$ are the age-specific regression coefficients for the transgenic mice. The parameters of primary interest- the age-specific contrasts γ_j, between transgenic and wildtype mice- are given by,

$$\gamma_j = \beta_{2j} - \beta_{1j}. \tag{3.5}$$

Multiple statistical tests are performed corresponding to the four timepoints for each parameter and ROI. Multiplicity adjustment can be performed at two levels: (1) adjustment for 4 tests in total corresponding to the tests between genotypes at 2, 4, 6 and 8 months and (2) taking into account that for a given MRI parameter, $4 \times 23 = 96$ tests are performed in the whole brain hence adjusting for 92 tests per parameter. Both approaches were considered and False Discovery Rate (FDR, Benjamini and Hochberg, 1995) adjustment was performed in order to control the false-positive error-rate.

3.3 The Motor Cortex Data

For illustration purposes, longitudinal analysis is first applied to the motor cortex data. The motor cortex is a region in the brain primarily involved in planning, control and execution of voluntary movements. In AD patients, the motor cortex suffers from neural degeneration, amyloid plaque deposition and neurofibrillary tangles (hyperphosphorylated tau protein formation) (Hol *et al.*, 2003). Depending on the degree of disease progression, there might be clinically observable changes in motor coordination in patients, although this might only occur at later stages of the disease. Note that the APP/PS1 mouse model presents no neurodegeneration.

The individual profile plots for the seven MRI parameters in the motor cortex are presented in Figure 3.2 which shows that, the different MRI parameters exhibit different evolution patterns over time. Figure 3.3 presents the average profile plots from which, potentially significant separation in the profiles is observed as from four months of age with MRI-MK, MRI-RK, MRI-AK and MRI-FA. For the diffusivity parameters MRI-MD, MRI-RD and MRI-AD, there is no indication of separability between the two genotypes as the animals grow older.

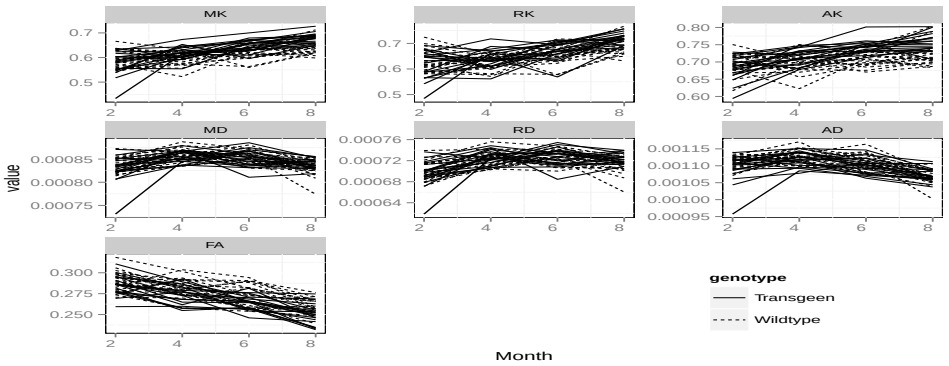


Figure 3.2: Individual profile plots for the seven MRI parameters in the motor cortex. DKI/DTI units: mm^2/s.

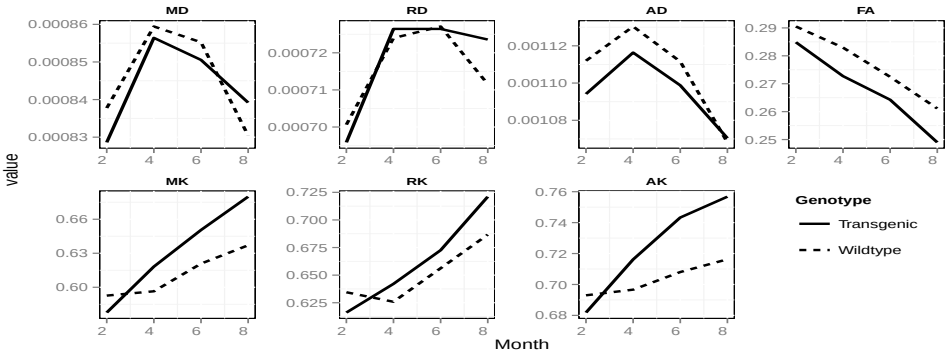


Figure 3.3: Average profile plots for MRI parameters in the motor cortex region. DKI/DTI units: mm^2/s.

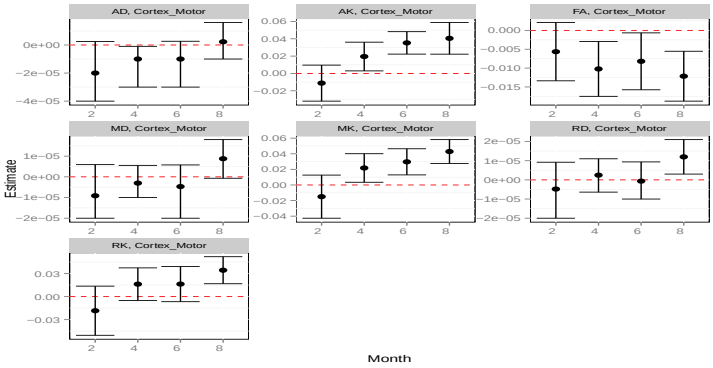


Figure 3.4: Estimated contrasts between genotypes with the 95% confidence interval included for the cortex motor region in the brain. Confidence intervals not including zero (dashed horizontal line) are statistically significant. DKI/DTI units: mm^2/s.

3.4 Analysis of Longitudinal MRI Data From the Motor Cortex

The goal of the analysis presented in this section is to identify the age at which significant differences between healthy and diseased mice can be detected, using MRI parameters. Each of the DKI (MRI-AK, MRI-RK, MRI-MK) and DTI (MRI-AD, MRI-RD, MRI-MD, MRI-FA) parameters is evaluated separately. Data was available from 19 transgenic and 20 wildtype mice, which were observed at all four planned timepoints. In Section 3.5, results of applying the methodology to all the 23 brain regions are graphically summarized.

Figure 3.4 presents the estimated contrasts between the genotypes at each age, with the corresponding 95% confidence interval. From these results, it is evident that DTI parameters (first row), have the least ability to discriminate between wildtype and transgenic mice even for older animals. DKI parameters on the other hand, show good separation between the genotypes especially for MRI-MK and MRI-AK. Finally, for these MRI parameters where significant separation is evident (see next section), the disease progression can be detected from the increase in magnitude of the disease effect on the MRI parameters.

3.4.1 Multiplicity Adjustment

To control the *type I* error rate, FDR adjustment on the set of four contrasts in the motor cortex region was performed for each MRI parameter separately. Figure 3.5 presents the FDR adjusted P-values for the test of a difference in the MRI readout between transgenic and wildtype at each age. Results indicate that, for MRI-AK, MRI-MK, and MRI-FA, significant differences between the genotypes are detectable at 4, 6 and 8 months in the motor cortex. With MRI-RD and MRI-RK however, significant differences are only detectable at 8 months of age.

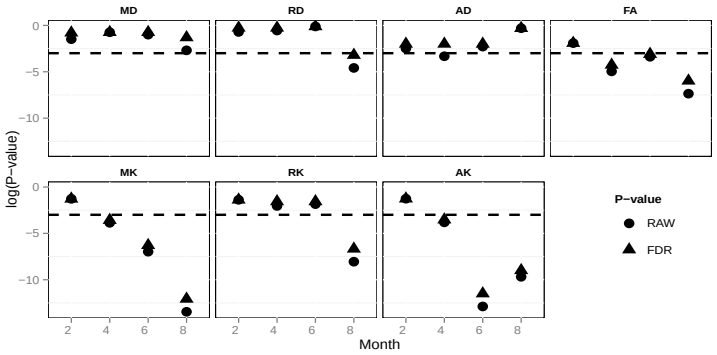


Figure 3.5: FDR-adjusted P-values for the contrasts between the two genotypes. Note that the natural logarithm of the P-values is plotted, hence significant P-values are below the dashed line. The FDR adjustment is performed for four tests per MRI parameter.

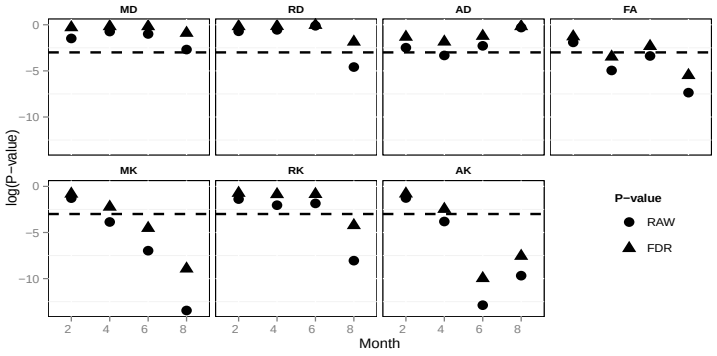


Figure 3.6: FDR-adjusted P-values for the contrasts between the two genotypes. Note that the natural logarithm of the P-values is plotted, hence significant P-values are below the dashed line. The FDR adjustment is performed for 92 tests per parameter.

Alternatively, for each MRI parameter, FDR adjustment can account for the fact that measurements are obtained from all the 23 ROI. In this case, there are $23 \times 4 = 92$ tests. FDR adjustment taking into account multiple testing across the 23 brain regions in addition to the four age groups resulted in more conservative results as shown in Figure 3.6. Significant differences between genotype are detectable at six and eight months with MRI-AK and MRI-MK, at four and eight months with MRI-FA, and at eight months only with MRI-RK.

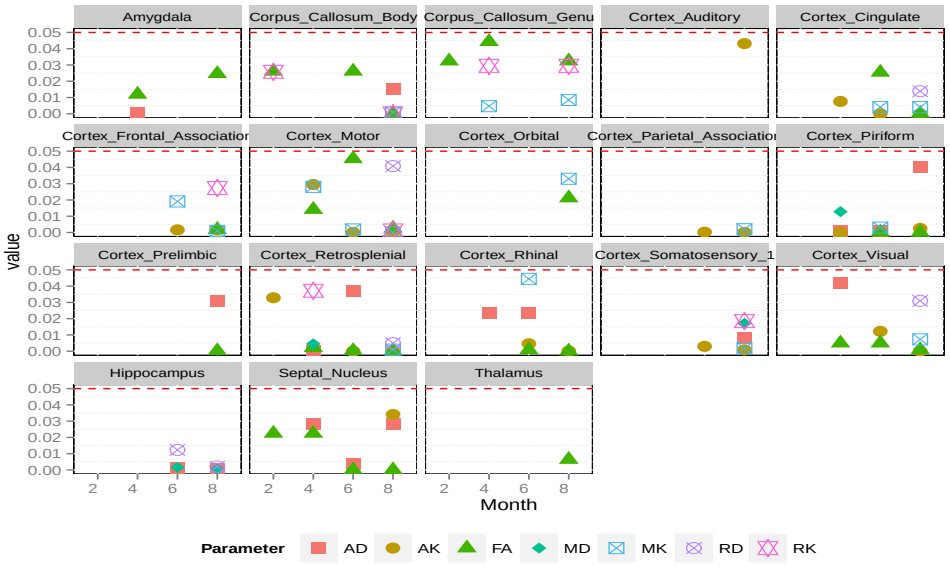


Figure 3.7: FDR-adjusted P-values for the contrasts between the two genotypes. Only significant p-values are plotted. FDR adjustment in this case only adjusts for 4 tests per MRI parameter.

3.5 Application of the Methodology to all Regions of Interest

Figure 3.7 shows the adjusted p-values for parameters and brain regions of interest in which statistically significant differences between genotype were detectable. Results indicate that statistically significant differences were detected in 18 of the 23 regions of interest in the brain. Most of the significant differences were detectable in the cortex and corpus callosum brain regions. Significance depends on the MRI parameter and age of the animals although most significant differences were obtained at 6 and 8 months.

A more conservative adjustment was performed by controlling for $23 \times 4 = 92$ tests corresponding to the fact that for each parameter, 4 tests were performed in 23 brain regions of interest. In Figure 3.8, results of the regions with statistically significant differences between genotype are presented. Significant differences were mainly detectable at 6 and 8 months although for some regions of interest (such as the motor cortex, amyglada, cortex visual and cortex orbital), significant differences were detectable at 4 months of age. Moreover, most of the significant differences are observed for MRI-AK, MRI-MK and MRI-FA.

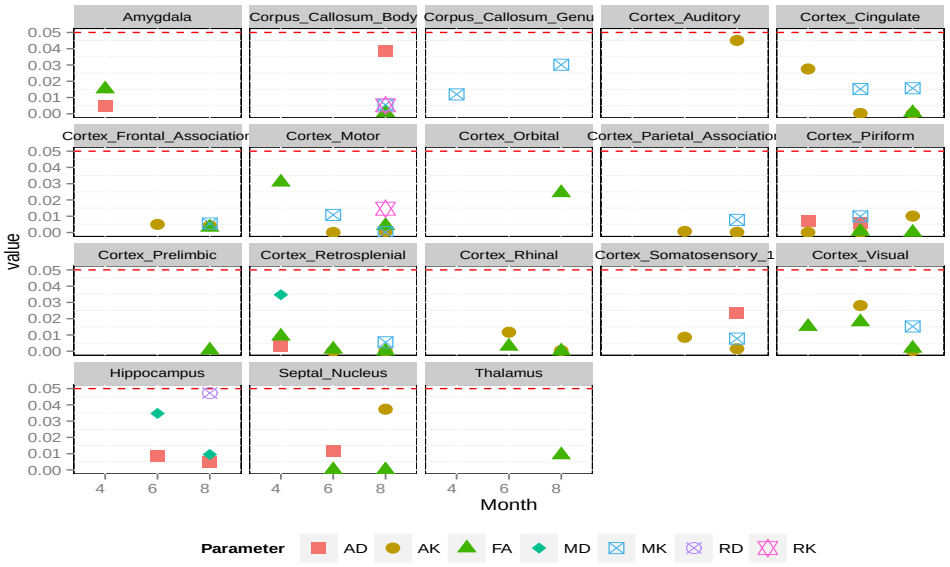


Figure 3.8: FDR-adjusted P-values for the contrasts between the two genotypes. Only significant p-values are plotted. FDR adjustment in this case adjusts for 92 tests per MRI parameter.

3.6 Discussion

Although the disease pathology is expected to evolve with time, hence a difference is expected between wildtype and transgenic mice, not all MRI parameters are able to detect this difference in evolution profiles. In particular, the ability for MRI to detect significant differences in the evolution profile depends on the specific MRI parameter used, the age at which the difference is evaluated and the region of interest. For some brain regions such as the caudate putamen and cerebellum, none of the MRI- parameters detects a significant difference between the genotype. Note that this does not necessarily imply that the regions do not have disease pathology developing. This will be further investigated in Chapter 6. In this chapter, we focused on longitudinal data analysis, while in the next chapter, we investigate if the data can be used in order to classify subjects to their genotypic group.

In this chapter, a marginal model for the MRI parameters was formulated. That is, given a vector of responses $\mathbf{Y_i}$ for the kth MRI parameter in the ith rat,

$$\mathbf{Y_i} \sim Normal(X_i\boldsymbol{\beta}, \Sigma). \tag{3.6}$$

The association between the measurements are captured by Σ. Alternatively, a Linear Mixed-effects Model (Verbeke and Molenberghs, 2000, LMM,) can be formulated. From the LMM, the marginal distribution of Y is given by,

$$\begin{aligned} &\mathbf{Y}_i = X_i\boldsymbol{\beta} + Z\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \\ &\mathbf{b}_i \sim Normal\left(0, D\right), \\ &\boldsymbol{\varepsilon}_i \sim Normal\left(0, \Sigma_i\right). \end{aligned} \tag{3.7}$$

Chapter 4

Classification of Alzheimer's Disease Status Using MRI Parameters

4.1 Introduction

Developing a predictive tool for Alzheimer's disease is of interest especially for the purpose of early diagnosis, disease follow-up and management. The analysis presented in Chapter 3 indicates that MRI parameters have the potential to discriminate between wildtype and transgenic mice, although this depends on the MRI parameter of interest, brain region of interest and age of the mice. The aim of the analysis presented in this chapter is to develop a classification tool, based on MRI parameters, to predict the mice genotype (diseased or healthy). In particular, the primary interest is to investigate whether either Diffusion Kurtosis Imaging (DKI) parameters or the Diffusion Tensor Imaging (DTI) parameters lead to a more accurate classification. Moreover, we evaluate the additive value of DKI parameters in the classification model.

In machine learning, two approaches to discovering clusters, patterns or trends in observed data are popular: supervised and unsupervised learning algorithms. In supervised learning (for instance, Linear Discriminant Analysis, LDA), the class membership is provided to the algorithm during the training phase. On the other hand, unsupervised learning algorithms such as clustering algorithms (such as K-means clustering and hierarchical clustering) evaluates the existence of groups of similarly associated observations without the need for providing the true class membership in the training phase. For both approaches, cross-validation methods are applied by splitting the data into training and test sets (Hastie *et al.*, 2003). The model is fitted on the training set and classification performed on the test set.

Development of a classifier for the MRI data poses some practical challenges. Based on the analysis presented in Chapter 3, it is plausible that the performance of the

classifier varies with the brain region under evaluation. This implies the need to perform the classification task by brain region of interest and evaluate the most promising regions for the classification. In order to find the age at which an accurate classification is possible, the classification procedure presented in this chapter is performed at each timepoint separately. It is expected that the misclassification error-rate will be lower for older animals as the disease pathology is more developed in older mice. Alternatively, model-based classification algorithms such as the functional data analysis (Ramsay *et al.*, 2009) and in particular, the functional linear discriminant analysis algorithm of Gareth M. James (2001), mixture of random-effects models for linear discriminant analysis by Komárek *et al.* (2010) amongst others accounts for the longitudinal effects in performing classification.

In this chapter, we apply Fisher's Linear Discriminant Analysis (LDA, Fisher, 1936) to MRI data at each age. At each age, LDA with the DKI parameters only (MRI-AK, MRI-MK, and MRI-RK), DTI parameters only (MRI-AD, MRI-MD, MRI-RD and MRI-FA) and finally, LDA with all seven MRI parameters is performed. Appropriate measures are defined to evaluate the added benefit DKI parameters have on the classification algorithm as compared to using only the DTI parameters. The analysis algorithm and results are illustrated with the motor cortex data, while the results of applying the methodology to all the 23 brain regions are presented in the final section of this chapter.

4.2 Genotype Classification Using Linear Discriminant Analysis (LDA)

4.2.1 Data Structure

Linear Discriminant Analysis (Hastie *et al.*, 2003, Mclachlan, 2004, LDA) is a supervised learning method for which a classifier is trained and test data applied in order to derive the class membership. Let the class membership for animal i be defined as

$$Y_i = \begin{cases} 1, \text{ Transgenic APP/PS1,} \\ 0, \text{ Wildtype.} \end{cases} \tag{4.1}$$

Given a set $\mathbf{X}$ of predictors, LDA defines a linear function associating the predictor matrix with the class membership. Further, a classification rule, $C(\mathbf{X})$, based on a threshold τ, is defined in order to predict the class membership (Hastie *et al.*, 2003) such that,

$$\begin{cases} C(\mathbf{X}) > \tau, \text{if subject is classified as Transgenic,} \\ C(\mathbf{X}) \leq \tau, \text{if subject is classified as Wildtype.} \end{cases} \tag{4.2}$$

For the MRI experiment, the predictor matrix comprises of DTI parameters and DKI parameters, respectively, given by,

$$\begin{aligned} \mathbf{X} &= [X_1 | X_2], \\ \mathbf{X}_1 &= [AD, FA, RD], \\ \mathbf{X}_2 &= [AK, MK, RK]. \end{aligned} \tag{4.3}$$

Table 4.1: Illustration of test data misclassification error based on LDA.

		True status	
		Transgenic	Wildtype
LDA	Transgenic	a	c
	Wildtype	b	d
		T	W

Note that MRI-MD is not used in building the classifier in $\mathbf{X_1}$, since by definition, MRI-MD is a linear function of MRI-AD and MRI-RD. For the classifiers, let $C(\mathbf{X_1})$ be the LDA classifier based on DTI parameters, $C(\mathbf{X_2})$ is the LDA classifier based on DKI parameters and $C(\mathbf{X_{12}})$ is the LDA classifier based on both DTI and DKI parameters.

4.2.2 Cross-Validation

Two-fold cross-validation was performed by randomly splitting the data into a training and test sample. The training dataset comprised of 10 transgenic and 10 wildtype mice, while the test dataset comprised of 9 transgenic and 10 wildtype mice. We use the misclassification Error (MCE) to evaluate the accuracy of the classification procedure on the test dataset. Table 4.1 presents a hypothetical classification problem.

From Table 4.1, the MCE is defined as

$$MCE = \frac{(b+c)}{T+W}. \tag{4.4}$$

The cross-validation procedure above is repeated for 1000 iterations, each resulting in a random sample of test and training datasets. At each iteration the classifier is trained on the training dataset and evaluated on the test dataset. Three different classifiers were developed: (1) a classifier based on DTI parameters MRI-AD, MRI-RD and MRI-FA, (2) a classifier based on DKI parameters MR-AK, MRI-RK, MRI-MK and (3) a classifier based on both DTI and DKI parameters. Furthermore, in order to assess the relative importance of DKI over DTI, several statistics were defined and discussed below.

4.2.3 Average Misclassification Error (AMCE)

As mentioned above, a loop of $B = 1000$ two-fold cross-validated samples was used resulting in $MCE_1 \dots MCE_B$ from which the average misclassification error (AMCE) is defined as

$$AMCE = \frac{1}{B}\sum_{b=1}^{B} MCE_b. \tag{4.5}$$

4.2.4 Relative Change in Misclassification Error

We define the Relative Change in Misclassification Error (RCMCE) at iteration b as,

$$RCMCE_{1,b} = \theta_{DKI,b} = \frac{MCE(\mathbf{X}_1)_b - MCE(\mathbf{X}_2)_b}{MCE(\mathbf{X}_1)_b},$$
$$RCMCE_{2,b} = \theta_{DKI+DTI,b} = \frac{MCE(\mathbf{X}_1)_b - MCE(\mathbf{X}_{12})_b}{MCE(\mathbf{X}_1)_b}. \tag{4.6}$$

Where $\theta_{DKI,b}$ is the change in misclassification error when the classifier is based on DKI parameters compared to the misclassification error when the classifier is based on DTI parameters. Similarly, $\theta_{DKI+DTI,b}$ is the relative change in the misclassification error when the classifier is based on both DKI and DTI parameters compared to the misclassification error when the classifier is based on DTI parameters.

4.2.5 Comparisons Between DKI and DTI

Let I_{b_1} and I_{b_2}, be an indicator variable such that

$$I_{b_1} = \begin{cases} 1, \text{if } \mathrm{MCE}\left(X_2\right) < \mathrm{MCE}\left(X_1\right), \\ 0, \text{otherwise.} \end{cases}$$
$$\text{and} \tag{4.7}$$
$$I_{b_2} = \begin{cases} 1, \text{if } \mathrm{MCE}\left(X_{12}\right) < \mathrm{MCE}\left(X_1\right). \\ 0, \text{otherwise.} \end{cases}$$

Note that the proportion of cross-validated samples for which $I_{b_1} = 1$ is a measure of the improvement in classification based on DKI parameters rather than a classification based on DTI alone.

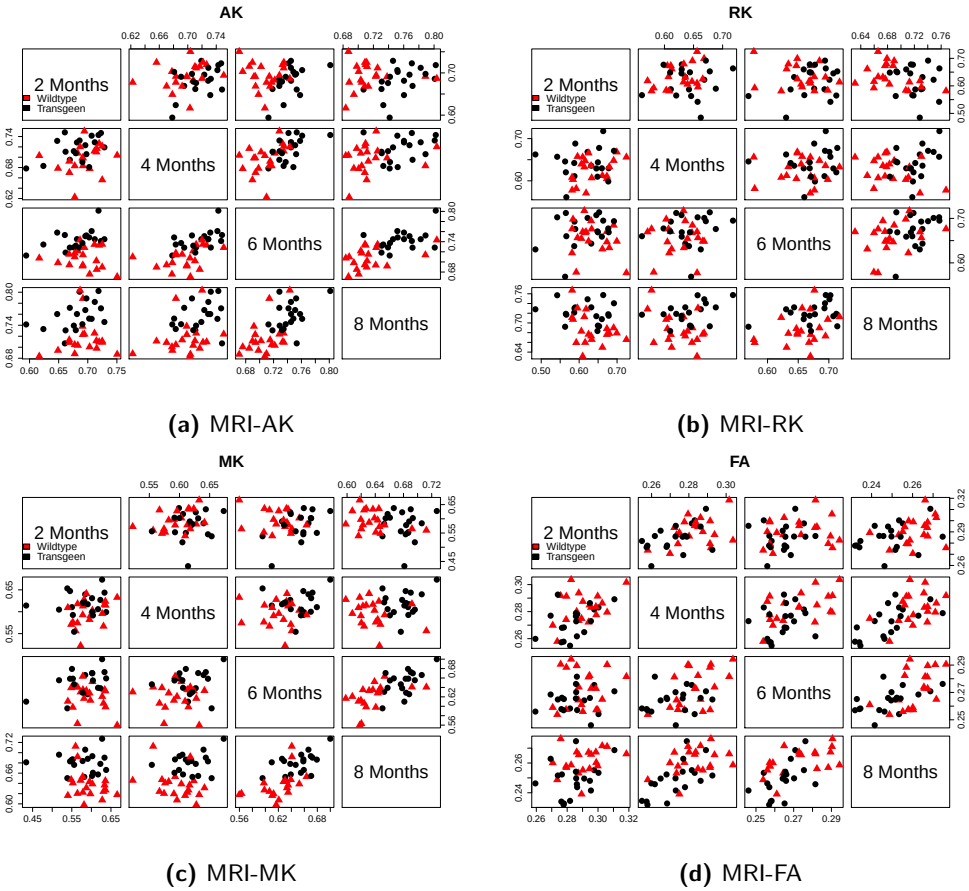
$$P(\theta_{DKI,b_1}) = \frac{1}{B}\sum_{b=1}^{B} I_{b_1}. \tag{4.8}$$

Similarly, the proportion of samples for which $I_{b_2} = 1$, is a measure of the improvement in classification when both DKI and DTI parameters are used in classification compared to when only DTI is used to build the classifier.

$$P(\theta_{DKI+DTI,b_2}) = \frac{1}{B}\sum_{b_2=1}^{B} I_{b_2}. \tag{4.9}$$

4.3 Classification Based on MRI Parameters in the Motor Cortex Region

The results of the analysis presented in Chapter 3 indicate that significant separation between the genotypes highly depends on the age of the animals. In the motor cortex



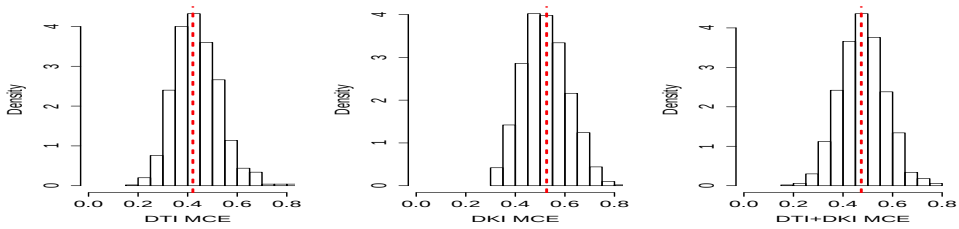
(a) MRI-AK **(b)** MRI-RK

(c) MRI-MK **(d)** MRI-FA

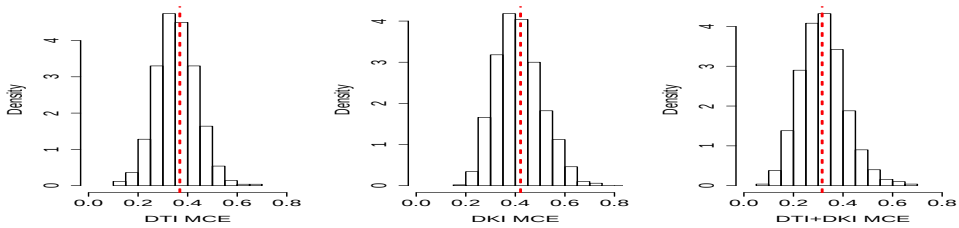
Figure 4.1: Scatterplot matrix for MRI-AK, MRI-RK, MRI-MK and MRI-FA in the motor cortex region. Triangles: wildtype. Circles: transgenic. DKI/DTI units: mm^2/s.

region, MRI-MK, MRI-RK, MRI-AK and MRI-FA showed significant differences at 8 months while at 6 months, only MRI-MK and MRI-AK were found to be statistically significant. The scatterplot matrix of the DKI parameters MRI-AK, MRI-RK, MRI-MK and MRI-FA is presented in Figure 4.1. The separation between transgenic and wildtype mice improves with age, with 6 versus 8 month data showing the best separation ability in all the parameters.

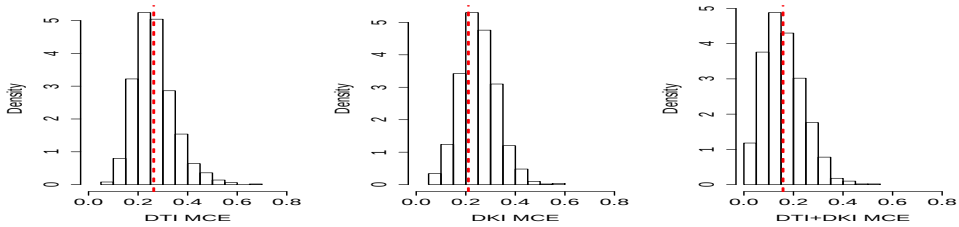
Linear discriminant analysis was performed on 2 months, 4 months, 6 months and 8 months data separately. For each age group, classification based on DTI parameters only, DKI parameters only or a combination of both DKI and DTI parameters was performed. Figure 4.2 presents the distribution of MCE for the test set in the motor cortex region. In all three classification models, the MCE reduces with increasing age, implying that better discrimination of the disease status is obtained in older animals than in younger animals.



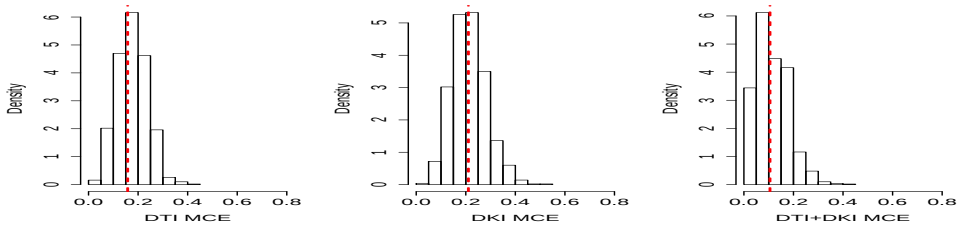
(a) Test data MCE for classification based on 2 months data.



(b) Test data MCE for classification based on 4 months data.



(c) Test data MCE for classification based on 6 months data.



(d) Test data MCE for classification based on 8 months data.

Figure 4.2: Distribution of misclassification error based on linear discriminant analysis at 2, 4, 6 and 8 months in the motor cortex. The MCE is measured in the test datasets in a two-fold cross validation sample. In total 1000 cross-validated samples were used. The vertical dashed line represents the average MCE. First column: LDA with DTI parameters only. Second column: LDA with DKI parameters only. Last column: LDA with both DTI and DKI parameters.

Table 4.2: Summary statistics for test misclassification error in the motor cortex region.

	AMCE			RCMCE		$P(\theta)$	
Age	DTI	DKI	DTI+DKI	DKI	DTI+DKI	DKI	DTI+DKI
2	0.43	0.52	0.48	-0.278	-0.142	0.183	0.266
4	0.35	0.41	0.31	-0.299	0.182	0.256	0.579
6	0.25	0.24	0.14	0.094	0.695	0.452	0.809
8	0.16	0.2	0.09	-0.364	0.624	0.259	0.701

Summary statistics of the misclassification errors for the test data at different months for the three LDA classifiers are shown in Table 4.2. As earlier stated, average MCE reduces as the animals grow older regardless of the LDA classifier applied. Moreover, although the average MCE from a model with only DKI is higher than that of DTI only, the average MCE improves substantially if both DTI and DKI are used in classification. On average, at 6 and 8 months, there is a change of misclassification error of 0.695 and 0.624, respectively, if both DKI and DTI parameters are used for classification. The change is observed in 80.9% (age 6 months) and 70.1% (age 8 months) of the cross-validated samples.

4.4 Summary of Results in all Brain Regions of Interest

LDA was applied to all regions of interest for every age. The test misclassification error based on two-fold cross-validation was computed for the three classifiers presented in Section 4.2. Figure 4.3 presents the AMCE for different regions of interest in the brain. As expected, the AMCE depends on the region of interest with cortex regions specifically motor cortex, retrosplenial cortex, piriform cortex and rhinal cortex having the largest change in misclassification rate as the animals age. Moreover, for all the ROI, classification based on DKI parameters is at least as good as classification based on DTI parameters with regards to AMCE and classification based on both DTI and DKI parameters provides the smallest misclassification error in some of the regions of interest. The relative change in misclassification error is presented in Figure 4.4, while Figure 4.5 presents the proportion of change over the cross-validated samples. The results indicate that only in a few ROI such as the motor cortex and cortex rhinal does DKI in combination with DTI provide a significant additional accuracy in classification.

4.5 Discussion

The classifiers presented in this chapter were based on LDA performed on data for each age separately. From the results, the age at which classification is performed influences the accuracy of the genotype classification whereby, classification improves for older mice. Note that different regions in the brain for which classification is performed have different misclassification error for the same classifier. From three classifiers based on DTI only,

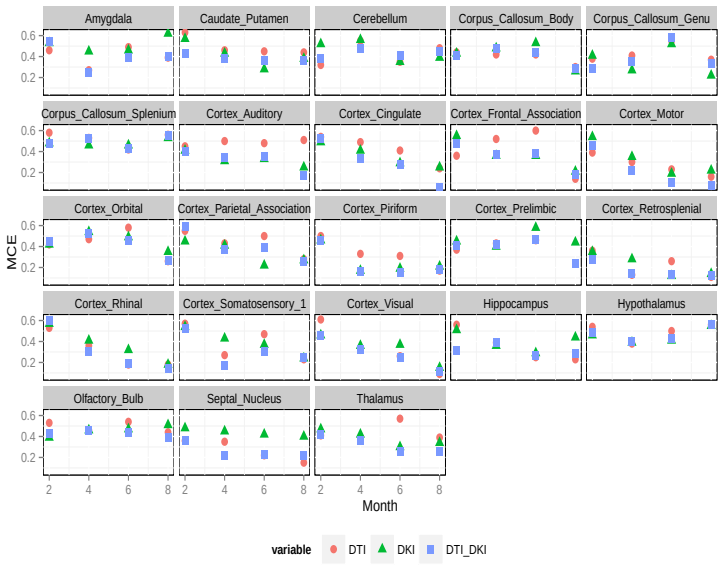


Figure 4.3: Misclassification error for all the regions of interest in the brain. Three LDA models are presented for each age.

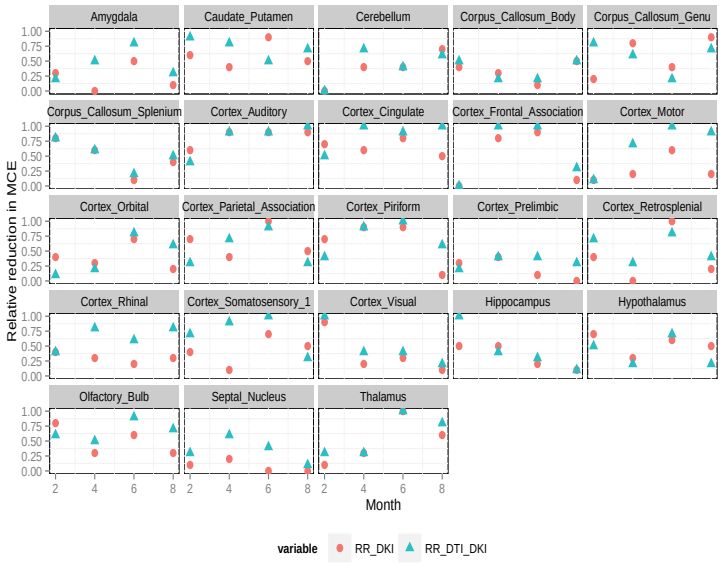


Figure 4.4: RCMCE for all ROI.

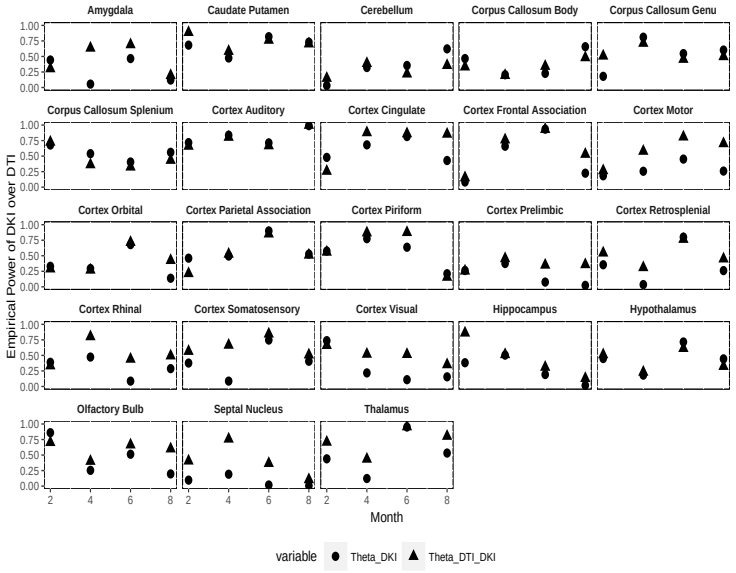


Figure 4.5: Proportion of improvement for all ROI.

DKI only or both DTI and DKI parameters, it was evident that the relative change in misclassification error was largest when DKI was used in combination with DTI. Alternatively, rather than use the observed MRI parameters data at each age for classification, an analysis based on the magnitude of the change from baseline (2 months) can be performed.

Chapter 5

Development of MRI Biomarkers for AD: A Graphical Tour

5.1 Surrogacy in the Context of AD

In the context of randomized clinical trials, biomarkers for 'true' clinical endpoints are often of interest to researchers (Burzykowski *et al.*, 2005). Several reasons may cause the need to establish a surrogate marker to a "true" endpoint. For example, some clinical endpoints take more time to achieve (for example, overall survival versus progression-free survival) or the "true" endpoint may be expensive to measure. Advances in the validation of surrogate endpoints as biomarkers for clinical endpoints have seen a wide array of measures being developed. For instance, Prentice (1989) presented a formal operational definition of a surrogate endpoint as a variable such that a test for the null hypothesis for no treatment effect is also valid for the hypothesis based on the "true" endpoint. Freedman *et al.* (1992) extended this notion by deriving the proportion of the treatment effect that a surrogate endpoint explained, thus allowing for the fact that a surrogate endpoint may not necessarily explain all the treatment effect observed in the "true" endpoint.

Buyse and Molenberghs (1998) further introduced quantities such as the relative effect of the treatment on the "true" and surrogate endpoints and a measure of the association between the surrogate and "true" endpoints after adjusting for the treatment. In order to take into account the additional complexity posed by clinical trials, for instance, multiple centres, Buyse *et al.* (2000) proposed a meta-analytic framework for the evaluation of surrogate endpoints. Under this framework, the individual-level surrogacy is obtained upon adjusting for the multiple experimental units, while the relative effect derived in a single trial setting is replaced with a trial-level surrogacy measure.

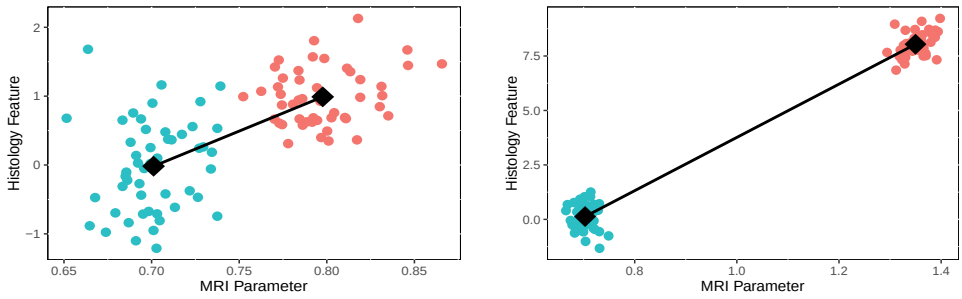
In the context of Alzheimer's disease, the validation of biomarkers has been of interest on several fronts. First, there is a need for the development of reliable diagnostic tests for Alzheimer's disease since currently, clinical evaluation of the disease relies on administration of psychological evaluation tools such as the Mini-mental State Examination, administration of family and primary caregiver questionnaires amongst others (Sabbagh *et al.*, 2010, McKhann *et al.*, 1984). These tools are complex to administer, especially while dealing with patients with cognitive challenges. Moreover, the early stages of Alzheimer's disease are difficult to distinguish from other forms of dementia, therefore rendering these tools less specific.

Laboratory diagnostic tests either based on Cerebral Spinal Fluid (CSF) markers such as amyloid-beta protein deposition, total tau and hyperphosphorylated tau are currently the most widely accepted markers for Alzheimer's disease diagnosis (Humpel, 2011). PET scan, which uses a radioactive tracer, is also used to visualize the plaque load. However, CSF collection is an invasive procedure with potential side effects involved, therefore hampering it's use as a patient follow-up tool. Although body fluids such as blood, saliva and urine-which can be easily collected- are commonly used in managing other diseases, such success has not been replicated in Alzheimer's disease. In fact, blood-based biomarkers with properties at least as superior as the CSF markers have not been successfully validated (Humpel, 2011, Doecke *et al.*, 2012, Snyder *et al.*, 2015).

In this chapter, we introduce the surrogacy setting in the context of AD in which MRI parameters are evaluated as potential biomarkers for histology. The surrogacy setting discussed in this chapter is slightly different from the surrogacy setting in randomised clinical trials. First, rather than multiple trials/centers, the replication unit of the experiment is the age of the animals. Moreover, often in clinical trials, interest is in validating a single surrogate endpoint, while in our case, multiple MRI markers are to be validated, for different parts of the brain separately. Despite these differences, we propose to use the surrogacy framework for normally distributed endpoints (Buyse *et al.*, 2000) in the evaluation of MRI parameters as biomarkers for specific histology features, in several regions of the brain. Using the terminology from the surrogacy framework, each histology feature denotes the "true endpoint", while each MRI parameter denotes a surrogate endpoint.

5.2 Two Levels of Surrogacy

Similar to the multi-centre clinical trial surrogacy setting, we define two measures of surrogacy for the Alzheimer's Disease (AD) MRI experiment . Panel a and b in Figure 5.1 show an illustrative example in which a histology feature is plotted against a specific MRI parameter at 2 and 8 months, respectively. The effect of the disease progression is translated into a shift in both the MRI parameter and histology feature in the transgenic group. Panel a corresponds to a scenario with a relatively small disease effect on the "true" endpoint, while Panel b denotes a scenario with a significant disease effect on both the "true" and the surrogate endpoints. Note that the slope of the lines connecting



(a) Simulated data at 2 months.

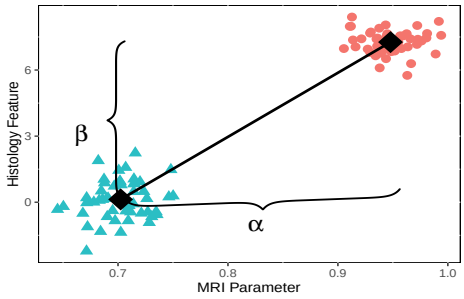
(b) Simulated data at 8 months.

Figure 5.1: Illustrative example. The effect of AD progression on an MRI parameter and a specific histology feature at two time points for simulated data (arbitrary units). The solid line connects means of the two genotypes. Blue symbols: wildtype mice. Red symbols: transgenic mice.

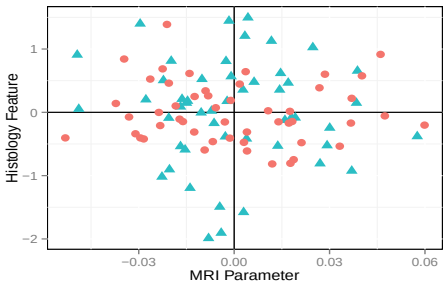
the means of the two clouds in each panel corresponds to the relative effect (RE) as defined by Buyse and Molenberghs (1998).

Figure 5.2 illustrates two aspects of the association between an MRI parameter and a given histology feature: the effect of AD progression (for instance, characterized by additional amyloid-beta deposition) and the correlation between the two variables. Panels a and b show the data and the residuals after subtracting the means, respectively. The AD progression effect on both endpoints (denoted by alpha and beta), can be seen clearly in panel a, while panel b indicates that, conditional on AD progression effect, the two variable are not correlated. Using surrogacy terminology, Figure 5.2b indicates that, on an *individual level*, MRI is a poor biomarker for histology. A second illustrative example is shown in Figure 5.2 c and d. For this example, the AD progression effect (shown in Figure 5.2c) has the same magnitude of the effects as in the first example. Figure 5.2d reveals a substantial difference between the two examples on an individual level. For the second example, after adjusting for the AD progression effects, MRI and histology are correlated (Figure 5.2d). Hence, at an individual level, MRI is a good biomarker for histology.

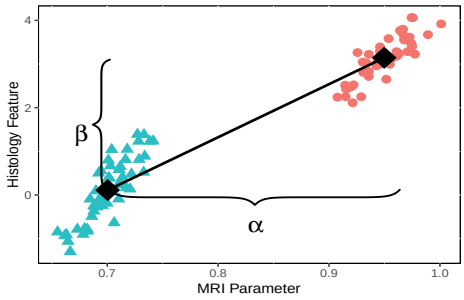
In the next example, presented in Figure 5.3 and 5.4, we "translate" the two aspects of the association between MRI and histology into two surrogacy measures: individual-level surrogacy and disease-level surrogacy. The latter corresponds to the trial-level surrogacy, discussed in Burzykowski *et al.* (2005). The examples presented in Figure 5.2 correspond to a single trial setting and allows us to evaluate the quality of MRI as a biomarker for histology only at individual level. On the other hand, the animal model for AD allows us to estimate surrogacy measures in both levels since MRI parameters and histology features are measured at 5 ages. Figure 5.3 shows a scenario in which an MRI parameter and a histology feature are not correlated, given the progression effect of AD but the disease effects (on both MRI and histology), shown in panel c, are correlated. This suggests a scenario for which the disease-level surrogacy is high (implying that there is an association between disease-effects as measured by the two endpoints) while



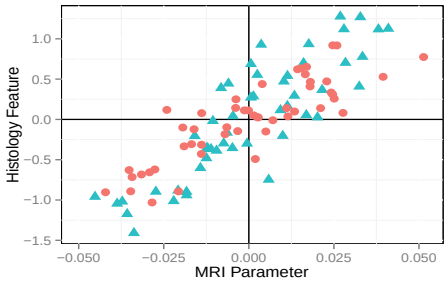
(a) Histology versus MRI.



(b) Residuals after subtracting the group means.

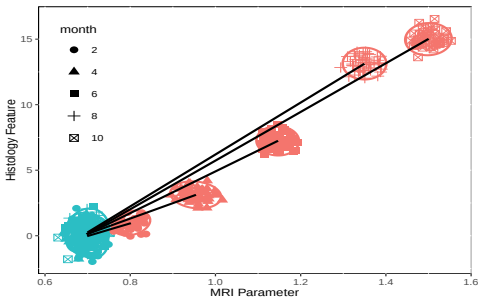


(c) Histology versus MRI.

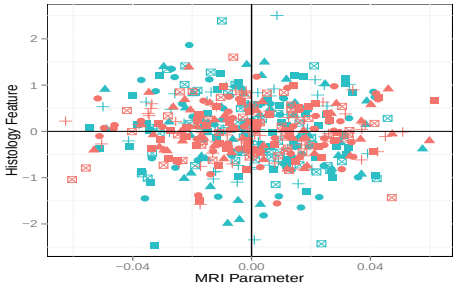


(d) Residuals after subtracting the group means.

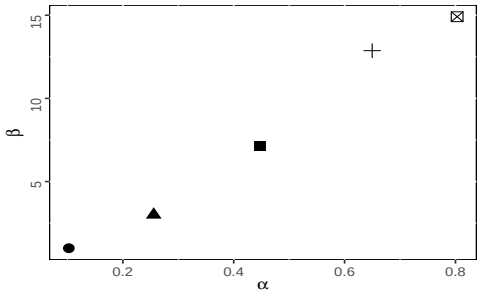
Figure 5.2: Illustrative examples demonstrating the effect of AD progression on an MRI parameter and a histology feature at two time points. Panel (a and b) without and with (c,d) correlation between MRI and histology parameters (arbitrary units). Larger symbols denote the group means. Blue symbols: wildtype mice. Red symbols: transgenic mice.



(a) Histology versus MRI.



(b) Residuals after subtracting the means.

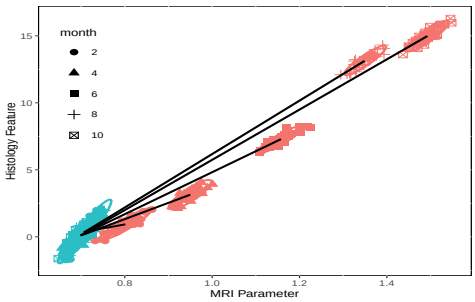


(c) Age dependent AD progression effect.

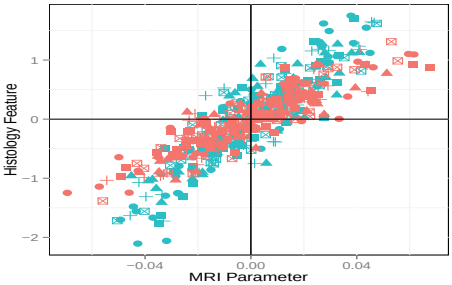
Figure 5.3: Illustration of simulated (arbitrary units) individual and disease-level surrogacy using the AD animal model for a scenario with low individual-level surrogacy. The solid lines in Panel a connects the means of the transgenic and wildtype groups at each age. The slope of these lines is equal to the RE (at each age). Panel c presents the disease effects β on a histology feature versus the disease effects α on an MRI parameter. Blue symbols: wildtype mice. Red symbols: transgenic mice.

individual-level surrogacy is low. In other words, the effect of AD progression on histology features can be predicted using the AD progression effects observed on MRI parameters while at individual level, MRI values are not predictive for histology features. A scenario in which MRI parameters and histology features are associated at both disease and individual levels is shown in Figure 5.4.

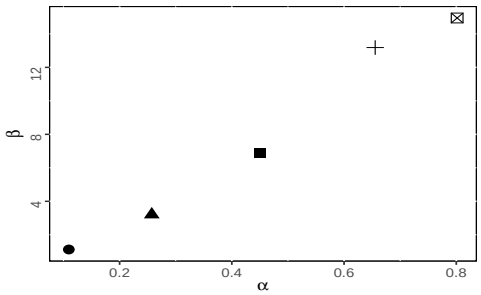
Other settings that may arise are shown in Figure 5.5. Panels on the left (a, c and e) corresponds to data with both low individual and disease-level surrogacy, while Panels b, d and f corresponds to a setting with high individual-level, but low disease-level surrogacy. In what follows, a two-stage modelling approach applied to the MRI data is presented in Chapter 6 while a one-stage model fitted within the hierarchical Bayesian framework is presented in Chapter 7.



(a) Histology versus MRI

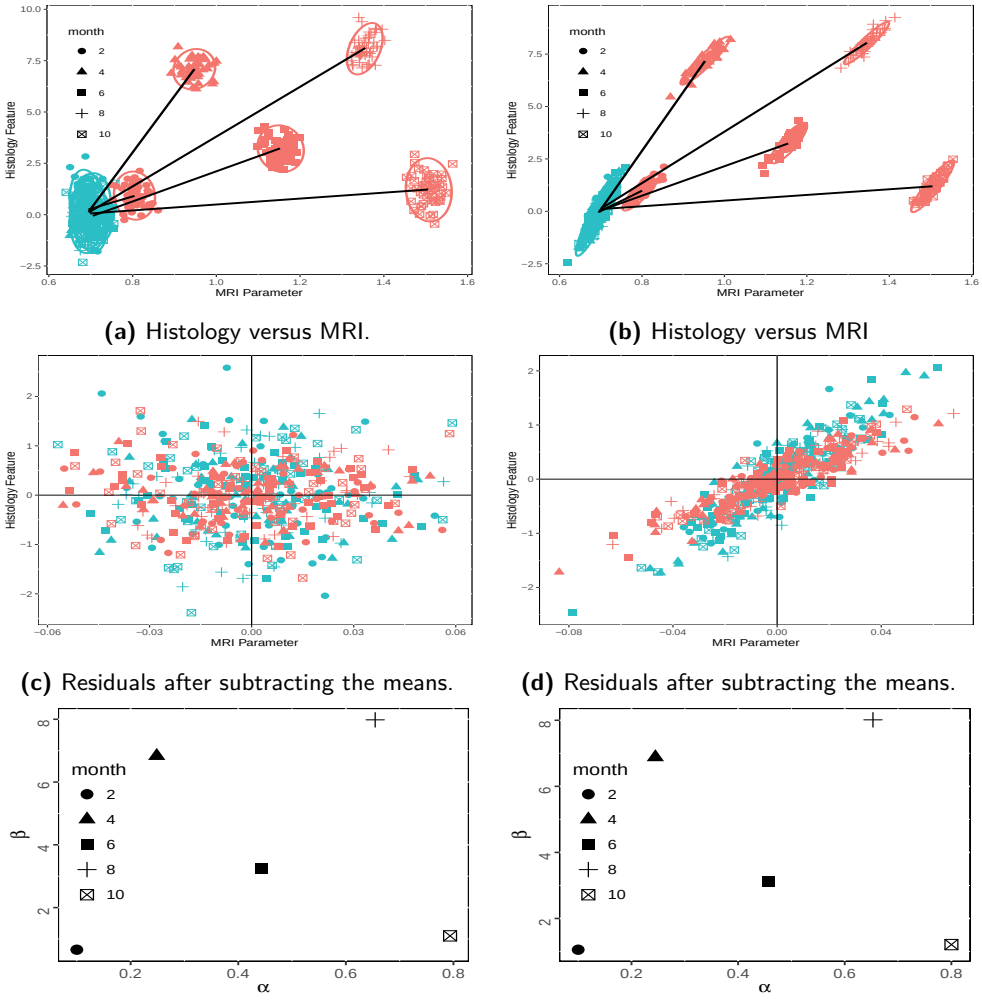


(b) Residuals after subtracting the means.



(c) Age dependent AD progression effect.

Figure 5.4: Illustration of simulated (arbitrary units) individual and disease-level surrogacy using the AD animal model for a scenario with high individual-level surrogacy. The solid lines in panel a connects the means of the transgenic and wildtype groups at each age. The slope of these lines is equal to the RE (at each age). Panel c presents the disease effects β on a histology feature versus the disease effects α on an MRI parameter. Blue symbols: wildtype. Red symbols: transgenic.



(a) Histology versus MRI.

(b) Histology versus MRI

(c) Residuals after subtracting the means.

(d) Residuals after subtracting the means.

(e) Age dependent AD progression effect.

(f) Age dependent AD progression effect.

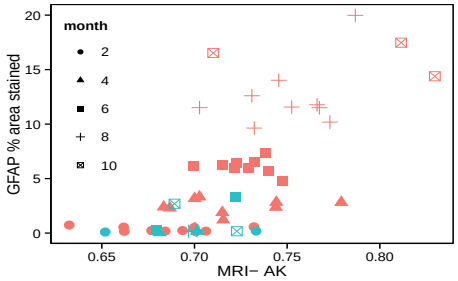
Figure 5.5: Illustration of simulated (arbitrary units) individual and disease-level surrogacy using the AD animal model for a scenario with both low individual-level surrogacy (left panel) and high individual-level surrogacy (right panel). Both settings correspond to data with low disease-level surrogacy. The solid lines in panel a and b connects the means of the transgenic and wildtype groups at each age. The slope of these lines are equal to the RE (at each age). Panel e and f presents the disease effects β on a histology feature versus the disease effects α on an MRI parameter. Blue symbols: wildtype mice. Red symbols: transgenic mice.

Chapter 6

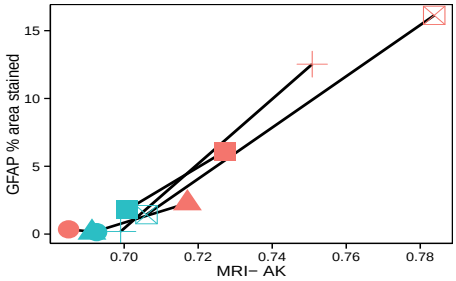
Evaluation of MRI as a Biomarker for Histology in Alzheimer's Disease

6.1 Introduction

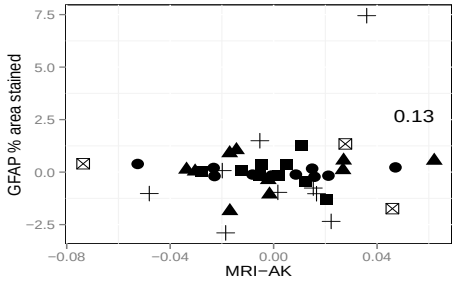
Similar to previous chapters, the methodology discussed in this chapter is first illustrated with the motor cortex data, a particular MRI parameter and histology feature. Subsequently, results of analysing data from all brain regions and MRI/histology parameters combination are presented. Following the procedure for pathological histology acquisition described in Section 2.4, four histology stains (MBP, GFAP, IBA1, and 4G8) were applied and several histology features derived. In this and subsequent chapters, the analysis only focuses on one histology feature (percentage of area stained). Figure 6.1 presents the motor cortex data for GFAP staining and MRI-AK. In Panel a, a clear age-dependent shift of the true and surrogate endpoints for the transgenic mice is evident. Moreover, Panel b shows the age-matched relative effects which also increase with age. Upon adjusting for the disease status, the association between GFAP percentage of area stained and MRI-AK is negligible (Panel c and d). The remainder of the chapter is arranged as follows: Section 6.2 presents the two-stage modelling approach while, Sections 6.3 and 6.4 are devoted to the application of the proposed methodology on the case study. Section 6.5 provides a discussion while software issues are discussed in Sections 6.6.1 and 6.6.2 .



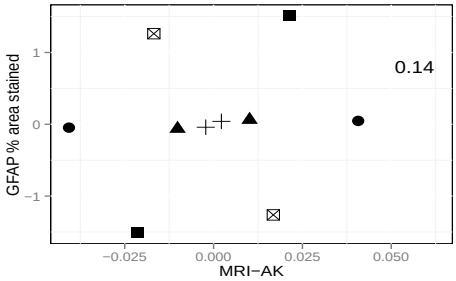
(a) Histology versus MRI.



(b) Age dependent AD progression effect.



(c) Transgenic: Residuals after subtracting the age-matched means.



(d) Wildtype: Residuals after subtracting the age-matched means.

Figure 6.1: The motor cortex data for MRI-AK and GFAP staining. The solid lines in panel b connects the age-matched means of the transgenic and wildtype groups at each age. The slope of these lines are equal to the RE (at each age). Blue symbols: wildtype mice. Red symbols: transgenic mice. DKI units: mm^2/s.

6.2 A Two-Stage Model for Biomarker Evaluation

6.2.1 A Joint Model for MRI and Histology

The analysis presented in this section consists of a region/MRI/histology-specific model. Hence, for each region, 4×7 models are fitted. Each model is used to evaluate one MRI parameter as a biomarker for one histology feature. An extension of the methodology presented in this chapter to evaluating multiple MRI parameters and/or histology features simultaneously can be performed, with appropriate adaptation. The observation unit for the analysis is (X_{ij}, Y_{ij}, Z_i) with X_{ij} being the MRI parameter for the ith animal, $i = 1, \ldots, N_j$ at age j, $j = 1, \ldots, J$, Y_{ij} is the histology feature of the ith animal at age j and Z_i is an indicator variable for the genotype the animal belongs to given by

$$Z_i = \begin{cases} 1, \text{ APP/PS1 Transgenic,} \\ 0, \text{ Wildtype.} \end{cases}$$

We assume that the mean structure for an MRI parameter and a histology feature, respectively, is given by

$$\begin{aligned} E\left(X_{ij}|Z_i\right) &= \mu_{X_j} + \alpha_j Z_i, \\ E\left(Y_{ij}|Z_i\right) &= \mu_{Y_j} + \beta_j Z_i. \end{aligned} \tag{6.1}$$

Here, μ_{X_j} and μ_{Y_j} are the age-specific means of the wildtype mice for the MRI parameter and histology feature, respectively. Note that for the wildtype mice group, we assume that the histology feature is constant over time since the disease pathology does not vary a lot for these young ages (2-10 months) in the wildtype mice. Thus, the mean structure in (6.1) can be simplified by having only one parameter for histology staining in wildtype mice, i.e, $\mu_{Y_j} = \mu_Y$. The age-specific parameters α_j and β_j correspond to the disease effect on MRI and histology at a given age, respectively. Further, we assume that the two endpoints (histology and MRI) follow a bivariate normal distribution with genotype-specific covariance matrices, that is,

$$\begin{pmatrix} X_{ij} \\ Y_{ij} \end{pmatrix} \sim Normal\left(\begin{bmatrix} \mu_{X_j} + \alpha_j Z_i \\ \mu_Y + \beta_j Z_i \end{bmatrix}, \Sigma\right). \tag{6.2}$$

Here, Σ is a 2×2 genotype-specific covariance matrix given for transgenic and wildtype mice, respectively, given by

$$\Sigma_{\mathrm{T}} = \begin{pmatrix} \sigma^2_{\mathrm{A_m}} & \sigma_{\mathrm{A_{hm}}} \\ \sigma_{\mathrm{A}_{hm}} & \sigma^2_{\mathrm{A}_h} \end{pmatrix} \text{ and } \Sigma_{\mathrm{W}} = \begin{pmatrix} \sigma^2_{\mathrm{W}_m} & \sigma_{\mathrm{W_{hm}}} \\ \sigma_{\mathrm{W_{hm}}} & \sigma^2_{\mathrm{W_h}} \end{pmatrix}. \tag{6.3}$$

From (6.3), the notation $\sigma^2_{A_h}$, $\sigma^2_{A_m}$, and $\sigma_{A_{hm}}$ corresponds to the variance of a histology feature, MRI parameter and the covariance between them, respectively, in transgenic mice. Similarly, for wildtype mice, $\sigma^2_{W_h}$, $\sigma^2_{W_m}$, and $\sigma^2_{W_{hm}}$ corresponds to variance of

a specific histology feature, MRI parameter and the covariance between them, respectively.

The joint model specified in (6.1) allows us to model two sources (or aspects) of the association between a specific histology feature and an MRI parameter: (1) the association between the disease evolution effects (with respect to age) of the two endpoints and (2) the association between the two endpoints adjusted for the time evolution of the disease. In what follows, we show that the two sources of association can be interpreted as individual and disease-level surrogacy. The later is similar to the trial-level surrogacy discussed in Burzykowski *et al.* (2005).

6.2.2 Genotype-Specific Individual-Level Surrogacy

Based on the covariance matrices specified in (6.3), we can derive the adjusted correlation between an MRI parameter and a specific histology feature for each genotype given by

$$\rho_{\mathrm{T}} = \frac{\sigma_{A_{hm}}}{\sqrt{\sigma^2_{A_h} \times \sigma^2_{A_m}}}, \text{ and } \rho_{\mathrm{W}} = \frac{\sigma_{W_{hm}}}{\sqrt{\sigma^2_{W_h} \times \sigma^2_{W_m}}}. \tag{6.4}$$

The genotype-specific adjusted correlations ρ_{W} and ρ_{T} measures the association between the two endpoints adjusted for the time evolution of the disease and can be interpreted in the same way as the adjusted association in the surrogacy model presented in Burzykowski *et al.* (2005), Muchene *et al.* (2016b). A large absolute values of the adjusted correlation imply better surrogacy at an individual level. Note that in contrast with the models discussed in Burzykowski *et al.* (2005), we do not assume that the association between MRI and histology is equal in the two groups.

6.2.3 Disease-Level Surrogacy

The joint model specified in (6.1) allows us to estimate the age-and genotype-specific parameters $\alpha_j = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ and $\beta_j = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$. Our aim is to establish a relationship between α_j and β_j and in particular, to assess whether AD evolution observed for the MRI parameter is predictive for the AD evolution observed for a particular histology feature. In other words, we wish to evaluate whether an MRI parameter can be used as a biomarker for a given histology feature in an AD mouse model at a disease level. Disease-level surrogacy can be measured using R^2 obtained from the regression model (6.5), whereby, η and γ are regression coefficients, while ε_j denotes the measurement error for the regression model.

$$\hat{\beta}_j = \eta + \gamma \hat{\alpha}_j + \varepsilon_j, \ j = 1...J. \tag{6.5}$$

Table 6.1: The motor cortex region: Parameter estimate (standard error) of AD progression effects. α: the disease effects on MRI-AK (mm^2/s). β: the disease effect on GFAP percentage of area stained.

Age	$\hat{\beta}$ (SE)	$\hat{\alpha}$ (SE)
2	-0.35 (0.51)	-0.01 (0.02)
4	1.58 (0.51)	0.03 (0.02)
6	5.43 (0.53)	0.03 (0.02)
8	11.83 (0.53)	0.05 (0.02)
10	15.43 (0.92)	0.08 (0.03)

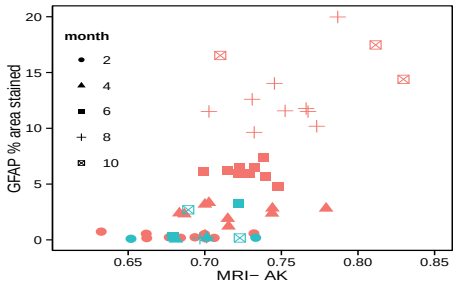
6.3 Application to the MRI Project Data: Examples of Region Specific Models

The joint model specified in (6.1) was applied for each combination of the seven MRI parameters and the four histology stains in each of the 23 ROI (shown in Figure 6.6). In this section, we discuss the results in the motor cortex (Section 6.3.1) and the caudate-putamen (Section 6.3.2) regions. Note that, as explained in Section 6.2, the joint model (6.1) is formulated with a constant effect of histology in the wildtype mice ($\mu_{Y_j} = \mu_Y$).

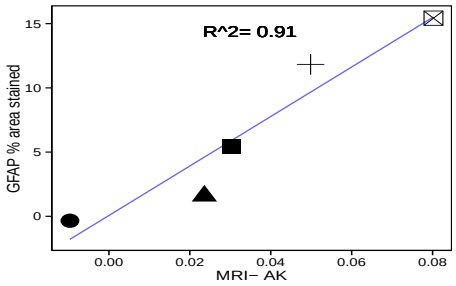
6.3.1 The Motor Cortex: GFAP Staining and MRI-AK

The observed data for GFAP percentage of area stained and MRI-AK are shown in Figure 6.2a where an age-dependent shift of MRI and histology measurements for older transgenic mice is observed. Parameter estimates obtained for the joint model are presented in Table 6.1. The estimated regression model, shown in Figure 6.2b is given by $\hat{\beta}_j = 0.064 + 192.675\hat{\alpha}_j$. The surrogacy measure at disease level $\hat{R}^2_D = 0.91$ indicates that MRI-AK is a good predictive biomarker for GFAP staining. At an individual level, after adjusting for the disease effect, there is low correlation between the residuals ($\hat{\rho_A} = \hat{\rho_W} = 0.13$) indicating a low individual-level surrogacy (Panel c and d).

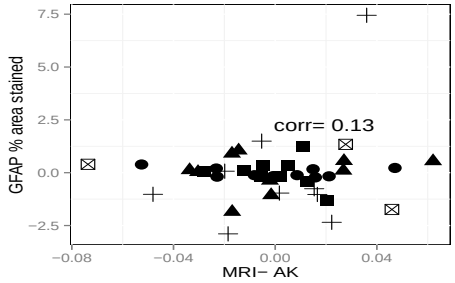
Figure 6.3a presents the disease-level surrogacy measures for all seven MRI parameters and four histology stains in the motor cortex region. Note that the MRI-AK parameter is found to be predictive for 4G8 staining ($R^2_D = 0.94$) and GFAP staining ($R^2_D = 0.91$), while it has relatively low predictive value for the IBA-1 ($R^2_D = 0.60$) and MBP ($R^2_D = 0.47$) staining. MRI-MD was found to be predictive at disease-level for 4G8 ($R^2_D = 0.87$), GFAP ($R^2_D = 0.90$), and IBA-1 ($R^2_D = 0.76$) stainings, respectively. In addition, MRI-RD was predictive for MBP staining with $R^2_D = 0.83$. The results for individual-level surrogacy are shown in Figure 6.3 which reveals that MRI parameters were not predictive for histology at an individual-level in both transgenic (Figure 6.3b) and wildtype mice (Figure 6.3c).



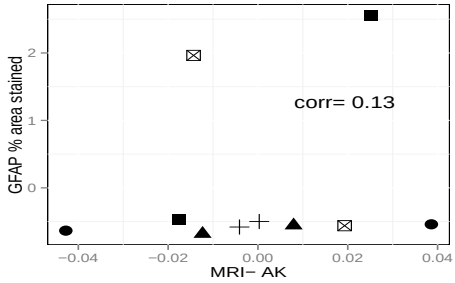
(a) Histology versus MRI.



(b) Age dependent AD progression effect.

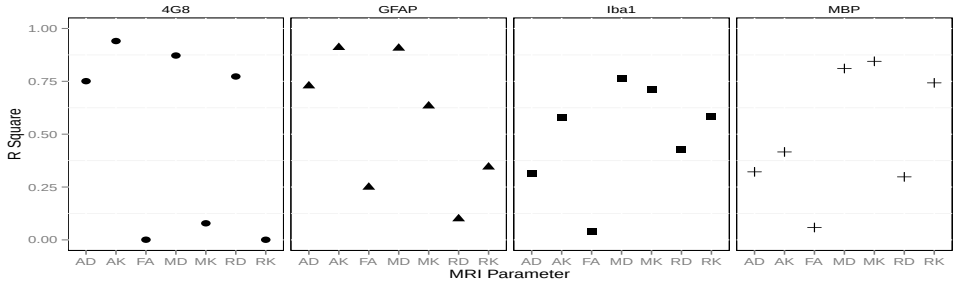


(c) Transgenic: Residuals after subtracting the age-matched group means.

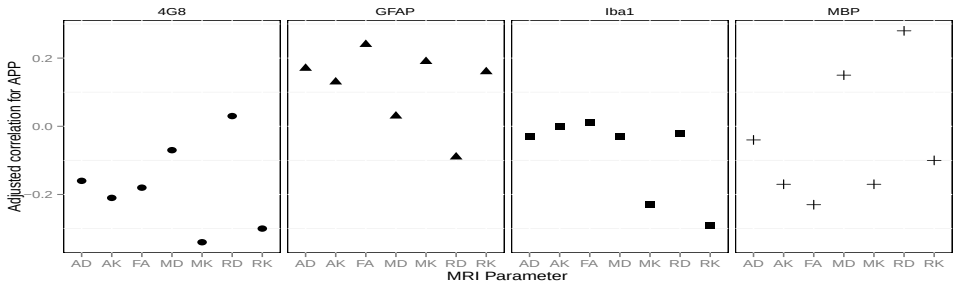


(d) Wildtype: Residuals after subtracting the age-matched group means.

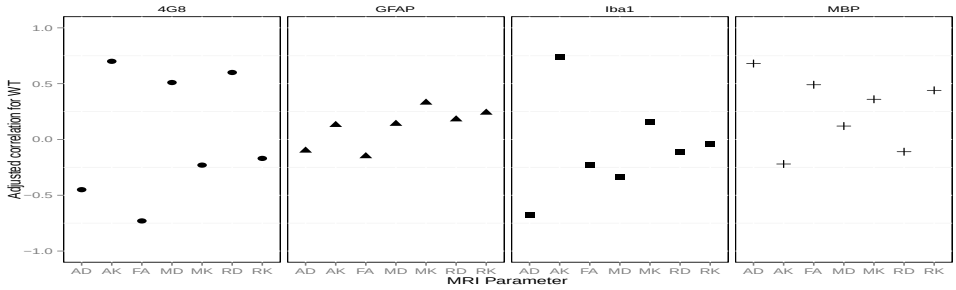
Figure 6.2: The motor cortex region: Evaluation of surrogacy for MRI-AK (mm^2/s) and GFAP percentage of area stained. Blue symbols: wildtype mice. Red symbols: transgenic mice.



(a) Disease-level surrogacy for all MRI parameters and histology staining.



(b) Transgenic mice: Individual-level surrogacy for all MRI parameters and histology staining.



(c) Wildtype mice: Individual-level surrogacy for all MRI parameters and histology staining.

Figure 6.3: The motor cortex region: Individual and disease-level surrogacy for all MRI parameters (mm^2/s) and histology stains (percentage of area stained).

Table 6.2: The caudate-putamen region: Parameter estimate of AD progression effects for GFAP percentage of area stained with MRI-AK. α: disease effect on MRI-AK (mm^2/s). β: disease effect on GFAP percentage of area stained.

Age	$\hat{\beta}$ (SE)	$\hat{\alpha}$ (SE)
2	0.35 (0.43)	-0.0001 (0.02)
4	1.05 (0.43)	0.02 (0.02)
6	3.99 (0.46)	-0.013 (0.024)
8	8.75 (0.46)	0.0001 (0.024)
10	16.37 (0.80)	0.09 (0.03)

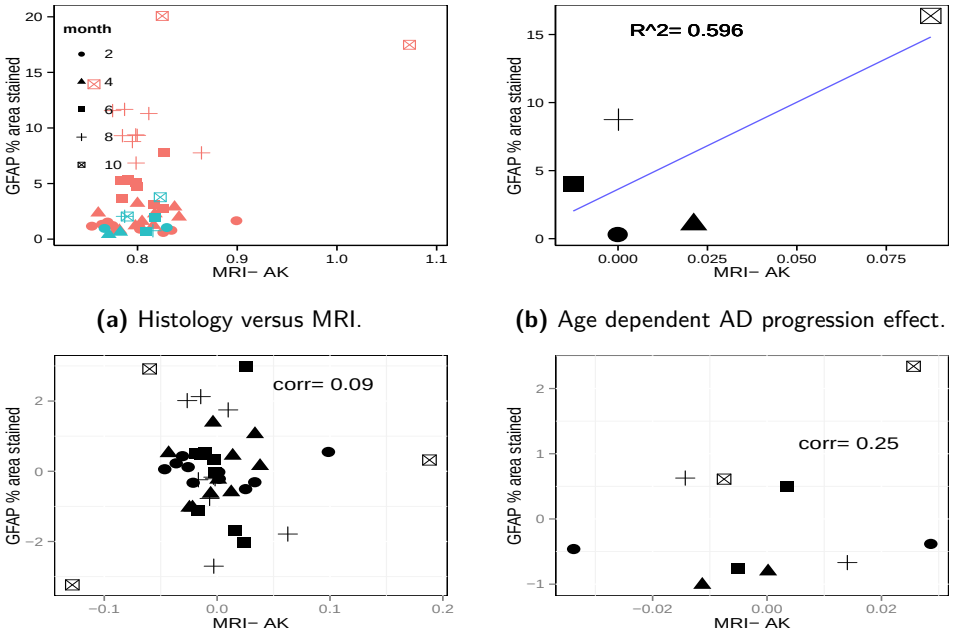
6.3.2 The Caudate-Putamen: GFAP Staining and MRI-AK

A similar analysis was conducted for the caudate-putamen region. The observed data for GFAP percentage of area stained and MRI-AK shown in Figure 6.4a indicates an age-dependent shift in histology values for transgenic mice and a relatively small shift in MRI-AK values. The estimated disease-level surrogacy is relatively low at $R_D^2 = 0.596$ (See Figure 6.4b). Moreover, although a large disease effect on histology is observed over time as shown in Table 6.2, the disease effect on MRI is relatively small, hence the low association between MRI and histology disease effects. Similar patterns can be observed for individual-level surrogacy of MRI-AK as a biomarker for GFAP percentage of area stained as shown in Figure 6.4c and 6.4d.

Figure 6.5a shows the surrogacy measures for all MRI parameters and histology stains (percentage of area stained) in the caudate-putamen region. For the 4G8 staining, the highest surrogacy measures at a disease level was found for MRI-AD and MRI-MD ($R_D^2 = 0.62$ and $R_D^2 = 0.61$, respectively). The MRI parameters with the highest disease-level surrogacy measures in the caudate-putamen region were MRI-MD, MRI-MK and MRI-RD for the GFAP staining with $R_D^2 = 0.83$, $R_D^2 = 0.87$, and $R_D^2 = 0.86$, respectively. Note that for the MBP staining, all MRI parameters except for MRI-AD ($R_D^2 = 0.35$), are good biomarkers at disease-level with $R_D^2 \geq 0.72$. For transgenic mice, individual-level surrogacy was very low for all MRI parameters and histology stains as shown in Figure 6.5b. On the other hand, individual-level surrogacy was found to be relatively high in the wildtype mice for MBP staining using MRI-AK ($\hat{\rho}_W = 0.95$), MRI-MK ($\hat{\rho}_W = 0.94$), MRI-RD ($\hat{\rho}_W = 0.94$) and MRI-RK ($\hat{\rho}_W = 0.71$).

6.4 The Surrogacy Map of the Brain

The joint model allows us to evaluate the surrogacy pattern in the brain for all combinations of MRI parameters and histology stains. Figure 6.6 shows a heatmap of the disease-level surrogacy in the 23 ROI. Clearly, surrogacy is highly dependent on the region, MRI parameter and histology stain. For regions such as amygdala and olfactory bulb, none of the MRI parameters is useful as a biomarker for any of the histology stains. GFAP percentage of area stained can be predicted by several MRI parameters in the caudate-putamen, cerebellum, and several cortex regions. For IBA-1 staining,



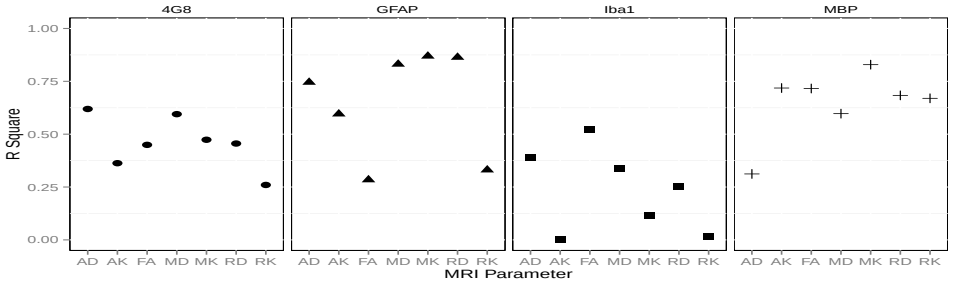
(a) Histology versus MRI.

(b) Age dependent AD progression effect.

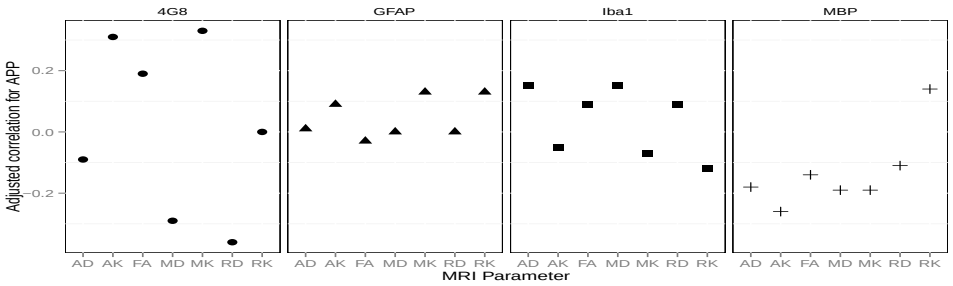
(c) Transgenic: Residuals after subtracting the mean.

(d) Wildtype: Residuals after subtracting the mean.

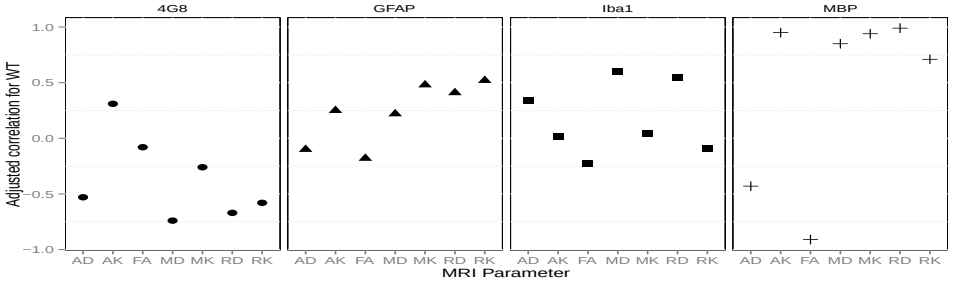
Figure 6.4: The caudate-putamen region: Evaluation of surrogacy with MRI-AK (mm^2/s) and GFAP percentage of area stained. Blue symbols: wildtype mice. Red symbols: transgenic mice.



(a) Disease level surrogacy.



(b) Individual-level surrogacy in transgenic mice.



(c) Individual-level surrogacy in wildtype mice.

Figure 6.5: The caudate-putamen region: Individual and disease-level surrogacy map for all MRI parameters (mm^2/s) and histology stains (percentage of area stained).

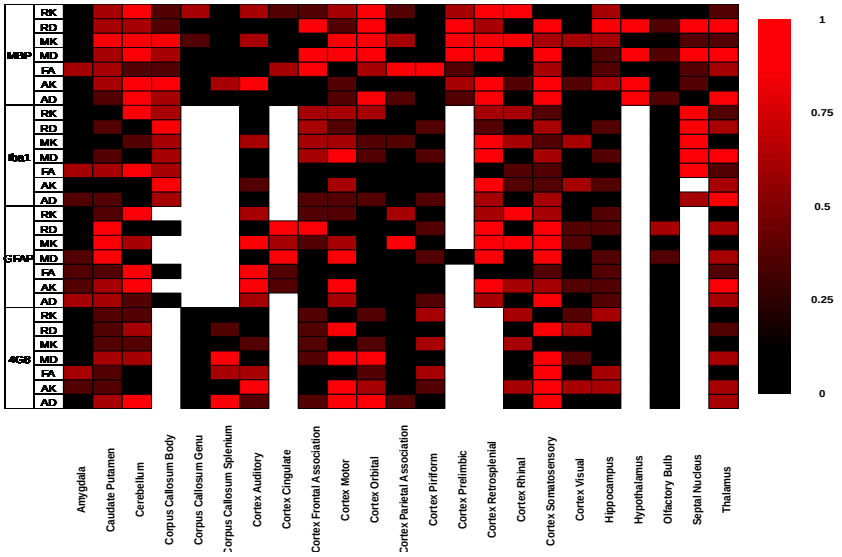


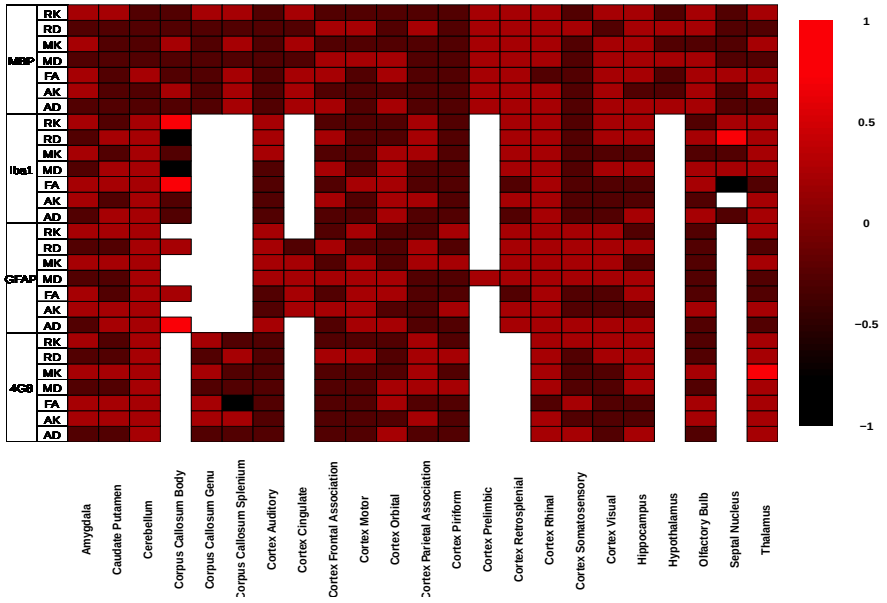
Figure 6.6: Disease level surrogacy in 23 regions of the brain. R_D^2 for each MRI parameter (units: mm^2/s) and histology (percentage of area stained). White fill: surrogacy measure not computed due to inadequate data for estimation.

good disease-level surrogacy was observed in the septal nucleus region using all MRI parameters (apart from MRI-AD and MRI-AK). Relatively high level of surrogacy of MRI parameters with 4G8 staining was observed in the cortex regions. Thus, since these four histology stains evaluate different aspects of the disease morphology, there is need to evaluate surrogacy at the different brain regions using MRI parameters with high surrogacy level for the particular histology stains.

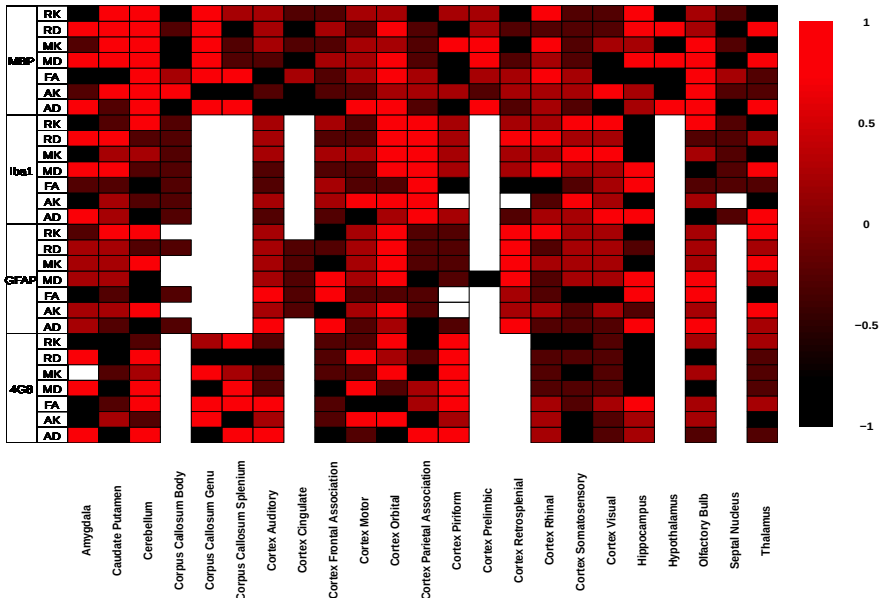
A surrogacy map for individual-level surrogacy in transgenic mice is presented in Figure 6.7a. Overall, a low individual-level surrogacy is observed, an indication that prediction of histology at an individual level is not practical using MRI parameters. For the wildtype mice however, individual-level surrogacy was relatively high in some brain regions, MRI parameters and histology stains (see Figure 6.7b). Note that this individual-level association between wildtype mice is less of interest since from a biological point of view, the endpoints in wildtype mice are expected to be "less" correlated since no disease-induced pathology is expected in wildtype mice.

6.5 Discussion

The joint model specified in Section 6.2 was developed in order to model the association between MRI and histology, taking into account the disease progression effects on both endpoints. The observation unit that we have used in this chapter is the triplet $(\text{Genotype}_i, \text{MRI}_{ij}, \text{Histology}_{ij})$. Figure 6.8 illustrates the two sources of association

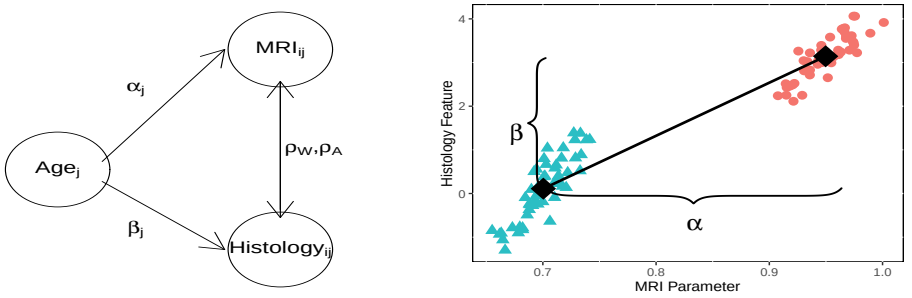


(a) Transgenic mice.



(b) Wildtype mice.

Figure 6.7: Individual level surrogacy in 23 regions of the brain. $\hat{\rho}_T$ and $\hat{\rho}_W$, respectively, for each MRI parameter (mm^2/s) and histology (percentage of area stained). White fill: surrogacy measure not computed.



(a) Association between the genotype and endpoints.

(b) Histology versus MRI at age j (arbitrary units).

Figure 6.8: Illustration of the joint modelling framework: The association between MRI and histology after adjusting for the disease effects.

presented in this chapter. For a given age, the effect of the disease on MRI α_j and the effect of the disease on histology β_j is represented by the shift in the distribution of both MRI and histology parameters as illustrated in panel 6.8b. Panel 6.8a illustrates the genotype-specific association in the residuals after adjusting for the disease effects α_j and β_j.

We have shown that, using a two stage approach, we can estimate a genotype-specific adjusted association $\hat{\rho}_W$ and $\hat{\rho}_A$ using the joint model (6.1) in the first stage, while the prediction of the disease progression effects on histology can be done in the second stage using linear regression model for $\hat{\beta}_j$ and $\hat{\alpha}_j$. Although, the experimental setting discussed in this chapter is completely different from the one encountered in clinical trials, the same association structure (as illustrate in Figure 6.8a) implies that the same modelling approach can be used in order to evaluate the quality of MRI as a biomarker for histology. We have shown that the use of MRI as a biomarker for histology depends on the brain region, MRI parameters and histology staining.

The joint model presented in this chapter assumes that the disease effect on MRI, α_j, at a particular age j can be used to predict the disease effect on histology, β_j at the same age. That is, there is a relationship between the disease effects on MRI and histology given by

$$\beta_j = f(\alpha_j). \tag{6.6}$$

A model whereby the disease-effect on histology β_j at a given age j can be predicted by the disease-effect on MRI at an earlier timepoint $t < j$, would be of much interest, from a diagnostic point of view. For such a model, the relationship is given by

$$\beta_j = f(\alpha_t), \quad t < j. \tag{6.7}$$

The case studies presented in this chapter posed two challenges with regards to sample size: (1) there were only five age groups, which implies that, estimation of the

linear regression line in the second stage is based on only five observations and (2) there were only two wildtype mice at each age group. Therefore, the genotype-specific coefficients μ_{Y_j} in (6.2) are based on two observations, hence they may have higher variability. Note that for the wildtype mice, we do not expect a correlation between the two endpoints and therefore, a model assuming the independence of the MRI parameter and histology feature in wildtype mice can be considered.

6.6 Appendix: Software

6.6.1 Implementation in SAS

In this section, we discuss the implementation of the joint model discussed in Section 6.2 in SAS using data from the motor cortex with a pair of MRI-AK (biomarker) and GFAP percentage of area stained (true endpoint).

Data Structure

The joint model, discussed in Section 6.2, was fitted using `PROCEDURE MIXED` in SAS 9.4. For each subject, measurements for both MRI and histology were available. Hence, data for MRI and histology parameters for a single subject appeared in subsequent rows. A partial print of the data is given in Panel 6.1.

Panel 6.1: Printout of a section of the SAS dataset.

```
PROC PRINT data=MriHistData;run;
animalid age     genotype      response    endpoint
1           2       TRANSGENIC 0.001128658     MRI
1           2       TRANSGENIC 0.126652588     Histology
2           2       WILDTYPE   0.189814745     Histology
2           2       WILDTYPE   0.001124777     MRI
```

Common Parameter for Histology in the Wildtype Group

As mentioned in Section 6.2, from a biological point of view, it is assumed that histology values of wildtype mice should remain constant between the age of 2-10 months, since there is no significant disease pathology (due to ageing) progression. Hence, in the model for histology, wildtype mice have a single parameter which does not change with age. In SAS, this can be achieved by defining a common parameter `CommonInt` for histology in wildtype as shown in Panel 6.2. Parameter estimate for the disease progression effects (in both endpoints) are shown in Figure 6.9.

Panel 6.2: SAS code for defining a common intercept the model.

```
DATA MriHistData;
SET MriHistData;
commonInt=age;
if treatment='WILDTYPE' and endpoint='Histology' then
    commonInt=0;
RUN;
```

The association between MRI and histology is modelled using the `REPEATED` statement. The option `GROUP=genotype` allows for genotype-specific covariance matrices (6.3). The estimated disease effects on both MRI and histology are output by passing the `SOLUTION` option and stored in a dataset named `fixedeffects` using the `SOLUTIONF` option in the `ODS OUTPUT` command. The complete SAS code used to fit the joint model is shown in Panel 6.3.

Panel 6.3: SAS code for implementing the joint model for surrogacy.

```
PROC MIXED DATA=surrogate;
CLASS genotype animalid endpoint commonint(ref='0');
MODEL response=commonint*endpoint
    commonint*endpoint*genotype/SOLUTION NOINT;
REPEATED endpoint/SUBJECT= animalid*commonint TYPE=UN GROUP=genotype
    R=1,4 RCORR=1,4;
ODS OUTPUT SOLUTIONF=fixedeffects;
RUN;QUIT;
```

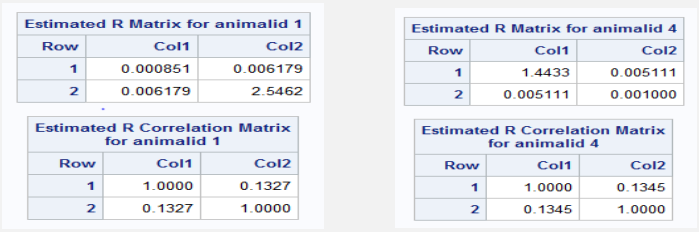
Figure 6.9: SAS fixed effects parameter estimates.

Solution for Fixed Effects

Effect	treatment	endpoint	int	Estimate	Standard Error	DF	t Value	Pr > |t|
int			2	0.6946	0.02220	4	31.29	<.0001
int			4	0.6935	0.02220	4	31.24	<.0001
int			6	0.6969	0.02220	4	31.39	<.0001
int			8	0.7010	0.02220	4	31.57	<.0001
int			10	0.7036	0.02220	4	31.69	<.0001
int			0	0.7304	0.3799	4	1.92	0.1269
treatme*endpoint*int	TRANSGENIC	surrogate	2	-0.00964	0.02404	31	-0.40	0.6910
treatme*endpoint*int	TRANSGENIC	surrogate	4	0.02359	0.02404	31	0.98	0.3341
treatme*endpoint*int	TRANSGENIC	surrogate	6	0.03035	0.02424	31	1.25	0.2198
treatme*endpoint*int	TRANSGENIC	surrogate	8	0.04985	0.02424	31	2.06	0.0482
treatme*endpoint*int	TRANSGENIC	surrogate	10	0.08022	0.02787	31	2.88	0.0072
treatme*endpoint*int	TRANSGENIC	true	2	-0.3455	0.5051	31	-0.68	0.4990
treatme*endpoint*int	TRANSGENIC	true	4	1.5811	0.5051	31	3.13	0.0038
treatme*endpoint*int	TRANSGENIC	true	6	5.4260	0.5324	31	10.19	<.0001
treatme*endpoint*int	TRANSGENIC	true	8	11.8291	0.5324	31	22.22	<.0001
treatme*endpoint*int	TRANSGENIC	true	10	15.4263	0.9215	31	16.74	<.0001
treatme*endpoint*int	WILDTYPE	surrogate	2	0	.	.	.	.
treatme*endpoint*int	WILDTYPE	surrogate	4	0	.	.	.	.
treatme*endpoint*int	WILDTYPE	surrogate	6	0	.	.	.	.
treatme*endpoint*int	WILDTYPE	surrogate	8	0	.	.	.	.
treatme*endpoint*int	WILDTYPE	surrogate	10	0	.	.	.	.
treatme*endpoint*int	WILDTYPE	true	0	0	.	.	.	.

The individual-level surrogacy for wildtype and transgenic mice is computed using the estimated covariance matrices shown in Figure 6.10.

Figure 6.10: SAS joint model covariance and correlation matrices estimates.



Estimated R Matrix for animalid 1

Row	Col1	Col2
1	0.000851	0.006179
2	0.006179	2.5462

Estimated R Correlation Matrix for animalid 1

Row	Col1	Col2
1	1.0000	0.1327
2	0.1327	1.0000

Estimated R Matrix for animalid 4

Row	Col1	Col2
1	1.4433	0.005111
2	0.005111	0.001000

Estimated R Correlation Matrix for animalid 4

Row	Col1	Col2
1	1.0000	0.1345
2	0.1345	1.0000

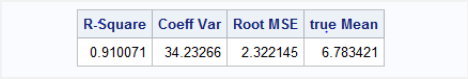
In order to compute disease-level surrogacy, a linear regression model is fitted using `PROC GLM` in SAS (Panel 6.4).

Panel 6.4: SAS code: Regression model for disease-level surrogacy.

```
PROC GLM DATA=fixedeffectsProcessed;
MODEL effectstrue=effectssurrogate;
RUN;QUIT;
```

In the above code, `effectstrue` corresponds to the estimated disease effects on the true endpoint $\hat{\boldsymbol{\beta}}$, while `effectssurrogate` corresponds to the estimated disease effects on the surrogate endpoint $\hat{\boldsymbol{\alpha}}$. The reported measure for disease-level surrogacy corresponds to R^2 obtained from the linear regression model is shown in Figure 6.11.

Figure 6.11: Disease level surrogacy: GLM regression model output.



R-Square	Coeff Var	Root MSE	true Mean
0.910071	34.23266	2.322145	6.783421

Age-Specific Parameters for Histology in the Wildtype Model

Rather than assume a common histology parameter in wildtype mice, an age-specific parameter can be specified by substituting `commonint` with `age`. In order to obtain age and genotype-specific estimates for the disease-effect on both MRI and histology, we include in the model an interaction term `age*endpoint*genotype` (See Panel 6.5).

Panel 6.5: SAS code for an age-specific joint model.

```
PROC MIXED DATA=MriHistData;
CLASS genotype animalid endpoint age;
MODEL response=age*endpoint age*endpoint*genotype/SOLUTION NOINT;
REPEATED endpoint/SUBJECT= animalid*age TYPE=UN GROUP=genotype
    R=1,4 RCORR=1,4;
ODS OUTPUT SOLUTIONF=fixedeffects;
RUN;QUIT;
```

6.6.2 Implementation in R

The proposed model can easily be implemented in R using the `gls` function from the `nlme` package.

Common Parameter for Histology in the Wildtype Group

We adopt a dummy coding for the variables of interest as shown in the partial print of the `MriHistData.dummy` data object in Figure 6.12.

Figure 6.12: Partial print of the R dataset with appropriately defined dummy variables.

```
> head(MriHistData.dummy, 10)
   animalId age   genotype  endpoint   response commonint commonint0 commonint2 commonint4 commonint6 commonint8 commonint10 commonint2.endpointtrue.genotypeTRANSGENIC
1        37   2 TRANSGENIC      true 0.57389482         2          0          1          0          0          0          0                                          1
52       37   2 TRANSGENIC surrogate 0.73209081         2          0          1          0          0          0          0                                          0
42       39   2   WILDTYPE      true 0.18981475         0          1          0          0          0          0          0                                          0
93       39   2   WILDTYPE surrogate 0.73324633         2          0          1          0          0          0          0                                          0
44        6   4   WILDTYPE      true 0.04817485         0          1          0          0          0          0          0                                          0
95        6   4   WILDTYPE surrogate 0.68116338         4          0          0          1          0          0          0                                          0
18       49   4 TRANSGENIC      true 2.39414174         4          0          0          1          0          0          0                                          0
69       49   4 TRANSGENIC surrogate 0.68341699         4          0          0          1          0          0          0                                          0
46       15   6   WILDTYPE      true 3.28905565         0          1          0          0          0          0          0                                          0
97       15   6   WILDTYPE surrogate 0.72204392         6          0          0          0          1          0          0                                          0
   commonint4.endpointtrue.genotypeTRANSGENIC commonint6.endpointtrue.genotypeTRANSGENIC commonint8.endpointtrue.genotypeTRANSGENIC commonint10.endpointtrue.genotypeTRANSGENIC
1                                           0                                           0                                            0                                             0
52                                          0                                           0                                            0                                             0
42                                          0                                           0                                            0                                             0
93                                          0                                           0                                            0                                             0
44                                          0                                           0                                            0                                             0
95                                          0                                           0                                            0                                             0
18                                          1                                           0                                            0                                             0
69                                          0                                           0                                            0                                             0
46                                          0                                           0                                            0                                             0
97                                          0                                           0                                            0                                             0
   commonint2.endpointsurrogate.genotypeTRANSGENIC commonint4.endpointsurrogate.genotypeTRANSGENIC commonint6.endpointsurrogate.genotypeTRANSGENIC
1                                                0                                                0                                                0
52                                               1                                                0                                                0
42                                               0                                                0                                                0
93                                               0                                                0                                                0
44                                               0                                                0                                                0
95                                               0                                                0                                                0
18                                               0                                                0                                                0
69                                               0                                                1                                                0
46                                               0                                                0                                                0
97                                               0                                                0                                                0
   commonint8.endpointsurrogate.genotypeTRANSGENIC commonint10.endpointsurrogate.genotypeTRANSGENIC
1                                                0                                                 0
52                                               0                                                 0
42                                               0                                                 0
93                                               0                                                 0
44                                               0                                                 0
95                                               0                                                 0
18                                               0                                                 0
69                                               0                                                 0
46                                               0                                                 0
97                                               0                                                 0
> |
```

The model can be fitted using the R code displayed in Panel 6.6.

Panel 6.6: R code for implementing the joint model.

```
library(nlme)
fit <- gls(response~-1+endpoint+mu_SURRO_wt4+mu_SURRO_wt6+
     mu_SURRO_wt8+mu_SURRO_wt10 +beta_TRANS_TRUE_2
     +alpha_TRANS_SURRO_4+alpha_TRANS_SURRO_6+alpha_TRANS_SURRO_8
     +alpha_TRANS_SURRO_10+beta_TRANS_TRUE_2
     +beta_TRANS_TRUE_4 +beta_TRANS_TRUE_6
     +beta_TRANS_TRUE_8 +beta_TRANS_TRUE_10,
           data=MriHistData.dummy,
           correlation=corSymm(form = ~ 1| animalid ),
           weight=varIdent(form=~1|endpoint*genotype))
```

By specifying `endpoint` in the right hand side of the `formula`, we allow for a common parameter estimate for histology (true endpoint) in wildtype as well as a parameter for MRI (surrogate endpoint) at 2 months for wildtype mice. The variables `alpha_TRANS_SURRO_2`- `alpha_TRANS_SURRO_10` corresponds to α_1-α_5 while `beta_TRANS_TRUE_2`- `beta_TRANS_TRUE_10` correspond to β_1-β_5 in (6.1).

The argument `correlation=...` allows for the specification of correlated outcomes within a subject. Further, we specify an unstructured correlation using the `corSym` construct. In order to define heterogeneous variances, that is, endpoint and genotype-specific variance covariance matrices as defined in (6.3), the argument `weight=varIdent(...)` is used. The output for the disease progression parameters is shown in Figure 6.13 below.

Figure 6.13: R `gls` output for the surrogacy model.

```
> round(summary(fit)$tTable,3)
                    Value Std.Error t-value p-value
endpointtrue        0.730     0.380   1.923   0.058
endpointsurrogate   0.695     0.022  31.281   0.000
mu_SURRO_wt4       -0.001     0.031  -0.036   0.971
mu_SURRO_wt6        0.002     0.031   0.076   0.939
mu_SURRO_wt8        0.006     0.031   0.202   0.840
mu_SURRO_wt10       0.009     0.031   0.286   0.775
alpha_TRANS_TRUE_2 -0.381     0.632  -0.604   0.548
alpha_TRANS_TRUE_4  1.545     0.632   2.444   0.017
alpha_TRANS_TRUE_6  5.390     0.654   8.237   0.000
alpha_TRANS_TRUE_8 11.793     0.654  18.022   0.000
alpha_TRANS_TRUE_10 15.390    0.997  15.436   0.000
beta_TRANS_SURRO_2 -0.010     0.024  -0.400   0.690
beta_TRANS_SURRO_4  0.024     0.024   0.982   0.329
beta_TRANS_SURRO_6  0.030     0.024   1.250   0.215
beta_TRANS_SURRO_8  0.050     0.024   2.057   0.043
beta_TRANS_SURRO_10 0.080     0.028   2.877   0.005
>
```

The estimated covariance matrices (6.3) with the corresponding correlation estimates for disease-level surrogacy (6.4) are shown in Panel 6.7 below.

Panel 6.7: R output for the joint model covariance matrices.

```
Transgenic.covmat <- getVarCov(fit, individual=1)
cov2cor(Transgenic.covmat )
Marginal variance covariance matrix
       [,1]    [,2]
[1,] 1.00000 0.13295
[2,] 0.13295 1.00000

Wildtype.covmat <- getVarCov(fit, individual=4)
cov2cor(Wildtype.covmat)
Marginal variance covariance matrix
       [,1]    [,2]
[1,] 1.00000 0.13355
[2,] 0.13355 1.00000
```

To obtain the disease-level surrogacy, a linear regression is fitted to the disease progression effects and the model R^2 obtained as shown in Panel 6.8.

Panel 6.8: Disease-level surrogacy measure from the R model fit.

```
summary(lm(alpha~beta, data=fixedEffectsProcessed))$r.squared
[1] 0.9101327
```

Age-Specific Parameters for Histology in the Wildtype Group

A partial print of the dataset for the model with age-specific parameters for histology in the wildtype group is shown in Panel 6.9.

The model is fitted using age and `endpoint` as factor variables. The interaction term `endpoint:age` denotes the age-specific average readout for MRI and histology in wildtype mice, while the three-way interaction term `endpoint:genotype:age` denotes the age-specific disease effect (transgenic compared to wildtype mice). The appropriate R code for model fit is shown in Panel 6.10. The resulting parameter estimates are shown in Figure 6.14.

Panel 6.9: Printout of a section of the dataset used in R model fit.

```
 > head(MriHistData)
   animalid age  genotype endpoint   response
30        1  10 TRANSGENIC    true  17.4715576
81        1  10 TRANSGENIC surrogate  0.8114934
31        2  10 TRANSGENIC    true  16.5277908
82        2  10 TRANSGENIC surrogate  0.7101217
32        3  10 TRANSGENIC    true  14.3902328
83        3  10 TRANSGENIC surrogate  0.8296881
```

Panel 6.10: R code for fitting an age-specific joint model.

```
fit2 <- gls(response~endpoint:age+endpoint:genotype:age -1,
    data=MriHistData,
        correlation=corSymm( form = ~ 1| animalid ),
        weight=varIdent(form=~1|endpoint*genotype))
```

The resulting parameter estimates are shown in Figure 6.14.

Figure 6.14: R gls output for the age-specific surrogacy model.

```
> round(summary(fit2)$tTable,3)
                                                Value Std.Error t-value p-value
endpointtrue:month2                             0.143     0.882   0.162   0.872
endpointsurrogate:month2                        0.693     0.022  30.955   0.000
endpointtrue:month4                             0.111     0.882   0.126   0.900
endpointsurrogate:month4                        0.691     0.022  30.900   0.000
endpointtrue:month6                             1.776     0.882   2.015   0.047
endpointsurrogate:month6                        0.701     0.022  31.318   0.000
endpointtrue:month8                             0.190     0.882   0.216   0.829
endpointsurrogate:month8                        0.699     0.022  31.246   0.000
endpointtrue:month10                            1.432     0.882   1.624   0.108
endpointsurrogate:month10                       0.706     0.022  31.558   0.000
endpointtrue:month2:genotypeTRANSGENIC          0.206     1.016   0.203   0.840
endpointsurrogate:month2:genotypeTRANSGENIC    -0.008     0.024  -0.313   0.755
endpointtrue:month4:genotypeTRANSGENIC          2.164     1.016   2.130   0.036
endpointsurrogate:month4:genotypeTRANSGENIC     0.026     0.024   1.065   0.290
endpointtrue:month6:genotypeTRANSGENIC          4.347     1.030   4.222   0.000
endpointsurrogate:month6:genotypeTRANSGENIC     0.027     0.024   1.092   0.278
endpointtrue:month8:genotypeTRANSGENIC         12.340     1.030  11.984   0.000
endpointsurrogate:month8:genotypeTRANSGENIC     0.052     0.024   2.122   0.037
endpointtrue:month10:genotypeTRANSGENIC        14.698     1.275  11.527   0.000
endpointsurrogate:month10:genotypeTRANSGENIC    0.078     0.028   2.776   0.007
>
```

Note that in this case, the coefficients for histology in wildtype are more imprecise (since at each age, only two observations are available). The estimation of disease-level surrogacy follows as in the previous case of the model with a common histology parameter in wildtype mice.

Chapter 7

Hierarchical Bayesian Model for the Evaluation of MRI as a Biomarker for Histology in Alzheimer's Disease

7.1 Introduction

In Chapter 6, we presented an adaptation of the meta-analytic approach for the evaluation of surrogacy of MRI parameters for particular pathological histology stains in Alzheimer's disease. A two-stage approach was proposed, whereby a joint model for a given MRI parameter and histology stain's feature was fitted in order to estimate the disease effects on both outcomes. In the second stage, a linear regression model was fitted and the disease-level surrogacy quantified by the model R^2. In this chapter, rather than the two-stage approach, we propose to combine the computation of both disease and individual-level surrogacy into one stage. To achieve this, a hierarchical Bayesian joint model is proposed whereby, the disease-level surrogacy evaluation is performed by specifying a joint prior distribution for the disease effects. The methodology, discussed in details in Section 7.2, is based on the hierarchical Bayesian model proposed by Shkedy and Barbosa (2005), with an adaptation for the AD surrogacy settings. Results of applying the methodology to selected parameters from the MRI study are presented in Section 7.3. In Section 7.4, a simulation study is performed to assess the performance of the methodology under different settings.

7.2 Methodology

Consider a histology feature (i.e. a "true" endpoint using surrogacy terminology) Y_{ij} and an MRI parameter (a biomarker) X_{ij} measured for subject $i = 1 \ldots N$ at time point $j = 1 \ldots 5$ (corresponding to $2, 4, 6, 8$ and 10 months of age). Further, let Z_i be an indicator variable for the mice genotype, where $Z_i = 0$ corresponds to a healthy control (i.e, wildtype mice, denoted by W) and $Z_i = 1$ corresponds to APP/PS1 mice (transgenic mice denoted by A). Buyse *et al.* (2000) proposed a two-stage model for the validation of surrogate endpoints in the meta-analytic framework. In the first stage, the linear predictor for the MRI and histology endpoints, respectively, is defined as

$$\begin{array}{l} E(X_{ij}|Z_i) = \mu_{X_j} + \alpha_j Z_i, \\ E(Y_{ij}|Z_i) = \mu_{Y_j} + \beta_j Z_i. \end{array} \tag{7.1}$$

Here, μ_{X_j} and μ_{Y_j} are the age-specific intercepts (means) in wildtype mice for the biomarker (MRI parameter) and "true" (histology feature) endpoints, respectively, while α_j and β_j are the age-specific effects of the genotype (transgenic) on the MRI parameter and histology endpoints, respectively. The association between the two endpoints is captured by specifying a bivariate normal distribution,

$$\left(\begin{array}{c} X_{ij} \\ Y_{ij} \end{array} \right) \sim Normal \left(\left[\begin{array}{c} \mu_{X_j} + \alpha_j Z_i \\ \mu_{Y_j} + \beta_j Z_i \end{array} \right], \Sigma_k \right). \tag{7.2}$$

The specification of the genotype specific covariance matrix $\Sigma_k,\ k = W, A$, allows for the association between the true and surrogate endpoints to be different in wildtype and transgenic mice. For the analysis presented in this chapter, Σ_k is a 2×2 covariance matrix given by,

$$\Sigma_{\mathrm{W}} = \left(\begin{array}{cc} \sigma^2_{\mathrm{W_X}} & \sigma_{\mathrm{W_{XY}}} \\ \sigma_{\mathrm{W}_{XY}} & \sigma^2_{\mathrm{W_Y}} \end{array} \right) \text{ and } \Sigma_{\mathrm{A}} = \left(\begin{array}{cc} \sigma^2_{\mathrm{A}_X} & \sigma_{\mathrm{A_{XY}}} \\ \sigma_{\mathrm{A_{XY}}} & \sigma^2_{\mathrm{A_Y}} \end{array} \right). \tag{7.3}$$

Similar to the previous chapter, two measures of surrogacy are computed: an individual-level and a disease-level surrogacy. The individual-level surrogacy measures the association between MRI and histology after adjusting for possible genotype effects, and is obtained from the covariance matrices defined in (7.3) such that, $R^2_{ind} = \rho^2_{W}$ and $R^2_{ind} = \rho^2_{A}$, for wildtype and transgenic mice, respectively,

$$\rho_{\mathrm{W}} = \frac{\sigma_{W_{XY}}}{\sqrt{\sigma^2_{W_X}.\sigma^2_{W_Y}}} \text{ and } \rho_{\mathrm{A}} = \frac{\sigma_{A_{XY}}}{\sqrt{\sigma^2_{A_X}.\sigma^2_{A_Y}}}. \tag{7.4}$$

The genotype-specific adjusted correlations ρ_{W} and ρ_{A} measures the association between the two endpoints adjusted for the time evolution of the disease. A large absolute value of the adjusted correlation, or equivalently, a large R^2_{ind} value imply better surrogacy at individual level.

As explained in Chapter 5 and 6, the second level of surrogacy, disease-level surrogacy, is related to the association between the disease evolution on both endpoints and in particular, it focuses on the question of whether the AD evolution on the MRI parameter

can predict the AD evolution on histology. In order to obtain the disease-level surrogacy, we specify a bivariate normal distribution prior for the disease effects given by,

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim Normal\left(\begin{bmatrix} \mu_\alpha \\ \mu_\beta \end{bmatrix}, D\right) \text{and } D = \begin{pmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{pmatrix}. \tag{7.5}$$

The disease-level surrogacy R_D^2, is defined as the square of the correlation coefficient $\rho_{_{\alpha\beta}}$ which is derived from (7.5) as follows:

$$R_{\mathrm{D}}^2 = \left(\rho_{_{\alpha\beta}}\right)^2 = \left(\frac{\sigma_{\alpha\beta}}{\sigma_\alpha \times \sigma_\beta}\right)^2. \tag{7.6}$$

A large R_D^2 is desirable. For a biomarker to be validated as a good surrogate at disease and/or individual level, it is required to have at least one or both of surrogacy measures (individual-level surrogacy and disease-level surrogacy) to be sufficiently high.

The following hyperprior distributions are specified for the covariance matrices in (7.3):

$$\Sigma_W^{-1} \sim Wishart\left(Q_W, \phi\right) \text{ and } \Sigma_A^{-1} \sim Wishart\left(Q_A, \phi\right). \tag{7.7}$$

Here, Q_W and Q_A are 2×2 diagonal matrices with diagonal elements equal to 0.001 and $\phi = 3$ are the degrees of freedom (Gelman and Hill, 2007, Lesaffre and Lawson, 2012). Similarly, for the disease-level surrogacy estimation, the priors for the covariance matrix of the disease-effects in (7.5) are given by,

$$D^{-1} \sim Wishart(Q_D, \phi). \tag{7.8}$$

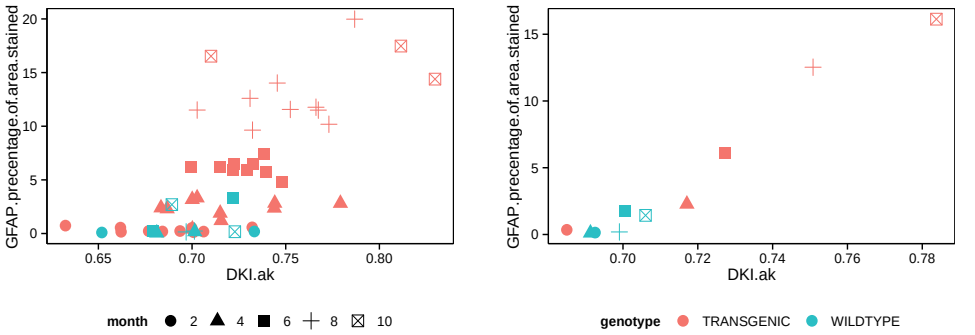
Where Q_D is a 2×2 diagonal matrix with elements equal to 0.001, $\phi = 3$ are the degrees of freedom. The hyperprior for the overall mean of the effects is given by

$$\mu_\alpha \sim \mathrm{Normal}\left(0.0, 1.0E-6\right) \text{ and } \mu_\beta \sim Normal\left(0.0, 1.0E-6\right). \tag{7.9}$$

To complete the specification of the hierarchical model, the priors for the intercepts in the linear predictors in (7.13) are given by,

$$\begin{aligned} &\mu_{X_j} \sim Normal\left(0.0, \tau_X\right), \\ &\mu_{Y_j} \sim Normal\left(0.0, \tau_X\right), \\ &\tau_{_X} \sim Gamma\left(0.001, 0.001\right), \\ &\tau_{_Y} \sim Gamma\left(0.001, 0.001\right). \end{aligned} \tag{7.10}$$

Finally, taking into account biological considerations, an age-invariant parameter for the histology endpoint in wildtype mice. That is, $\mu_{_{Y_j}} = \mu_{_Y}$, is specified since for healthy mice, there is no disease pathology developing due to ageing. Therefore, histology staining is expected to be constant even as the animal grows older (Kohama *et al.*, 1995).



(a) Histology versus MRI. **(b)** Age dependent disease progression effects.

Figure 7.1: The motor cortex region: GFAP percentage of area stained versus MRI-AK (mm^2/s). Blue symbols: wildtype mice. Red symbols: transgenic mice.

7.3 Application to the Data

To illustrate the methodology discussed above, MRI-AK and MRI-RD parameters from the motor cortex region in the brain are used to evaluate the validity of these endpoints as biomarkers for a GFAP histology staining feature (percentage of area stained).

7.3.1 Motor Cortex: MRI-AK with GFAP Staining

The motor cortex data for MRI-AK and GFAP histology staining is shown in Figure 7.1a. The "true" endpoint (GFAP percentage of area stained) shown on the Y-axis seems to be constant for wildtype mice, while it increases with age for transgenic mice. This is in agreement with the biological expectation, hence the choice of a common intercept $\mu_{Y_j} = \mu_Y$ as explained above. Figure 7.1b presents the genotype and age-specific means of true and surrogate endpoints. There is a clear age-dependent genotype (disease) effect on both the true and surrogate endpoints.

The proposed joint hierarchical Bayesian model formulated in Section 7.2 was fitted using a Markov Chain Monte Carlo (MCMC) simulation as implemented in R `runjags` package (Plummer, 2015). Three chains were run in parallel for 60000 iterations with a burn-in period of 30000. Chain mixing and convergence diagnostics were performed (Lesaffre and Lawson, 2012), whereby the Gelman and Rubin potential scale reduction factor close to one was obtained, while the autocorrelation at lag 60 was very small for all parameters. The parameter estimates for the posterior mean of the disease effects are shown in Table 7.1 and Figure 7.2a. For the biomarker (MRI-AK), large disease effects are observed at eight and ten months, while for histology (GFAP percentage of area stained), large disease effects are observed as from four months onwards. Note that the disease effect increases with age in both the histology and MRI-AK endpoints.

Table 7.1: Motor cortex. Posterior mean estimates (95% credible intervals) of the disease effect on true (GFAP staining) and surrogate (MRI-AK) endpoints.

Age	GFAP % area stained	MRI-AK (mm^2/s)
2	0.318 (-0.981, 1.574)	-0.001 (-0.027, 0.025)
4	2.262 (0.969, 3.504)	0.016 (-0.009, 0.044)
6	5.795 (3.932, 7.076)	0.029 (0.001, 0.058)
8	12.393 (10.999, 13.668)	0.058 (0.023, 0.092)
10	15.637 (13.41, 17.566)	0.077 (0.036, 0.12)

The disease effects on both endpoints are visualised from the plots shown in Figure 7.2. The posterior distribution of the histology endpoint (Figure 7.2e) shows little overlap in the distributions with increasing age, compared to that of the surrogate endpoint (Figure 7.2f). For both endpoints, a clear shift to the right (with respect to age) of the posterior distribution is observed.

The posterior mean estimates of the residuals (upon adjusting for the disease effects) covariance matrices (7.3) are given by,

$$\hat{\Sigma}_{\mathrm{W}} = \begin{pmatrix} 0.001 & 0.004 \\ 0.004 & 1.437 \end{pmatrix} \text{and } \hat{\Sigma}_{\mathrm{A}} = \begin{pmatrix} 0.001 & 0.006 \\ 0.006 & 2.557 \end{pmatrix}. \tag{7.11}$$

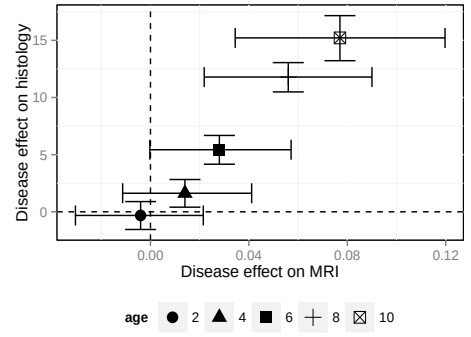
From these, the posterior means for the individual-level correlation coefficients (95% credible interval) were given by $\rho_{_W} = 0.149\ (-0.504,\ 0.769)$ and $\rho_{_A} = 0.133\ (-0.186,\ 0.427)$ (see Figure 7.2c and 7.2d), while the posterior medians were $\rho_{_W} = 0.162$ and $\rho_{_A} = 0.134$. The posterior mean of the individual-level surrogacy, R^2_{ind}, is $0.112\ (0.0001,\ 0.473)$ and $0.042\ (0.00005,\ 0.181)$ for wildtype and transgenic mice, respectively, (posterior medians are 0.07 and 0.02, respectively). This implies that MRI-AK is not a valid biomarker to GFAP percentage of area stained at individual level.

For the disease-level surrogacy, the posterior mean estimate of the correlation between the disease effects $\rho_{_{\alpha\beta}} = 0.760\ (-0.354,\ 0.989)$ as shown in Figure 7.2b, while the posterior mean of the disease-level surrogacy $R^2_D = 0.689\ (0.017,\ 0.979)$ (posterior median for $\rho_{\alpha\beta} = 0.890$ and $R^2_D = 0.792$). This suggests that for the motor cortex, MRI-AK is a valid biomarker for GFAP (percentage of area stained) at the disease-level. However, the precision is low.

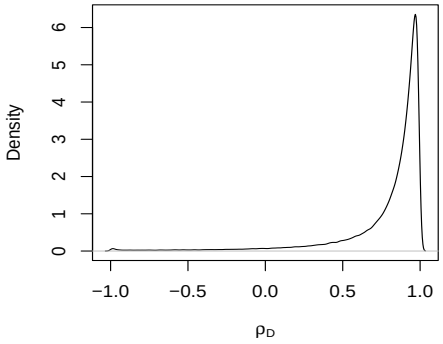
7.3.2 Motor Cortex: MRI-RD with GFAP Staining

The data for MRI-RD with GFAP staining in the motor cortex is shown in Figure 7.3. Although the histology endpoint (GFAP percentage or area stained) seems to vary with age (Figure 7.3a), there was no clear separation of genotype in the MRI-RD axis. The genotype and age-specific means are shown in Figure 7.3b, which indicate relatively low association between the disease effects for the two endpoints.

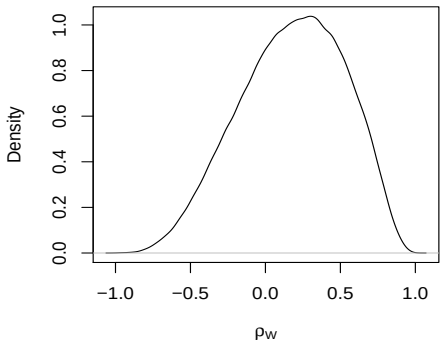
Parameter estimates of the posterior mean of the disease effects are shown in Table 7.2. As can be seen in Figure 7.4a, the posterior mean estimate of the disease effects



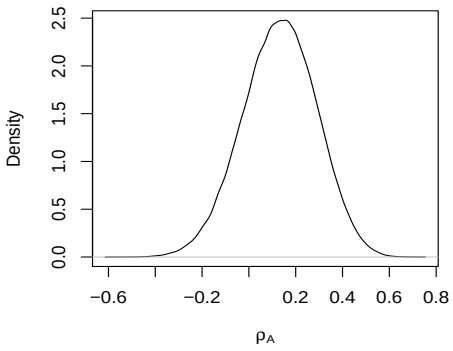
(a) Disease effects and 95% credible intervals.



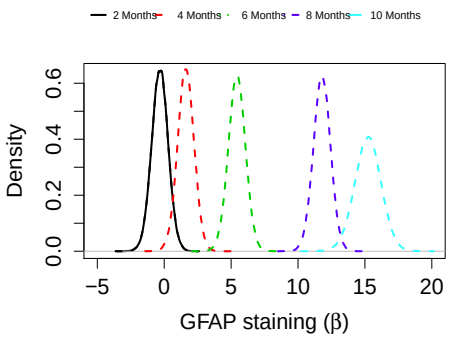
(b) Disease-level correlation, ρ_D.



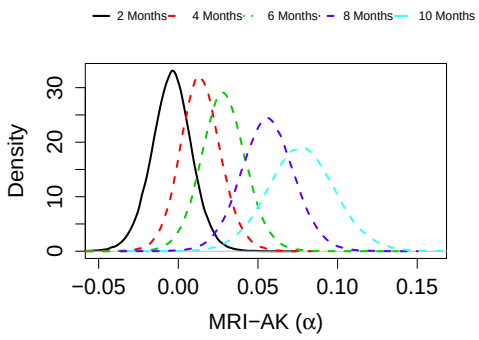
(c) Wildtype: individual-level correlation, ρ_W.



(d) Transgenic: individual-level correlation, ρ_A.

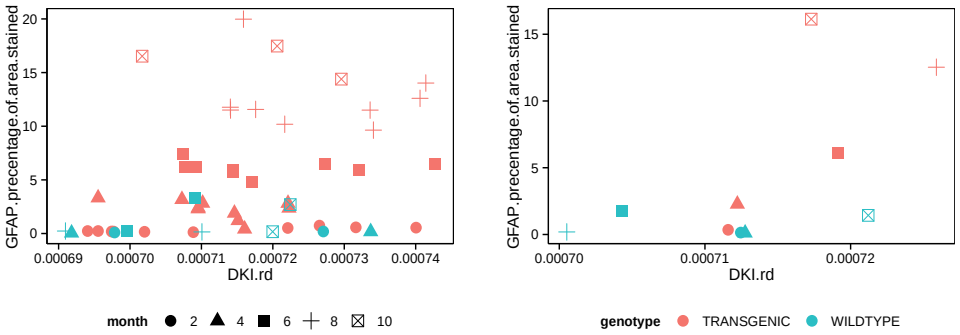


(e) Disease effect on true endpoint.



(f) Disease effect on the surrogate.

Figure 7.2: Motor cortex: GFAP staining (percentage of area stained) and MRI-AK (mm^2/s). Posterior mean estimates for the disease effect on the true and surrogate endpoints, with the corresponding posterior densities.



(a) Histology versus MRI. **(b)** Age dependent disease progression effects.

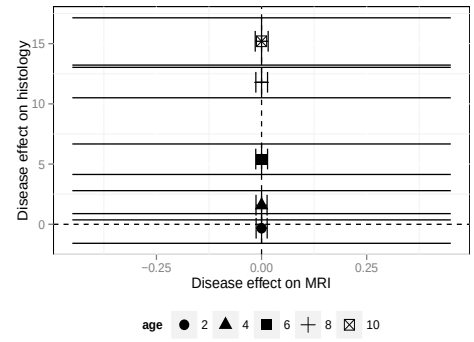
Figure 7.3: Motor cortex: evaluation of surrogacy for MRI-RD (mm^2/s) and GFAP percentage of area stained. Blue symbols: wildtype mice. Red symbols: transgenic mice.

Table 7.2: Motor cortex: posterior mean estimates (95% credible intervals) of the disease effect on "true" (GFAP staining) and surrogate (MRI-RD) endpoints.

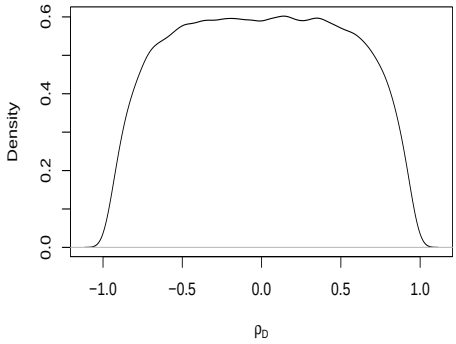
Age	GFAP % area stained	MRI-RD (mm^2/s)
2	0.337 (-1.006, 1.632)	0.00008 (-0.015, 0.015)
4	2.258 (0.918, 3.565)	0.00003 (-0.014, 0.015)
6	5.543 (3.653, 7.004)	0.0001 (-0.016, 0.016)
8	12.368 (10.94, 13.683)	0.00006 (-0.014, 0.015)
10	15.368 (13.029, 17.45)	0.00015 (-0.017, 0.017)

indicate a negligible disease effect on MRI-RD in the motor cortex (see also Figure 7.4b). The kernel density estimates for the disease effects on the histology (GFAP percentage of area stained) and MRI-RD endpoints are presented in Figure 7.4e and 7.4f. Although there is a clear age-dependent separation of genotype with GFAP staining (Figure 7.4e), MRI-RD (Figure 7.4f) cannot distinguish between the genotypes for all ages.

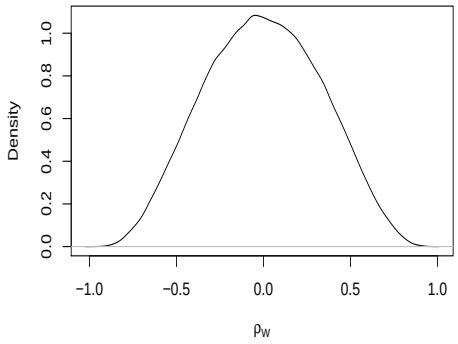
Figure 7.4c and 7.4d shows the posterior density estimate for the residual correlation after adjusting for the disease effects, which indicate weak surrogacy of MRI-RD for GFAP staining at the individual level. Furthermore, the posterior mean of individual-level surrogacy R^2_{ind} is $0.105\ (0.0001,\ 0.45)$ and $0.026\ (0.000025,\ 0.125)$ for wildtype and transgenic mice, respectively, an indication that MRI-RD is not a valid surrogate for GFAP staining in the motor cortex at individual level. Similarly, the posterior mean estimate of the disease-level correlation $\rho_{_D} = 0.002\ (-0.860,\ 0.860,\ \text{see Figure 7.4b})$, while the disease-level surrogacy measure $R^2_{_D} = 0.255\ (0.001,\ 0.814)$ is relatively low, implying that MRI-RD was not a valid biomarker for GFAP staining in the motor cortex at disease-level as well.



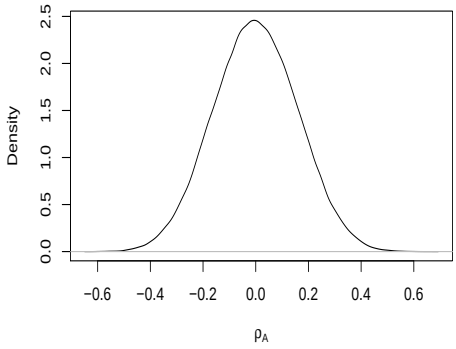
(a) Disease effects (with 95% credible intervals).



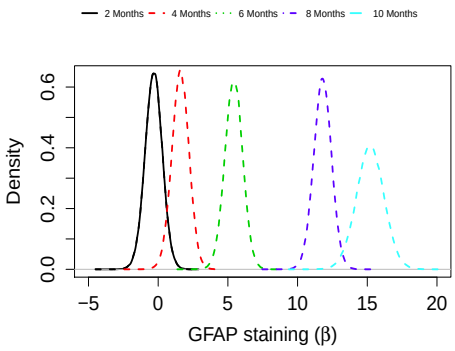
(b) Disease-level correlations.



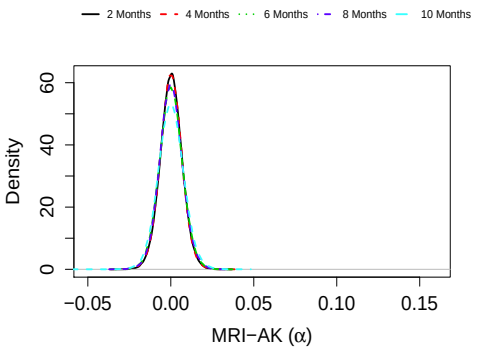
(c) Wildtype: individual-level correlation.



(d) Transgenic: individual-level correlation.



(e) Disease effect on true endpoint.



(f) Disease effect on the surrogate.

Figure 7.4: Motor cortex: GFAP staining and MRI-RD (mm^2/s). Posterior mean estimates for the disease effect on the true and surrogate endpoints, with the corresponding posterior densities.

7.4 Simulation Study

In order to evaluate the performance of the hierarchical Bayesian model presented in Section 7.2, a simulation study was performed. Two different aspects of the study design were investigated: (1) the impact of the sample size of each treatment group (number of transgenic and wildtype mice) and (2) the number of trials (corresponds to number of age groups in the Alzheimer's disease case study). In both cases, the effect on the estimation of individual-level and disease-level surrogacy measures (of varying magnitude) was evaluated.

7.4.1 Simulation Setting

Parameters that were varied during the simulation (shown in Table 7.3) include:

- Disease-level correlation $\rho_{_D} = (0.2, 0.5, 0.85)$, individual-level correlation $\rho_{_W} = \rho_{_A} = (0.2, 0.5, 0.85)$.
- Number of wildtype mice $(5, 10, 20)$ and transgenic $(5, 10, 20)$ mice in the sample.
- The number of trials for which data is obtained $(3, 5, 10)$.

In total, 243 settings were considered. In addition, parameters which were fixed for all simulation settings included $\sigma^2_{W_Y} = \sigma^2_{W_X} = 0.01, \sigma^2_{A_Y} = 0.75, \sigma^2_{A_X} = 0.01$, for the individual-level surrogacy covariance matrices, $\mu_\alpha = 0.25, \mu_\beta = 11.1875, \sigma^2_\alpha = 0.1$ and $\sigma^2_\beta = 47.69$, for the disease-level D components. Note that we fix $\mu_{Xj} = 0.70$ and $\mu_{Yj} = 0.50$, for all $j = 1, 2, \ldots J$, where J is the number of trials (number of time points in which data are observed).

7.4.2 Data Generation and Model Fitting

To generate a dataset, we first sample the disease effects α_j and β_j from a multivariate normal distribution with mean μ_α and μ_β, respectively, and covariance matrix D (according to the prior model specified in (7.12)). Once the values of α_j and β_j are obtained, we sample the individual observations according to model (7.13). Subsequently, data was generated using the procedure described in Section 7.4.2. Note that the individual level correlation was set to be equal in both wildtype and transgenic mice, i.e., $\rho_A = \rho_W$ for all datasets.

Data Generation

The parameter grid for the first 25 settings out of the 243 settings considered in the simulation is presented in Table 7.7. The data generation and estimation of the correlation parameters is performed as follows:

Table 7.3: Simulation study: parameter values for the key parameters explored in the simulations.

	Description	Parameter values
No. trials	Age	3, 5, 10
Sample size	Wildtype	5, 10, 20
	Transgenic	5, 10, 20
Correlation	Individual-level	0.2, 0.5, 0.85
	Disease-level	0.2. 0.5, 0.85

- First, sample the "true" disease effects α_j and β_j from a bivariate normal distribution given by,

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim Normal\left(\begin{bmatrix} \mu_\alpha \\ \mu_\beta \end{bmatrix}, D\right) \text{and } D = \begin{pmatrix} \sigma_{\alpha\alpha} & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_{\beta\beta} \end{pmatrix}. \tag{7.12}$$

  From this step, we estimate the Pearson correlation coefficient, $\rho_{sim} = corr\,(\boldsymbol{\alpha}, \boldsymbol{\beta})$, corresponding to the "true" simulated correlation between the vector of disease effects.

- The disease effect on MRI (α_j) and histology (β_j) are used in order to generate the observations for each rat which are sampled from a bivariate normal distribution as follows:

$$\begin{pmatrix} X_{ij} \\ Y_{ij} \end{pmatrix} \sim Normal\left(\begin{bmatrix} \mu_{x_j} + \alpha_j Z_i \\ \mu_{Y_j} + \beta_j Z_i \end{bmatrix}, \Sigma_k\right). \tag{7.13}$$

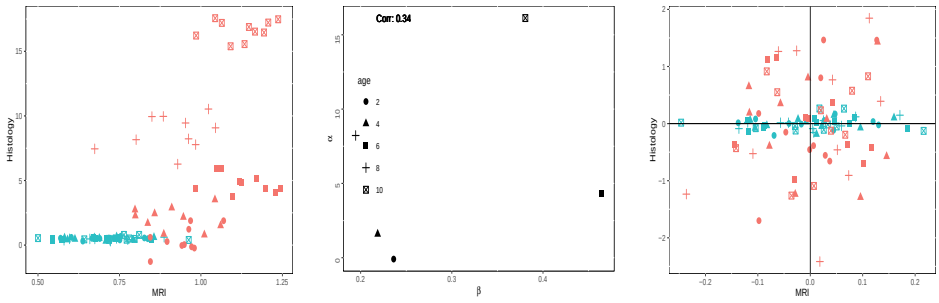
  Note that the dataset obtained from (7.13) is the one to which the proposed hierarchical Bayesian model is fitted.

A selection of simulated datasets is presented in Figure 7.5. Four scenarios are illustrated by varying the disease-level correlation and individual-level correlation between the two extremes (low=0.2 and high=0.85). For each simulation setting, 1000 datasets were generated, and the hierarchical Bayesian model described in Section 7.2 was fitted (with 60000 iterations of which the first 30000 were considered as burn-in period).

The key parameters of interest include the individual-level correlation parameters $\rho_{_A}$ and $\rho_{_W}$ and the disease-level correlation $\rho_{_D}$. In addition, since the "true" disease effects in (7.12) are sampled from a bivariate normal distribution, the simulation study allows us to estimate an additional parameter $\rho_{sim} = cor(\alpha_j, \beta_j)$ which denotes the correlation between the simulated disease effects. Note that this is not be possible from the case study data. This enables us to evaluate the association between the disease effect estimated by the Bayesian hierarchical model, $\hat{\rho_{_D}}$ and the "true" correlation between the disease effects ρ_{sim}.

7.4.3 Simulation Results

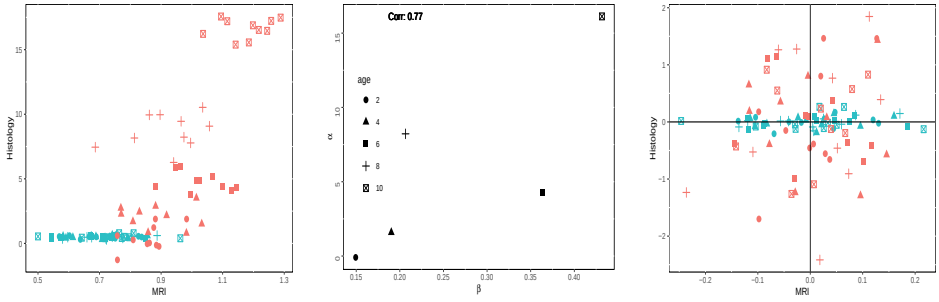
Upon fitting the model, the posterior means of the individual-level correlation $\rho_{_A}$ and $\rho_{_W}$, the posterior mean of the disease-level correlation $\rho_{_D}$ as well as for other parameters



(a) $\rho_D = 0.2, \rho_{ind} = 0.2$

(b) Disease effects.

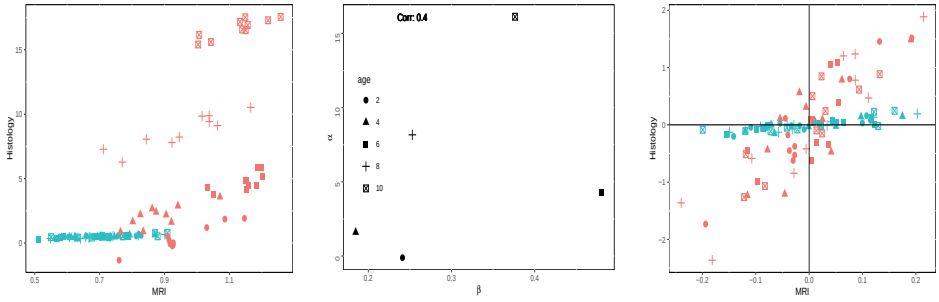
(c) Residuals.



(d) $\rho_D = 0.85, \rho_{ind} = 0.2$

(e) Disease effects.

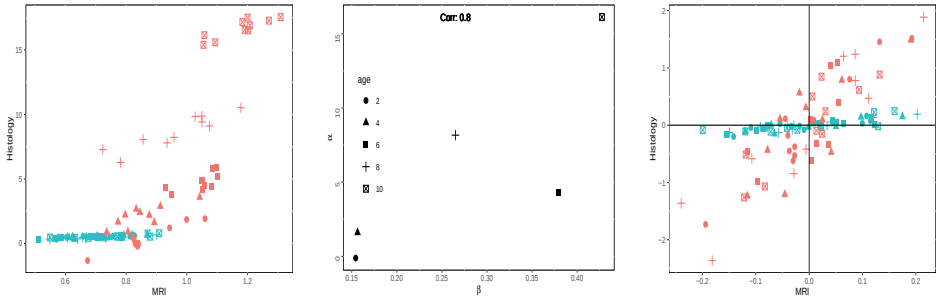
(f) Residuals.



(g) $\rho_D = 0.2, \rho_{ind} = 0.85$

(h) Disease effects.

(i) Residuals.



(j) $\rho_D = 0.85, \rho_{ind} = 0.85$

(k) Disease effects.

(l) Residuals.

Figure 7.5: Selected simulated datasets: only the correlation is varied. The number of trials=5 and sample size=10 for each genotype. Red symbols: transgenic mice. Blue symbols: wildtype mice. All units are arbitrary.

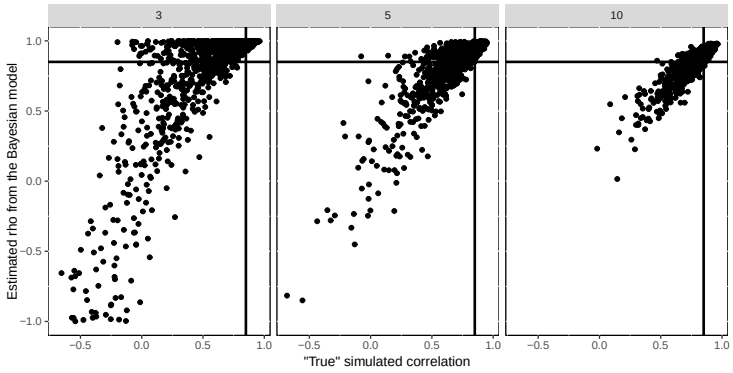


Figure 7.6: Scatterplot between the simulated and the estimated correlations for the disease effect. The panels corresponds to cohorts of 3, 5, and 10 age groups, respectively. The true underlying correlation is 0.85 (see also Figure 7.2 in the appendix for this chapter).

of interest are obtained from the Gibbs sampler.

Disease-Level Surrogacy

For the simulation study discussed in Section 7.4.2, estimates of the two correlation parameters for the disease-level model ($\rho_{_D}$ and ρ_{sim}) were obtained. Figure 7.6 shows the scatterplot of $\rho_{_D}$ against ρ_{sim} for various sample sizes (individual level correlation, $\rho_A = \rho_W = 0.85$). It is clear that the accuracy of both correlation estimates increases as the number of age points increase. Note that the "true" underlying correlation was 0.85 (solid lines). Table 7.4, 7.5, 7.6 and Figure 7.7 show the results for disease-level surrogacy for data with 5, 10 and 20 mice of each genotype. Additional output for different levels of individual and disease-level correlation is shown in the appendix.

While estimating the correlation, there is minimal effect of the composition of wildtye and transgenic mice in the sample. However, the number of age points for which data is obtained has an impact whereby, with three trials, the "true" underlying correlation is under-estimated and as expected, estimation is improved with increasing number of trials. Moreover, for a given trial size, the magnitude of the "true" underlying correlation has an impact on the estimation accuracy, with a high correlation in general being under-estimated (additional output are presented in the supplementary materials).

Individual-Level Surrogacy

Figure 7.8 and Table 7.4, 7.5 and 7.6 present the results for the individual level surrogacy. The results shown in the tables are for data with 5, 10 and 20 mice of each genotype. The "true" individual-level correlation is estimated more precisely as the sample size of the respective genotype increases. Moreover, for a given sample size, an increase in the number of trials for which data was obtained (age groups) resulted in more accurate estimates for the "true" individual-level correlation. In general, larger

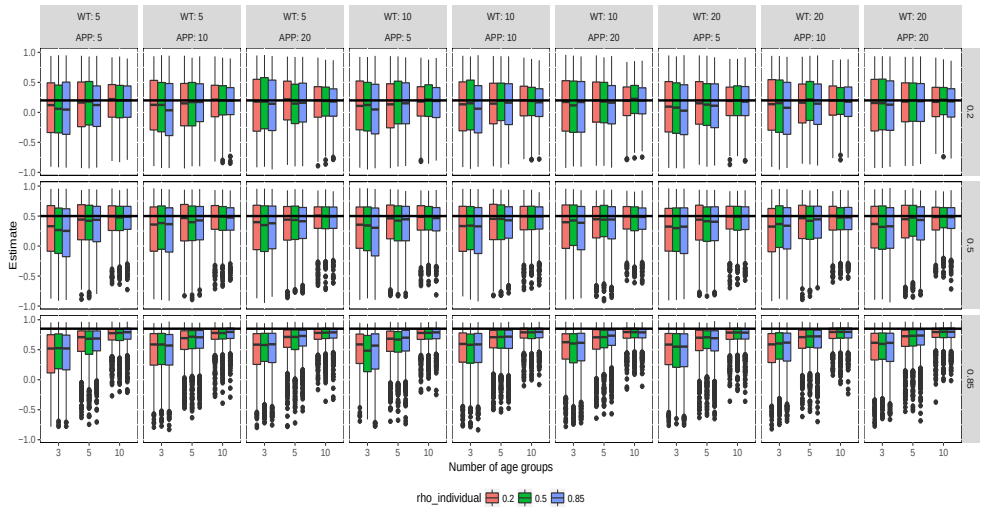


Figure 7.7: Disease-level surrogacy: boxplots of the posterior mean estimates for $\rho_D = 0.2, 0.5, 0.85$. WT: number of wildtype mice in the sample. APP: number of transgenic mice in the sample. The boxplot fill corresponds to individual-level correlations, $\rho_A = \rho_W$. The X-axis shows the number of trials.

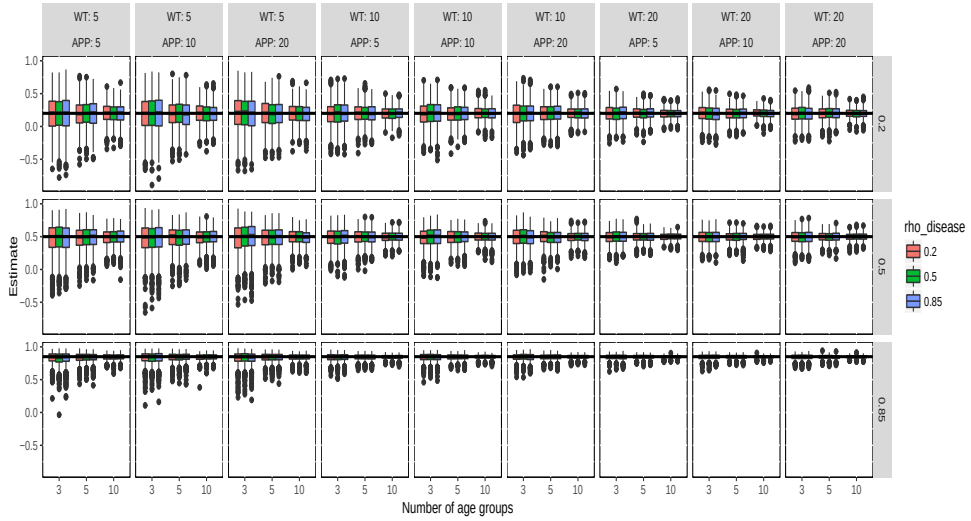
Table 7.4: Posterior means (standard error) of the correlation parameters under different settings, for a sample with 5 transgenic and 5 wildtype mice.

ρ_D	ρ_W	3 age groups			5 age groups			10 age groups		
		$\hat{\rho_D}$	$\hat{\rho_W}$	$\hat{\rho_A}$	$\hat{\rho_D}$	$\hat{\rho_W}$	$\hat{\rho_A}$	$\hat{\rho_D}$	$\hat{\rho_W}$	$\hat{\rho_A}$
0.2	0.2	0.082 (0.48)	0.195 (0.262)	0.197 (0.282)	0.12 (0.454)	0.184 (0.2)	0.193 (0.218)	0.18 (0.366)	0.203 (0.146)	0.2 (0.153)
	0.5	0.05 (0.481)	0.468 (0.23)	0.479 (0.231)	0.135 (0.454)	0.473 (0.172)	0.483 (0.169)	0.173 (0.365)	0.486 (0.122)	0.496 (0.124)
	0.85	0.049 (0.49)	0.83 (0.086)	0.828 (0.096)	0.097 (0.431)	0.834 (0.07)	0.845 (0.064)	0.168 (0.351)	0.842 (0.045)	0.846 (0.047)
0.5	0.2	0.254 (0.475)	0.188 (0.263)	0.196 (0.284)	0.361 (0.397)	0.199 (0.204)	0.19 (0.212)	0.441 (0.301)	0.198 (0.141)	0.196 (0.149)
	0.5	0.22 (0.465)	0.474 (0.222)	0.477 (0.229)	0.347 (0.405)	0.482 (0.164)	0.489 (0.169)	0.437 (0.296)	0.492 (0.116)	0.486 (0.125)
	0.85	0.199 (0.473)	0.816 (0.102)	0.83 (0.101)	0.337 (0.407)	0.836 (0.068)	0.84 (0.071)	0.437 (0.296)	0.842 (0.045)	0.845 (0.045)
0.85	0.2	0.413 (0.395)	0.193 (0.275)	0.194 (0.269)	0.604 (0.296)	0.19 (0.214)	0.18 (0.215)	0.728 (0.172)	0.197 (0.148)	0.196 (0.143)
	0.5	0.436 (0.388)	0.471 (0.226)	0.48 (0.223)	0.582 (0.305)	0.484 (0.164)	0.479 (0.172)	0.727 (0.176)	0.495 (0.121)	0.493 (0.121)
	0.85	0.418 (0.397)	0.825 (0.098)	0.83 (0.092)	0.612 (0.269)	0.836 (0.067)	0.837 (0.074)	0.741 (0.167)	0.845 (0.044)	0.841 (0.046)

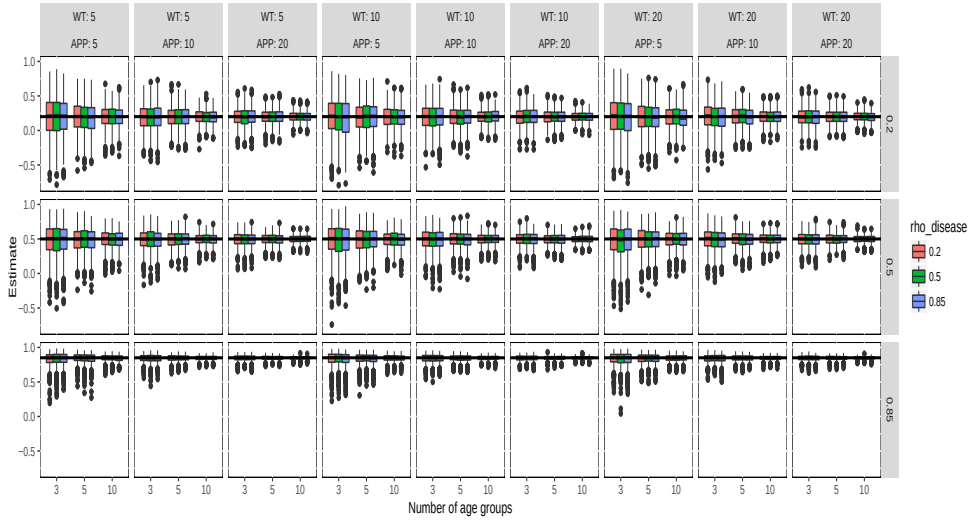
individual-level correlation values are estimated more accurately than smaller correlations. The magnitude of disease-level surrogacy does not have an influence on the estimation of the individual-level surrogacy. Estimates of the mean square error, variance and bias of the individual-level correlation are presented in the supplementary materials.

7.5 Discussion

The Bayesian approach to surrogate validation was adopted and applied on selected MRI parameters in the motor cortex. For the two examples presented in this chapter, individual-level surrogacy was too low. Prediction of GFAP staining given MRI-AK or MRI-RD values of a subject is therefore not possible. On the other hand, predicting



(a) Wildtype mice.



(b) Transgenic mice.

Figure 7.8: Individual-level surrogacy. Boxplots of the posterior mean estimates for $\rho_D = 0.2, 0.5, 0.85$. WT: number of wildtype mice in the sample. APP: number of transgenic mice in the sample. Ind: individual-level correlation $\rho_A = \rho_W$. Number of trials: number of time points for which data is available.

Table 7.5: Posterior means (standard error) of the correlation parameters under different settings, for a sample with 10 transgenic and 10 wildtype mice.

ρ_D	ρ_W	3 age groups			5 age groups			10 age groups		
		$\hat{\rho_D}$	$\hat{\rho_W}$	$\hat{\rho_A}$	$\hat{\rho_D}$	$\hat{\rho_W}$	$\hat{\rho_A}$	$\hat{\rho_D}$	$\hat{\rho_W}$	$\hat{\rho_A}$
0.2	0.2	0.096 (0.489)	0.185 (0.183)	0.196 (0.184)	0.125 (0.428)	0.186 (0.143)	0.198 (0.144)	0.176 (0.334)	0.2 (0.104)	0.2 (0.099)
	0.5	0.111 (0.487)	0.477 (0.151)	0.491 (0.144)	0.16 (0.414)	0.487 (0.119)	0.497 (0.112)	0.178 (0.337)	0.499 (0.078)	0.496 (0.08)
	0.85	0.049 (0.486)	0.838 (0.059)	0.84 (0.059)	0.126 (0.427)	0.844 (0.043)	0.845 (0.043)	0.154 (0.325)	0.848 (0.028)	0.846 (0.032)
0.5	0.2	0.257 (0.46)	0.197 (0.178)	0.194 (0.184)	0.363 (0.402)	0.195 (0.143)	0.185 (0.142)	0.44 (0.29)	0.199 (0.101)	0.196 (0.104)
	0.5	0.262 (0.455)	0.492 (0.147)	0.487 (0.149)	0.363 (0.401)	0.491 (0.114)	0.493 (0.114)	0.446 (0.283)	0.496 (0.08)	0.496 (0.079)
	0.85	0.237 (0.469)	0.84 (0.061)	0.84 (0.06)	0.359 (0.376)	0.843 (0.044)	0.844 (0.045)	0.433 (0.286)	0.848 (0.03)	0.847 (0.031)
0.85	0.2	0.5 (0.35)	0.199 (0.183)	0.197 (0.182)	0.633 (0.259)	0.196 (0.141)	0.199 (0.143)	0.753 (0.148)	0.198 (0.102)	0.203 (0.101)
	0.5	0.482 (0.353)	0.493 (0.148)	0.488 (0.151)	0.639 (0.258)	0.494 (0.115)	0.489 (0.113)	0.75 (0.147)	0.495 (0.08)	0.497 (0.079)
	0.85	0.49 (0.354)	0.838 (0.058)	0.838 (0.061)	0.636 (0.26)	0.844 (0.043)	0.844 (0.044)	0.759 (0.141)	0.847 (0.03)	0.849 (0.03)

Table 7.6: Posterior means (standard error) of the correlation parameters under different settings, for a sample with 20 transgenic and 20 wildtype mice.

ρ_D	ρ_W	3 age groups			5 age groups			10 age groups		
		$\hat{\rho_D}$	$\hat{\rho_W}$	$\hat{\rho_A}$	$\hat{\rho_D}$	$\hat{\rho_W}$	$\hat{\rho_A}$	$\hat{\rho_D}$	$\hat{\rho_W}$	$\hat{\rho_A}$
0.2	0.2	0.119 (0.493)	0.19 (0.123)	0.194 (0.124)	0.153 (0.417)	0.2 (0.102)	0.196 (0.097)	0.162 (0.321)	0.199 (0.068)	0.2 (0.069)
	0.5	0.108 (0.49)	0.491 (0.099)	0.495 (0.099)	0.156 (0.409)	0.497 (0.074)	0.498 (0.079)	0.178 (0.317)	0.498 (0.054)	0.498 (0.052)
	0.85	0.102 (0.486)	0.845 (0.038)	0.845 (0.037)	0.144 (0.414)	0.847 (0.029)	0.848 (0.031)	0.156 (0.318)	0.849 (0.02)	0.849 (0.02)
0.5	0.2	0.285 (0.444)	0.198 (0.127)	0.193 (0.13)	0.373 (0.38)	0.202 (0.101)	0.198 (0.096)	0.437 (0.271)	0.194 (0.07)	0.201 (0.071)
	0.5	0.268 (0.453)	0.49 (0.1)	0.492 (0.099)	0.386 (0.374)	0.493 (0.075)	0.492 (0.078)	0.454 (0.25)	0.499 (0.056)	0.498 (0.054)
	0.85	0.265 (0.443)	0.845 (0.039)	0.846 (0.041)	0.354 (0.374)	0.847 (0.031)	0.847 (0.03)	0.443 (0.258)	0.848 (0.02)	0.847 (0.021)
0.85	0.2	0.507 (0.347)	0.197 (0.126)	0.195 (0.128)	0.65 (0.246)	0.2 (0.101)	0.195 (0.097)	0.756 (0.153)	0.197 (0.066)	0.195 (0.069)
	0.5	0.517 (0.327)	0.491 (0.104)	0.493 (0.101)	0.662 (0.246)	0.499 (0.078)	0.497 (0.075)	0.763 (0.137)	0.499 (0.054)	0.497 (0.055)
	0.85	0.51 (0.34)	0.845 (0.038)	0.844 (0.039)	0.671 (0.23)	0.844 (0.028)	0.846 (0.028)	0.766 (0.133)	0.848 (0.021)	0.848 (0.02)

the disease effect on the GFAP staining from the disease effect on MRI-AK can be performed since the disease-level surrogacy for MRI-AK parameter was estimated to be equal to $\rho_D = 0.76$. Note that the disease-level surrogacy of MRI-RD for GFAP staining in the motor cortex is relatively low, hence MRI-RD cannot be a biomarker for GFAP staining in motor cortex. These results (and others not shown here) indicate that, the validity of MRI parameters as surrogates for disease pathology largely depends on the MRI parameter being evaluated.

Simulation studies reveals that, the estimation of disease-level surrogacy is only affected by the trial size (i.e., the number of age points in which MRI and histology are measured) and the magnitude of the "true" underlying disease-level correlation. As the number of trials increases, there is a reduction in the magnitude of the under-estimation of the "true" disease-level correlation. Moreover, for trials of the same size, if the "true" disease-level correlation is large, the magnitude of under-estimation is larger compared to the setting in which the underlying "true" disease-level correlation is low. This therefore implies that for the case studies presented in this Chapter, the reported disease-level surrogacy might be higher than the estimated value. The under-estimation may result in an MRI parameter with moderate disease-level correlation being declared an invalid biomarker for further consideration. The accuracy of the estimation of individual-level surrogacy is not only affected by the magnitude of the "true" individual-level correlation, but also by the sample size of the respective genotype. In contrast to the disease-level surrogacy, there is minimal bias in the estimation of the individual-level surrogacy.

7.6 Appendix

7.6.1 Disease-Level Surrogacy

In order to fully understand the final simulation output, a smaller simulation study was performed on the performance of the correlation parameter for different variances and sample sizes. We sampled data from a bivariate normal distribution with mean $\mu = (0.25, 11.1875)$. The covariance matrix had a correlation $\rho = 0.85$, while the variance components were set to $\sigma_1^2 = \sigma_2^2 = 0.1$ or $\sigma_1^2 = \sigma_2^2 = 1.5$. In both cases, 1000 datasets were generated for samples of size $3, 5, 10$ or 20. Figure 7.9 shows that only the sample size of a dataset has an influence on the distribution of the sampled correlation.

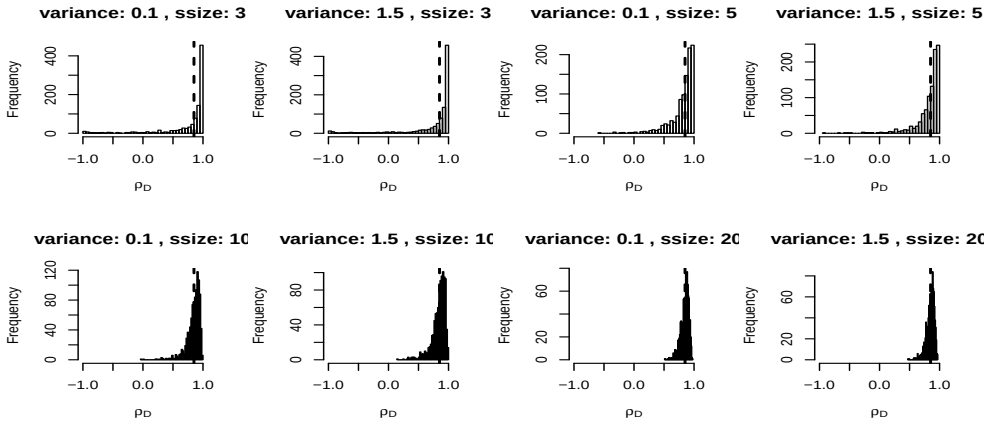


Figure 7.9: Density estimate for the distribution of the correlation. The broken line denotes the true correlation (0.85).

7.6.2 Additional Output for the Simulation Study

Scatter plots of the estimated correlation $(\rho_{_D})$ versus the simulated correlation (ρ_{sim}) between the disease effects are shown in Figure 7.10, 7.11 and 7.12 for trials of size $3, 5$ and 10, respectively. For a given trial size, the genotype composition and the magnitude of the individual-level surrogacy does not affect the accuracy of the estimation. Note that the correlation between ρ_D and ρ_{sim} increases as the number of trials (i.e. age points in which MRI and histology are measured) increases. As we mentioned in the main paper, the correlation ρ_{sim} can be estimated only in the simulation study and not in real data. The mean, variance and bias plots for disease-level surrogacy are shown in Figure 7.13. For the individual-level surrogacy, the MSE, variance and bias plots for wildtype and transgenic mice are presented in Figure 7.14 and 7.15, respectively.

Table 7.7: Simulation study: parameter grid of all the simulation settings considered. WT: wildtype mice sample size. APP: transgenic mice sample size. Trials: number of trials.

Setting	WT	APP	Trials	ρ_D	$\rho_A = \rho_W$
1	5	5	3	0.2	0.2
2	10	5	3	0.2	0.2
3	20	5	3	0.2	0.2
4	5	10	3	0.2	0.2
5	10	10	3	0.2	0.2
6	20	10	3	0.2	0.2
7	5	20	3	0.2	0.2
8	10	20	3	0.2	0.2
9	20	20	3	0.2	0.2
10	5	5	5	0.2	0.2
11	10	5	5	0.2	0.2
12	20	5	5	0.2	0.2
13	5	10	5	0.2	0.2
14	10	10	5	0.2	0.2
15	20	10	5	0.2	0.2
16	5	20	5	0.2	0.2
17	10	20	5	0.2	0.2
18	20	20	5	0.2	0.2
19	5	5	10	0.2	0.2
20	10	5	10	0.2	0.2
21	20	5	10	0.2	0.2
22	5	10	10	0.2	0.2
23	10	10	10	0.2	0.2
24	20	10	10	0.2	0.2
25	5	20	10	0.2	0.2

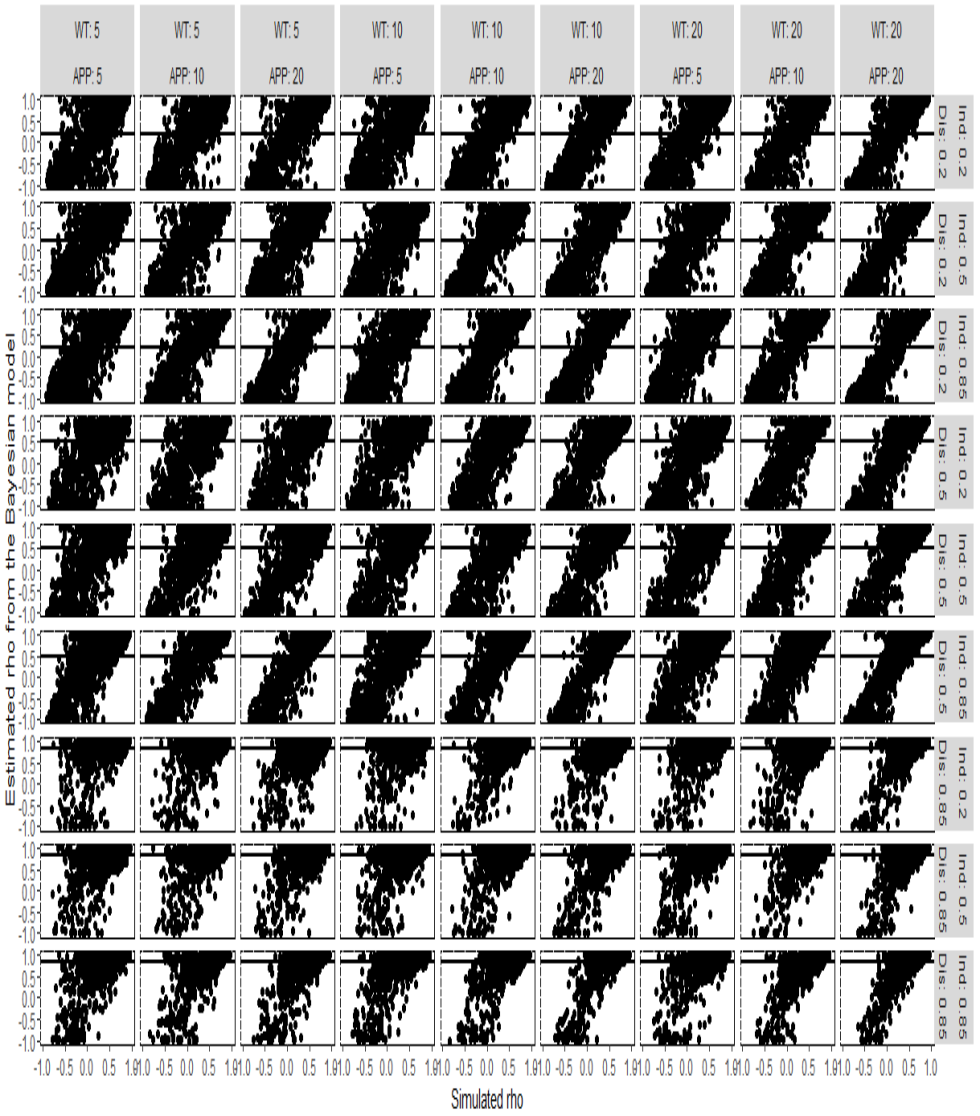


Figure 7.10: Disease-level surrogacy for trial of size 3. Scatter plot of the simulated and estimated correlation between the disease effects. The true disease-level correlation, ρ_D is shown by the solid line. Ind: the true individual-level correlation, $\rho_A = \rho_w$. WT: wildtype mice. APP: transgenic mice.

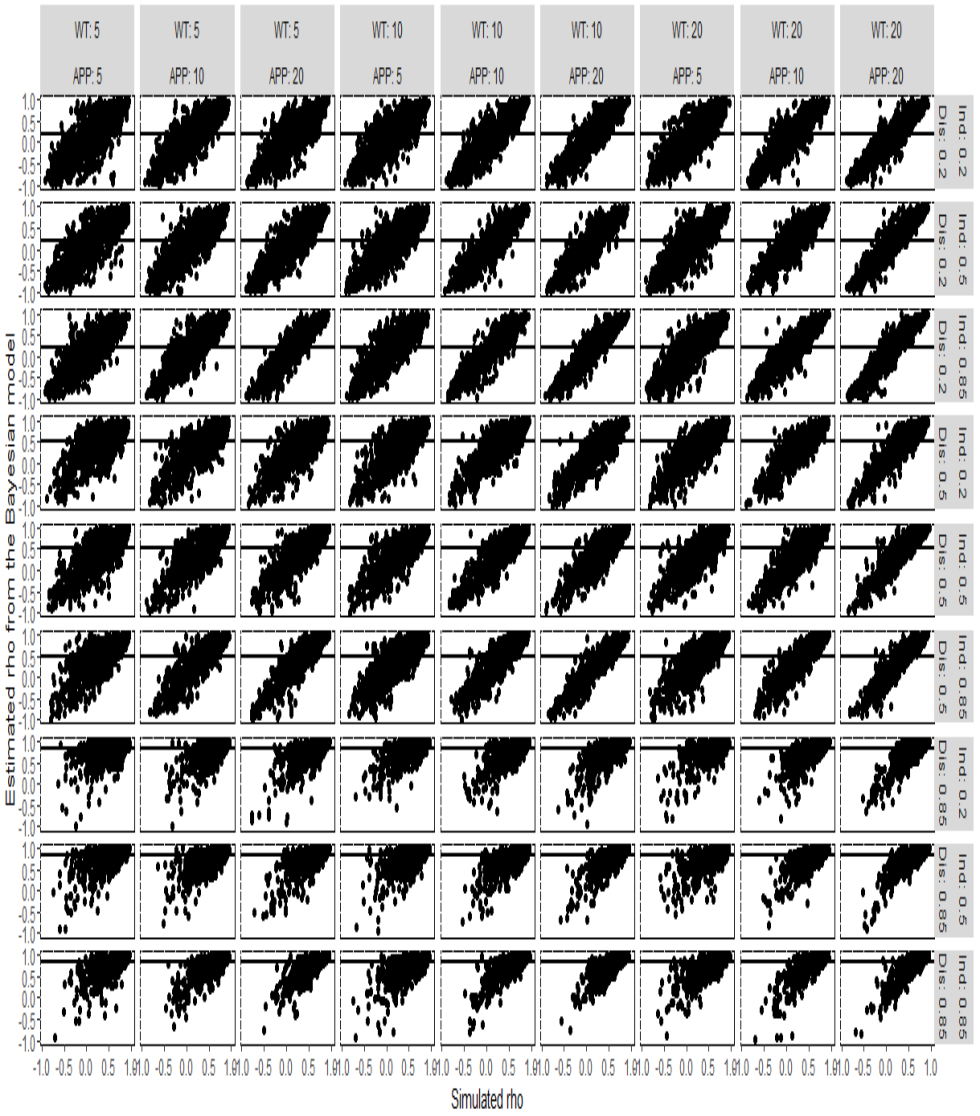


Figure 7.11: Disease-level surrogacy for trial of size 5. Scatter plot of the simulated and estimated correlation between the disease effects. The true disease-level correlation, ρ_D is shown by the solid line. Ind: the true individual-level correlation, $\rho_A = \rho_w$. WT: wildtype mice. APP: transgenic mice.

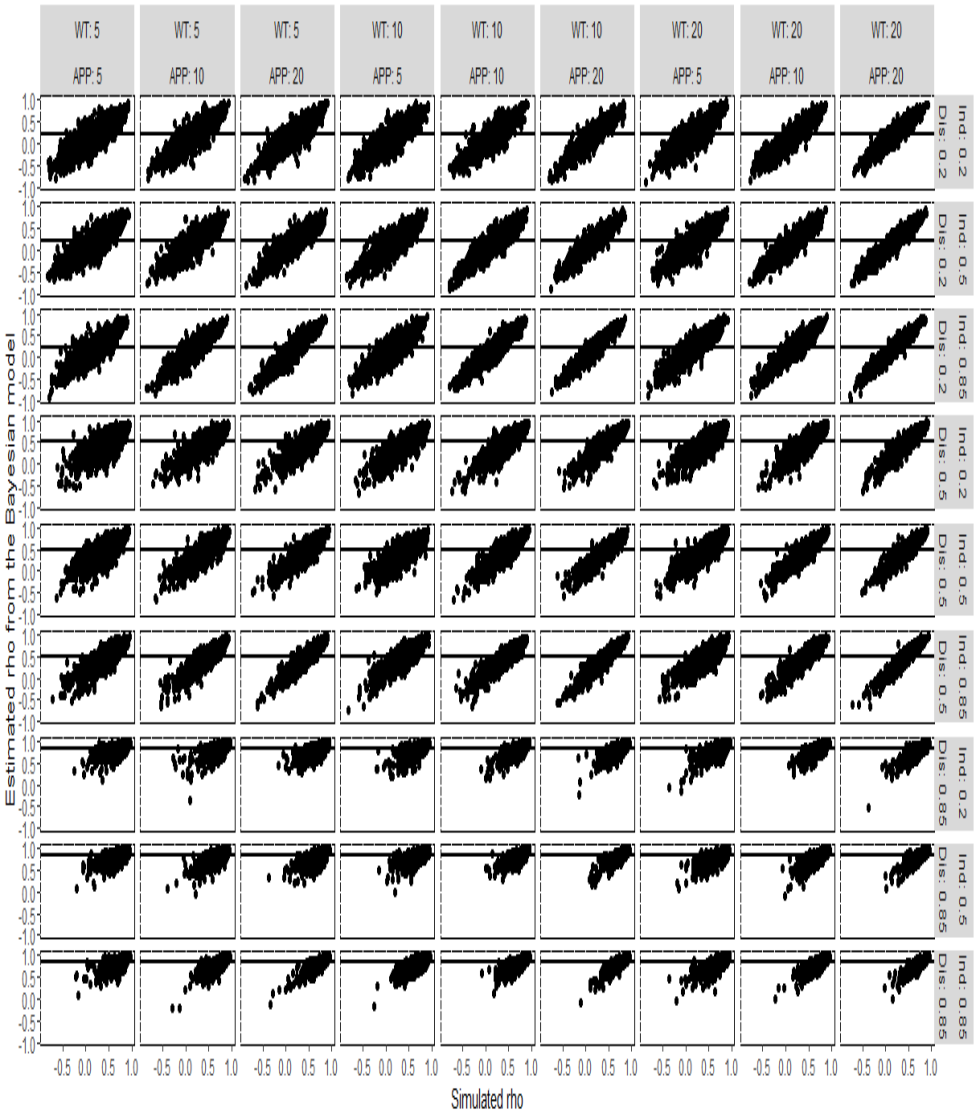


Figure 7.12: Disease-level surrogacy for trial of size 10. Scatter plot of the simulated and estimated correlation between the disease effects. The true disease-level correlation, $\rho_D$is shown by the solid line. Ind: the true individual-level correlation, $\rho_A = \rho_w$. WT: wildtype mice. APP: transgenic mice.

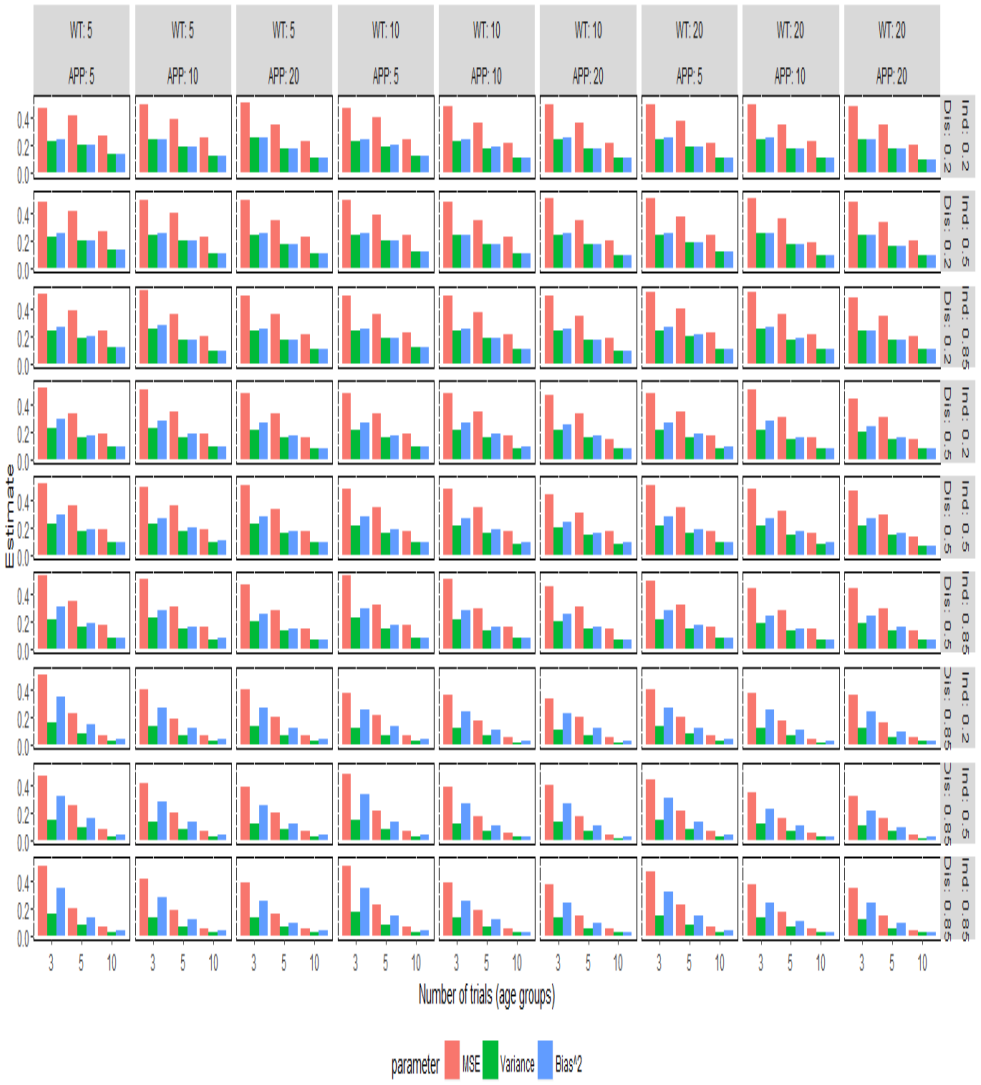


Figure 7.13: Disease-level surrogacy. Simulation performance measures for $\rho_D = 0.2, 0.5, 0.85$. WT: number of wildtype mice in the sample. APP: number of transgenic mice in the sample. Ind: individual-level correlation $\rho_A = \rho_W$. Number of trials: number of time points for which data is available.

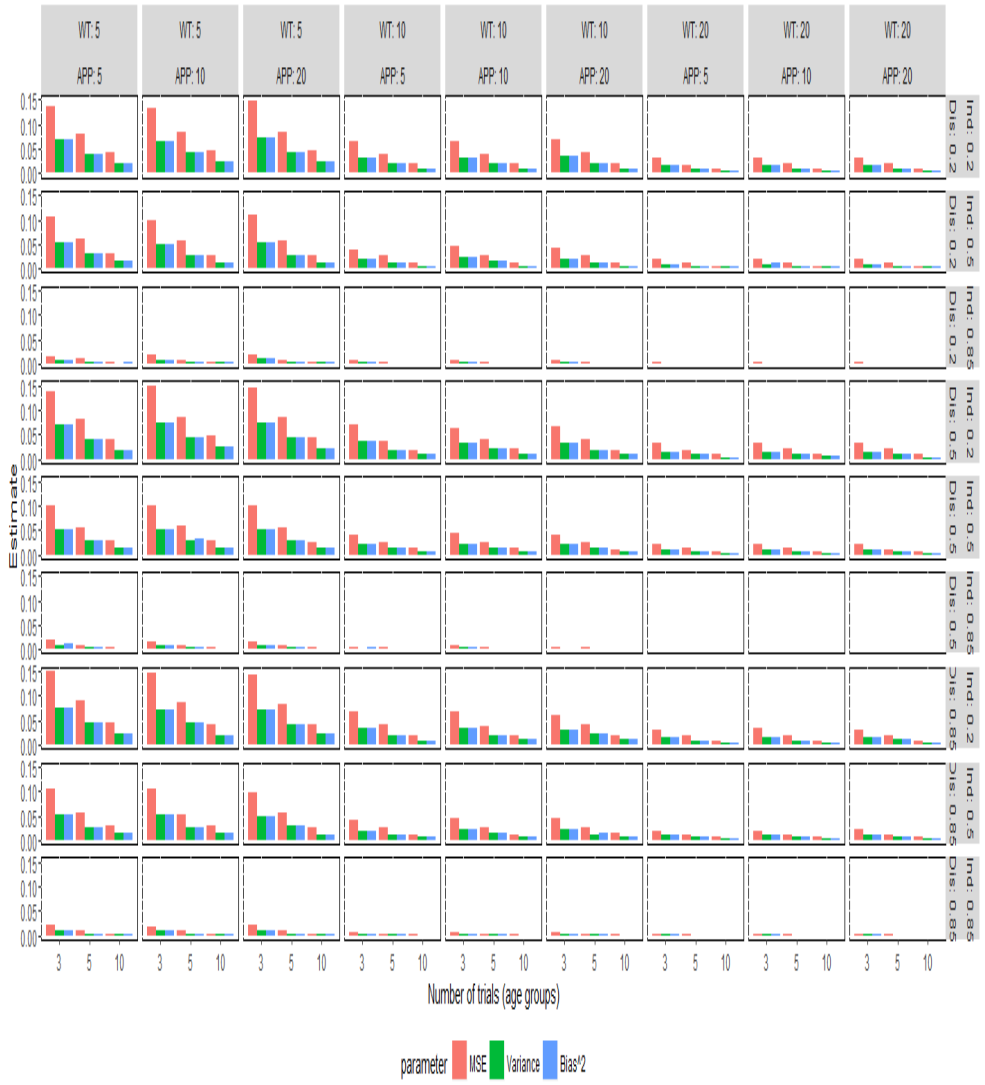


Figure 7.14: Wildtype individual-level surrogacy. Simulation performance measures for $\rho_W = 0.2, 0.5, 0.85$. WT: number of wildtype mice in the sample. APP: number of transgenic mice in the sample. Ind: individual-level correlation $\rho_A = \rho_W$. Number of trials: number of time points for which data is available.

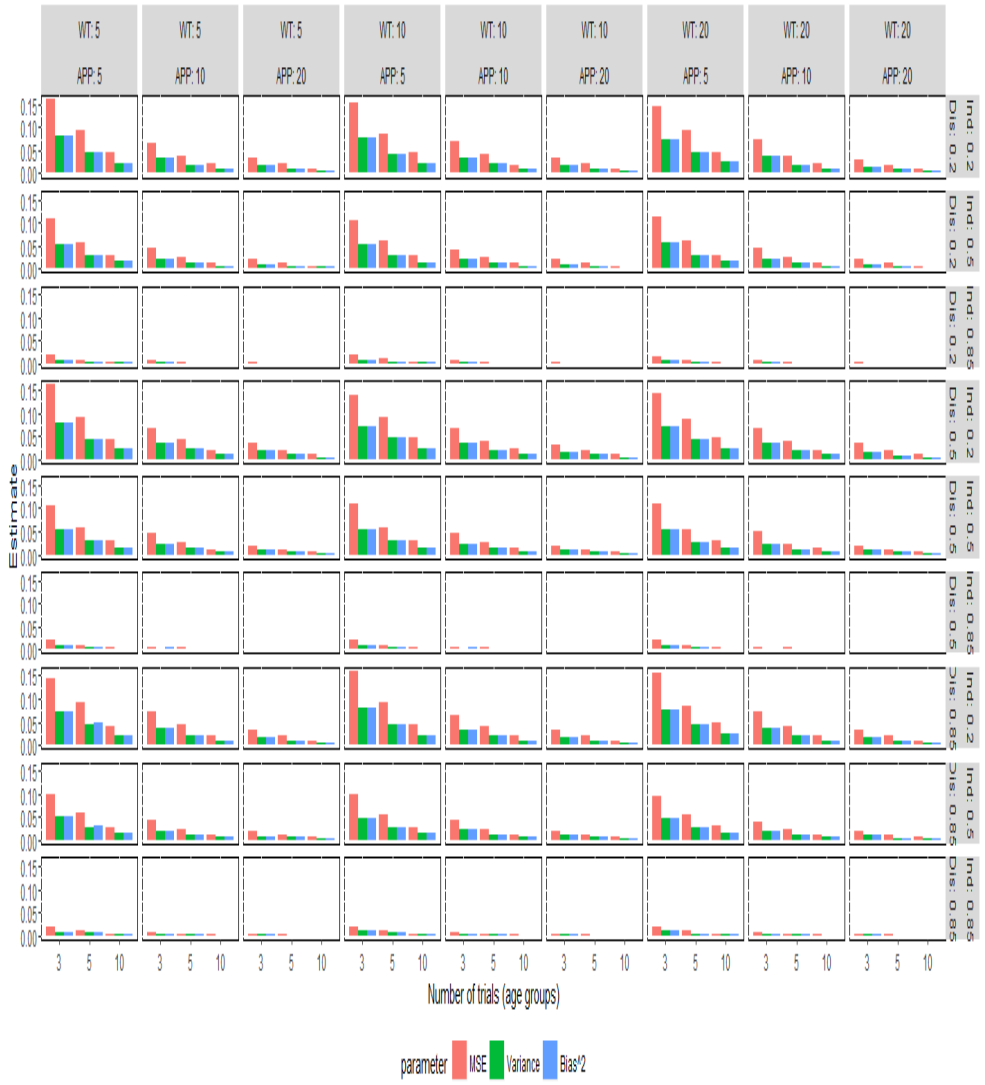


Figure 7.15: Transgenic individual-level surrogacy. Simulation performance measures for $\rho_A = 0.2, 0.5, 0.85$. WT: number of wildtype mice in the sample. APP: number of transgenic mice in the sample. Ind: individual-level correlation $\rho_A = \rho_W$. Number of trials: number of time points for which data is available.

Part II

Joint Modelling of Correlated Data: Overdispersion, Bayesian Variable Selection and Order-Restricted Hypotheses

Chapter 8

Model Selection and Uncertainty in Dose-Response Modelling

8.1 Introduction

Scientific research is often characterized by presence of competing models (and theories) from which the most probable model (and theory) given observed data is selected. In drug development for instance, different doses of a drug may be administered and the effect of the treatment investigated. Dose-response studies often presumes that there is a monotone relationship between administered doses and the observed response. However, the choice of an appropriate model to describe a biological process and the inference resulting from the parameter estimates is always associated with uncertainty regarding their suitability in capturing the underlying process (Briggs *et al.*, 2012, Klingenberg, 2009). Traditionally, uncertainty in parameter estimates has been quantified by confidence (credible) intervals around a point estimate, which provide information about the accuracy of the estimates.

Ordinarily, scientists/statisticians may hypothesise several potential models for the underlying biological process and apply model selection criteria such as Akaike information criterion (AIC, Akaike, 1973), likelihood-based ratio test or Deviance Information Criterion (DIC, Spiegelhalter *et al.*, 2002) to choose the "best-fitting model" (Zucchini, 2000). However, the shortcoming with these measures is that, they do not directly quantify the uncertainty associated with the chosen model, therefore ignoring the fact that such a model was selected from a set of *a priori* defined models. It is necessary to account for and quantify this uncertainty since inference on the resulting parameter estimates is dependent on the selected model (Raftery, 1995).

8.1.1 Bayesian Model Averaging

One method that accounts for model uncertainty is model averaging. In this framework, a weight is assigned to each model in the set of plausible models. Further, the resulting parameter estimates are a weighted average (using the model weights) of the estimates from each model. There exists both frequentist and Bayesian approaches to model averaging and quantification of the model uncertainty (Buckland *et al.*, 1997, Cripps *et al.*, 2005, Claeskens and Hjort, 2008, Burnham and Anderson, 2003). In this part of the thesis, we focus on the Bayesian approach applied to dose-response models with and without order constraints on some of the parameters. Kato and Hoijtink (2006), Klugkist and Hoijtink (2007) and Claeskens and Hjort (2008) proposed a Bayesian approach whereby, the Bayes Factor (BF) and an appropriate encompassing prior were used to compute the posterior model probabilities of the competing models.

Let $g_0 \dots g_R$ be the set of plausible models. The posterior model probability for model g_r, assuming that they were all equally probable *a priori* i.e $P(g_r) = 1/R$, for $r = 0, ..., R$, is based on the Bayes factor BF_{r0}. Let BF_{00} be the Bayes factor of the null model that assumes no dose effect, let D represent the data and $BF_{00} = 1$. The posterior probability is given by

$$P(g_r|D, g_0 \dots g_R) = \frac{BF_{r0}}{BF_{00} + BF_{10} + ... + BF_{R0}}. \tag{8.1}$$

Whitney and Ryan (2009) used an alternative computation for the posterior model probability for model g_r based on

$$\begin{aligned} P(g_r|D, g_0 \dots g_R) &= \frac{P(D|g_r)p(g_r)}{\sum_{i=0}^{R} P(D|\boldsymbol{\theta}_r g_r)p(g_r)}, \\ P(D|g_r) &= \int P(D|\theta_r, g_r)P(\boldsymbol{\theta}_r|g_r)d\boldsymbol{\theta}_r. \end{aligned} \tag{8.2}$$

Here, $P(g_r)$ is the prior probability of the rth model. Note that a non-informative prior, $P(g_r) = 1/R$ is used, $P(D|g_r)$ is the model likelihood, while $\boldsymbol{\theta_r}$ is the vector of parameters in the rth model, with appropriately defined non-informative priors, $P(\boldsymbol{\theta}_r|g_r)$. Similar to the usual model averaging approach (Burnham and Anderson, 2003, Claeskens and Hjort, 2008), an approximation to the posterior probability in (8.2) can be obtained using model information criteria such as Schwarz's Bayesian Information Criterion (BIC, Schwarz, 1978). In this case, the posterior model probability is given by

$$P(g_r|D, g_0 \dots g_R) = \frac{[\exp(-0.5BIC(g_r))]}{\sum_{i=0}^{R} [\exp(-0.5BIC(g_r))]P(g_r)}. \tag{8.3}$$

Note that other information criteria such as Akaike Information Criterion (AIC) and Deviance Information Criterion (DIC) can be used.

8.1.2 Bayesian Variable Selection

Given a set of potential predictors, $X_1, \ldots, X_p$, the relationship between the outcome Y_i and the predictors $\mathbf{X}$, is given by

$$g(Y_i) = \sum_{j=1}^{p} X_{ij}\beta_j. \tag{8.4}$$

Here, β_j are regression coefficients, and $g(.)$ is an appropriate link function.

Model selection entails identifying a subset of predictors, $X_1^*, \ldots, X_q^*$, corresponding to the "best" model. Selection of the "best" model out of the 2^p submodels can be performed using information criteria such as AIC and BIC. However, the number of submodels increases with the predictor set $\mathbf{X}$, therefore increasing the computational requirements.

George and McCulloch (1993) proposed a Bayesian approach to variable selection whereby, the hierarchical model is extended with an indicator variable

$$\mathrm{z} = \begin{cases} 1, \text{if } \beta_j \text{ sampled from } f_2, \\ 0, \text{if } \beta_j \text{ sampled from } f_1. \end{cases} \tag{8.5}$$

The prior distribution of β_j is a mixture of two normal distributions, f_1 and f_2, given by

$$\beta_j | z_j \sim (1 - z_j) f_1 + z_j f_2. \tag{8.6}$$

The posterior probability of z_j is the posterior inclusion probability, which is an indicator of whether X_j should be included in the model or not. Both f_1 and f_2 are assumed to have zero mean. Note that the variance in f_1 is set to be very small such that, if $z_j = 0$, β_j is very small and can be safely assumed to be zero. The "best" model is the one for which the subset, $X_1^*, \ldots, X_q^*$, has a high posterior mean of $\bar{z_1}, \ldots, \bar{z_q}$. The posterior mean estimates effect are obtained using Gibbs sampling (Dellaportas *et al.*, 2002).

A second approach for bayesian variable selection (O'Hara and Sillanpaa, 2009) assume two independent priors for β_j and z_j,

$$\begin{aligned} \beta_j &\sim Normal\left(0, \sigma^2\right), \\ z_j &\sim Bernoulli\left(\pi_j\right), \\ \pi_j &\sim Uniform\left(0, 1\right). \end{aligned} \tag{8.7}$$

For the prior specification in (8.7), the linear predictor is given by

$$\sum_{j=1}^{p} X_{ij} z_j \beta_j. \tag{8.8}$$

Note that, β_j is included in the model whenever $z_j = 1$. In the context of dose-response modelling, Otava *et al.* (2013), Lin *et al.* (2012) and Kasim *et al.* (2012)

applied Bayesian Variable Selection (O'Hara and Sillanpaa, 2009) methodology in model selection for gene-expression profiles. We adapt this later approach, which we apply to several case studies both for continuous, binomial and Poisson outcomes. In all the case studies, BVS is performed only on a selected subset of the parameters.

8.2 Gaussian Case Studies: Longitudinal Dose-Response Data

8.2.1 The Wistar Rat Data

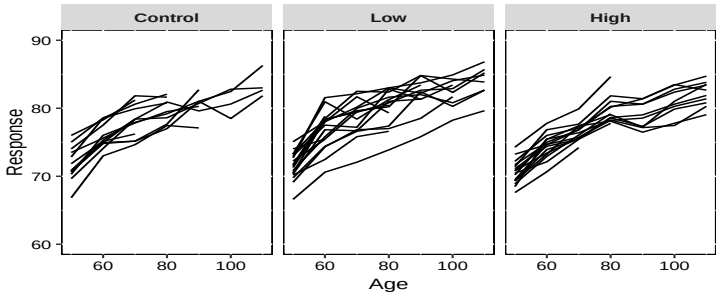
Verdonck *et al.* (1998) described an experiment to evaluate the effect of inhibiting the production of testosterone on craniofacial growth of Wistar rats. In their study, 50 male Wistar rats were randomized to either control, or one of two treatment groups. For the treatment, a low or high dose of Decapeptyl, which inhibits the production of testosterone, were administered to rats aged 45 days and measurements taken every 10 days thereafter (the first measurement was taken at 50 days). Rats were anaesthetized, an X-ray of the skull taken, and the the distance (in pixels) between two predefined points recorded. The measurement characterizing the height of the skull will be analysed. In Figure 8.1a, the individual rat profiles are presented from which, there is evidence of drop-out at later time points mainly attributable to deaths due to anaesthesia administration. The average evolution profiles are shown in Figure 8.1b.

8.2.2 The Milk Protein Content Trial

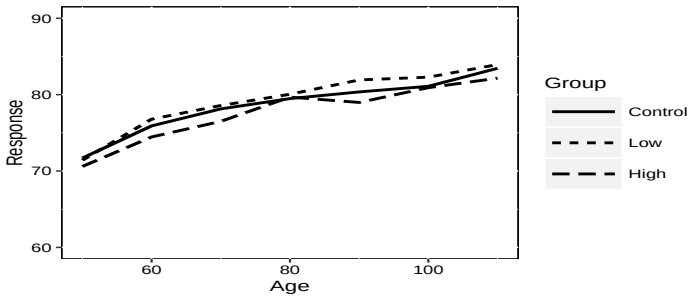
The study was designed in order to assess the impact of diet on a cow's milk protein content in Australia (Verbyla and Cullis, 1990). In total, 79 Australian cows were randomized into a diet of barley only (25), lupins only (27) or a barley-lupins mixture (27), after calving. Subsequently, assays of the milk's protein content were taken weekly for 19 weeks. The cow-specific evolution profiles are shown in Figure 8.2a, while the mean evolution for each diet is presented in Figure 8.2b. Note that although data was available from almost all cows for the first 14 weeks, subsequent weeks had data from only $59, 50, 46, 46$ and 41 cows, respectively. Statistical anaylses taking into account this 'dropout' can be found in Verbeke and Molenberghs (2000). Moreover, according to Verbyla and Cullis (1990), the first three weeks were considered to be settling-in period hence, will be omitted from subsequent analyses.

8.3 Non-Gaussian Case Studies: DRL-72 Data

In behavioral experiments, a standard protocol commonly applied for testing potential clinically active antidepressants is the Differential Reinforcement Low-Rate 72 seconds schedule (DRL-72, Evenden *et al.*, 1993). In this protocol, rats are trained to press a lever in order to obtain a reward (for instance, a pellet). A rat only obtains a reward if the inter-response time (time between subsequent lever presses) is at least 72 seconds. Each rat was subjected to several doses of a candidate anti-depressant treatment and

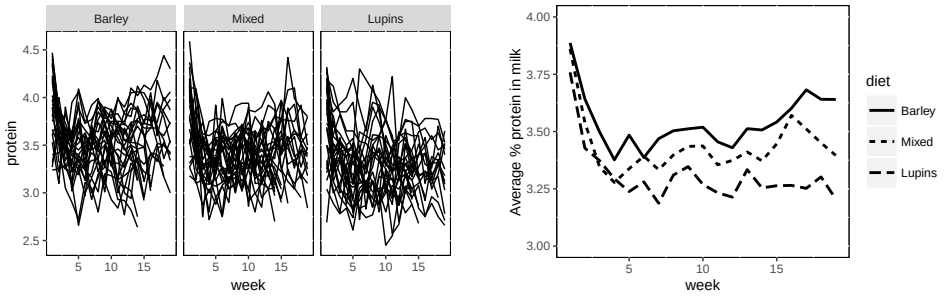


(a) Individual rat profiles per dose.



(b) Average rat profiles per dose.

Figure 8.1: The rat data. Observed evolution profiles.



(a) Individual evolution profiles. **(b)** Average evolution profiles.

Figure 8.2: Milk protein content trial. Observed data profiles.

for each administered dose, the rat was monitored for 60 minutes. In this duration, the inter-response times and the number of rewards obtained were recorded. Subsequently, the number of times the rat pressed the lever (response) was derived as the number of recorded inter-response times. There is interdependence of the two outcomes in that, the number of rewards (the binomial outcome) depends on the number of responses (the Poisson outcome) since both outcomes are measured on the same subject. Two DRL-72 case studies are analysed, the first involving a parallel drug administration, while the second has a crossover design.

8.3.1 Parallel Design DRL-72 Case Study

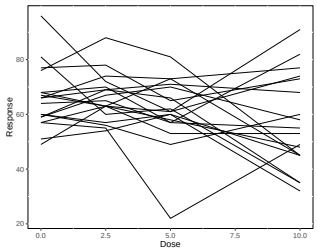
In the parallel design case study, 18 rats were each subjected to four doses of a treatment (0.0, 2.5, 5.0 and 10.0mg/kg) administered via an injection through the skin, resulting in a cluster of four measurements per outcome for each rat. Note that in this dissertation, the term "parallel design" refers to the fact that the same sequence of the four doses was administered to each rat in the study. An adequate washout period was provided for hence, no carrover effect of the previously administered doses was expected. Panels a and b in Figure 8.3 present the individual response and rewards profiles, respectively, while Panels c and d shows the average number of responses and rewards, respectively, across all four doses.

8.3.2 Crossover Design DRL-72 Case Study

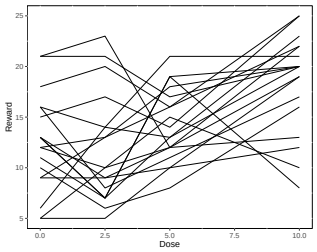
The second case study comes from a crossover DRL-72 experiment design, previously analysed by Shkedy *et al.* (2005). In their study, a crossover design (Jones and Kenward, 1989) comprising of four blocks, three periods and 5 active treatments per block was used. In total, 20 rats were randomised into a three-period sequence and the respective doses administered. Further, for every period, each day of treatment administration was preceded by a training day in which only the placebo was administered. Note that in a crossover design, there is potential carryover effect of a treatment between subsequent periods. The observed data for the crossover case study is presented in Figure 8.4 (panel a and b shows the individual profile plots for the number of responses and number of rewards, respectively, while panel c and d presents the respective average profiles).

8.3.3 Analysis Plan

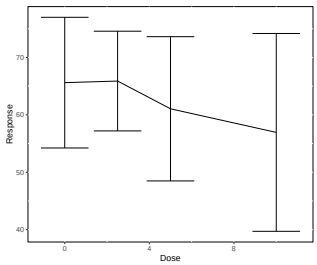
An overview of the analysis plan for the dose-response datasets is shown in Figure 8.5. When the outcome in dose-response experiments is Poisson or binomial, there is potential overdispersion which ought to be accounted for. Chapter 9 presents a hierachical Bayesian joint model to account for overdispersion in Poisson and binomial outcomes. In order to account for model uncertainty, we present the Bayesian variable selection methodology applied to hierarchical joint dose-response models with and without overdispersion (Chapter 10) and to order-restricted dose-response models whereby, the dose-response profile is



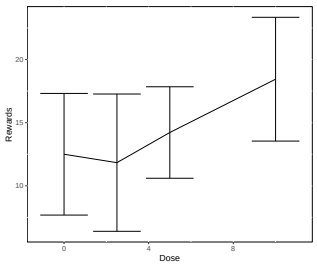
(a) Responses: individual profile plots.



(b) Rewards: individual profile plot.

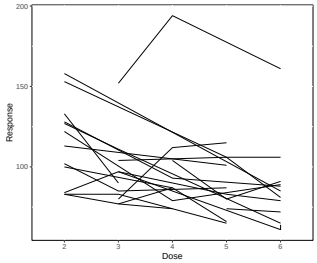


(c) Average number of responses.

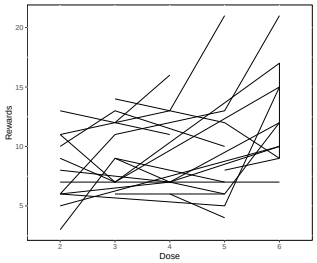


(d) Average number of rewards.

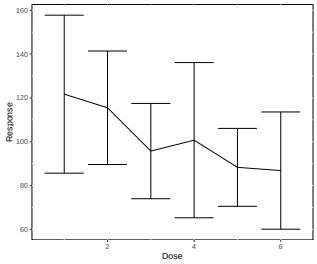
Figure 8.3: Parallel design DRL-72 experiment observed data. Subject-specific and average dose-response profiles. For the average evolution, one standard deviation error bars are included.



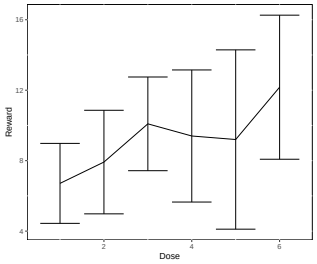
(a) Response: individual profile plot.



(b) Rewards: individual profile plot.



(c) Average number of responses.



(d) Average number of rewards.

Figure 8.4: Crossover DRL-72 experiment observed data: subject-specific and average dose-response profiles. For the average evolution, one standard deviation error bars are included.

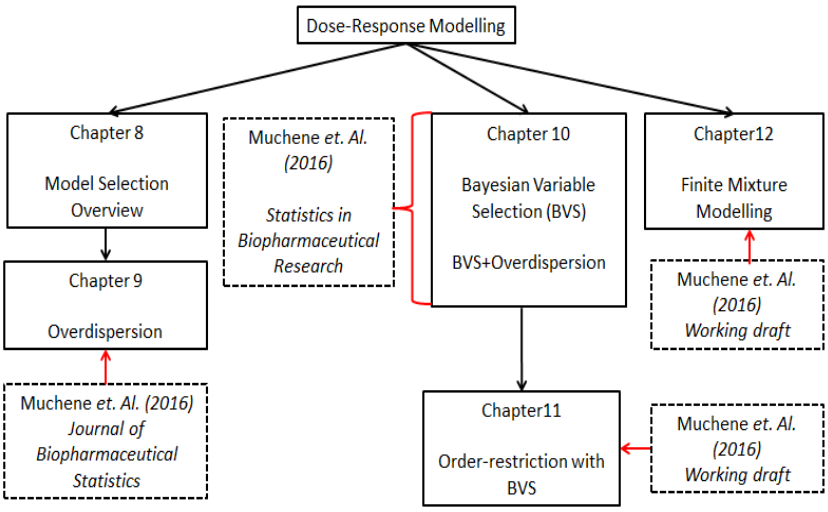


Figure 8.5: Overview of the analysis plan for the dose-response data and publication strategy.

assumed to be monotone (Chapter 11). Finally, in Chapter 12, an alternative modelling approach for the DRL-72 studies is provided.

Chapter 9

Overdispersion in Hierarchical Bayesian Joint Models for Correlated Data

9.1 Introduction

In drug discovery research, it is common to measure outcomes that are binary in nature, such as the success or failure of a treatment in inducing a desirable effect, as well as count outcomes such as the number of cells affected by a toxic effect. For these outcomes, the distributional assumption is often binomial and Poisson distributions, respectively. One fundamental feature of these distributions is that, the variance is a function of the mean. For the Poisson distribution, both mean and variance are assumed to be equal, while for the binomial distribution, the variance is a multiplicative function of the mean. Over/under-dispersion in a Poisson distribution occurs when the population variance is larger/smaller than the population mean (McCullagh and Nelder, 1989). If over/under-dispersion is not accounted for, the resulting inference for parameters of interest may be misleading since the variance will be over/under-estimated.

Several models have been proposed to account for overdispersion. For example, in the Poisson case, Nelder and Wedderburn (1972) proposed specifying the mean-variance relationship in a generalized linear model and applying the iteratively-weighted least squares algorithm to estimate the scaling parameter. A scaling parameter greater than one indicates presence of overdispersion while less than one indicates underdispersion. An alternative approach is to introduce a multiplicative random effect with unit mean and the variance quantifying the amount of heterogeneity in the observed Poisson counts. This implies that the count variable follows a negative binomial distribution (Breslow, 1984). Zero-inflated Poisson and hurdle models can be applied to count data when overdispersion occurs due to excessive zeros (Lambert, 1992). For the case of binomial data, the beta-binomial model (Skellam, 1948) allows the probability of an event to be

randomly sampled from a beta distribution with unit mean.

To account for extra variability within the setting of Poisson and binomial longitudinal/correlated data, Molenberghs *et al.* (2007, 2010) proposed the so-called combined models, which account for clustering of outcomes and incorporate parameters for overdispersion as well. Del Fava *et al.* (2014) applied the combined models to a joint model for a HCV and HIV longitudinal study with binary outcomes, while Milanzi *et al.* (2012) assessed the impact of ignoring overdispersion as well as the sensitivity to misspecification of the underlying distribution for the overdispersion random-effect within the combined modelling framework. Although there have been many developments and extensions of the combined model proposed by Molenberghs *et al.* (2007), the method has not yet been explored and applied to parallel and crossover dose-response trials requiring joint modelling of outcomes such as behavioural experiments in drug discovery.

In this chapter, we extend the hierarchical Bayesian model to jointly model the Poisson and binomial outcomes in the two DRL-72 case studies discussed in Chapter 8. The proposed model allows for additive or multiplicative overdispersion on the Poisson and binomial outcomes separately as well as on both outcomes simultaneously. The chapter is arranged as follows; the statistical methodology and model formulation are presented in Section 9.2. Application of the statistical methodology to the two case studies is presented in Section 9.3. Finally, we close with a discussion in Section 9.4.

9.2 A Joint Model for the Number of Responses and Rewards

9.2.1 Hierarchical Bayesian Binomial-Poisson Model for the DRL-72 Experiment

The first model we consider in this chapter is a hierarchical Bayesian binomial-Poisson model. Let n_{ij} be the number of times rat $i = 1, 2, \ldots, I$, pressed the lever while under dose $j = 1 \ldots J$, and Y_{ij} be the number of rewards obtained. The distribution assumptions are

$$\begin{aligned} Y_{ij} &\sim Binomial(n_{ij}, \pi_{ij}), \\ n_{ij} &\sim Poisson(\lambda_{ij}). \end{aligned} \tag{9.1}$$

The linear predictors for the mean response and reward are given, respectively, by

$$\begin{aligned} \log(\lambda_{ij}) &= \mathbf{X}_i\boldsymbol{\alpha} + a_i, \\ \text{logit}(\pi_{ij}) &= \mathbf{X}_i\boldsymbol{\beta} + b_i. \end{aligned} \tag{9.2}$$

Here, $\mathbf{X}_i$ is an $n_i \times p$ design matrix, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are vectors of unknown coefficients considered to be fixed parameters, while a_i and b_i are the subject-specific random intercepts for the ith rat.

The two case studies, based on parallel and crossover designs, differ in their specification of model (9.2) for the number of responses and number of rewards obtained. These differences are highlighted in Sections 9.2.1 and 9.2.1.

Model Formulation: A DRL-72 Experiment with a Parallel Design

For the parallel design of the DRL-72 experiment, all doses of the drug were administered to each rat resulting in a set of associated outcomes. As explained in the previous chapter, in this dissertation, the term "parallel design" refers to the fact that the same sequence of the four doses was administered to each rat in the study. The fixed effects component in the linear predictor $\mathbf{X}_i\boldsymbol{\alpha}$ and $\mathbf{X}_i\boldsymbol{\beta}$ in (9.2) can be denoted as

$$\begin{aligned}\mathbf{X}_i\boldsymbol{\alpha} &= \alpha_0 + \sum_{j=1}^{3} \alpha_{1j} d_{ij},\\ \mathbf{X}_i\boldsymbol{\beta} &= \beta_0 + \sum_{j=1}^{3} \beta_{1j} d_{ij}.\end{aligned} \tag{9.3}$$

Here, α_0 and β_0 are the placebo effects, while α_j and β_j are the dose effects, $j = 1, 2, 3$, for the number of responses and number of rewards, respectively, and $d_{ij} = 1$ if dose j is given to the ith subject and zero otherwise. Note that $j = 1, 2, 3$, corresponding to 2.5, 5 and 10 mg/kg, respectively.

Model Formulation: A DRL-72 Experiment with a Crossover Design

The crossover design requires an adjustment in the mean structure for carryover and the period effects in addition to the dose effect.The mean structures for the number of responses and reward in the crossover model is denoted as

$$\begin{aligned}\mathbf{X}_i\boldsymbol{\alpha} &= \alpha_0 + \alpha_{1j} + \alpha_{2k} + \alpha_{3\ell},\\ \mathbf{X}_i\boldsymbol{\beta} &= \beta_0 + \beta_{1j} + \beta_{2k} + \beta_{3\ell}.\end{aligned} \tag{9.4}$$

The linear predictors specified in (9.4) take the form for dose j in period k, with dose ℓ administered in the preceding period. Moreover, α_0 and β_0 are the placebo effects in period 1, α_{1j} and β_{1j} the effects of dose j, $j = 0, 1.25, 2.5, 5.0, 10$ mg/kg, α_{2k}, and β_{2k} the effect of period k, $k = 1, 2, 3$, $\alpha_{3\ell}$ and $\beta_{3\ell}$ the carryover effects of dose ℓ, $\ell = 0, 1.25, 2.5, 5, 10$ mg/kg, which was administered in the preceding period.

A Joint Modelling Approach

A joint model for responses and rewards is formulated to capture the association between responses and rewards. This can be achieved by imposing a joint distribution on the random effects:

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim Normal\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{D}\right),\ \mathbf{D} = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}. \tag{9.5}$$

Under the joint bivariate normal distribution specification of the prior model (9.5), the resulting correlation coefficient between the random effects is given by

$$\rho_{ab} = \frac{\sigma_{ab}}{\sqrt{\sigma_a^2.\sigma_b^2}}. \tag{9.6}$$

A negative correlation implies that, at an individual level, low number of responses corresponds to high number of rewards. To complete the model specification, non-informative priors and hyperpriors (for the precision parameters) are defined as

$$\begin{aligned} &\boldsymbol{\alpha} \sim Normal(0, \tau_\alpha), \\ &\boldsymbol{\beta} \sim Normal(0, \tau_\beta), \\ &\tau_\alpha \sim Gamma(0.001, 0.001), \\ &\tau_\beta \sim Gamma(0.001, 0.001). \end{aligned} \tag{9.7}$$

A Wishart prior distribution is specified for the inverse covariance matrix

$$\mathbf{D}^{-1} \sim Wishart(R_{\mathbf{D}}, k). \tag{9.8}$$

Here, the degrees of freedom *k* is set to be at least equal to the rank of $\mathbf{D}^{-1}$ and the scale matrix $R_{\mathbf{D}}$ is a 2×2 diagonal matrix with off-diagonal elements set to 0.001 (Gelman and Hill, 2007, Lesaffre and Lawson, 2012).

9.2.2 Hierarchical Bayesian Joint Model with Overdispersion Parameters

Molenberghs *et al.* (2007, 2010) and Aregay *et al.* (2015) reported an additional gain in accounting for overdispersion in hierarchical models in addition to accounting for clustering. They proposed the introduction of a set of random effects for overdispersion, which can be assumed to follow a normal or gamma distribution. It can be shown (see Molenberghs *et al.*, 2007, 2010) that a multiplicative random effect in a Poisson model results in a negative-binomial distribution while in a binomial model, the beta-binomial distribution is derived. In their model, Molenberghs *et al.* (2007, 2010) assumed an independent multiplicative overdispersion random effect and specified the cluster random effects (9.2) to capture the intra-cluster correlation. On the other hand, Aregay *et al.* (2015) discussed the case of additive overdispersion random effects for both outcomes that can be added to model (9.2) as well. In what follows, we discuss the formulation of combined models with multiplicative and additive overdispersion parameters in the setting of dose-response modelling with parallel and crossover designs.

Multiplicative Overdispersion Models

In this section, we reformulate the joint model given in (9.1)-(9.8) by including multiplicative overdispersion random effects. Let Θ_{1ij} and Θ_{2ij} be subject and dose-specific overdispersion parameters with prior specification given by

$$\begin{aligned} &\Theta_{1ij} \sim Gamma(\theta_1, \theta_1), \\ &\Theta_{2ij} \sim Beta(\theta_2, \theta_2). \end{aligned} \tag{9.9}$$

The joint model formulated in (9.1)-(9.8) can be re-written to account for overdispersion as follows:

$$\begin{aligned} &Y_{ij} \sim Binomial(n_{ij}, \Theta_{2ij}\pi_{ij}), \\ &n_{ij} \sim Poisson(\Theta_{1ij}\lambda_{ij}), \end{aligned} \tag{9.10}$$

The linear predictors for π_{ij} and λ_{ij} remain the same as specified in (9.2). Note that specifying a gamma distribution with both parameters equal results in an overdispersion parameter distribution with mean$= 1$ and variance$= 1/\theta_1$. In order to have positive scale and shape parameters for both the gamma and beta distributions, truncated normal priors for $\theta_1 \sim Normal(0, 10^{-6})T(0,)$ and $\theta_2 \sim Normal(0, 10^{-6})T(0,)$, with precision 10^{-6} are defined. Alternatively, $\theta_1 \sim \text{Uniform}(0, 100)$ and $\theta_2 \sim \text{Uniform}(0, 100)$ may be specified. Note that $\Theta_{1ij} \equiv 1$ implies lack of overdispersion in the Poisson outcome, while $\Theta_{2ij} \equiv 1$ is an indication of no overdispersion in the binomial outcome.

Additive Overdispersion Models

Rather than specifying multiplicative random effects at the mean level, the overdispersion random effects can be incorporated to the linear part of the model as an additive term. Additive overdispersion random effects not only allow for possible correlation between the Poisson and binomial overdispersion parameters, but also, models with additive overdispersion are less computationally intensive. The linear predictors for the additive overdispersion model are given by

$$\begin{aligned} \text{log}(\lambda_{ij}) &= \mathbf{X}_i\boldsymbol{\alpha} + a_i + \Theta_{1ij}, \\ \text{logit}(\pi_{ij}) &= \mathbf{X}_i\boldsymbol{\beta} + b_i + \Theta_{2ij}. \end{aligned} \tag{9.11}$$

Note that we can specify independent priors for Θ_{1ij} and Θ_{2ij}, that is,

$$\begin{aligned} \Theta_{1ij} &\sim Normal(0, \sigma^2_{\theta_1}), \\ \Theta_{2ij} &\sim Normal(0, \sigma^2_{\theta_2}). \end{aligned} \tag{9.12}$$

Alternatively, a joint prior distribution for the overdispersion parameters can be specified by

$$\begin{pmatrix} \Theta_{1ij} \\ \Theta_{2ij} \end{pmatrix} \sim Normal\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{D}_{\boldsymbol{\Theta}} \right), \ \mathbf{D}_{\boldsymbol{\Theta}} = \begin{pmatrix} \sigma^2_{\theta_1} & \sigma_{\theta_1\theta_2} \\ \sigma_{\theta_1\theta_2} & \sigma^2_{\theta_2} \end{pmatrix}. \tag{9.13}$$

From (9.13), the correlation between overdispersion parameters in the Poisson and binomial outcome is obtained from the covariance matrix estimates, $\mathbf{D}_{\boldsymbol{\Theta}}$. Note that the case for which D_Θ is a 2×2 diagonal matrix implies that the models specified in (9.12) and (9.13) are identical.

9.3 Application to the Data

For each of the two case studies, three types of models were considered: (1) basic hierarchical Bayesian binomial-Poisson model (BP) without overdispersion, (2) BP model with additive overdispersion and (3) BP model with multiplicative overdispersion. Markov Chain Monte Carlo (MCMC) simulation was used to estimate the unknown parameters via the `rjags` package (Plummer, 2015). For each model, three chains each with 60000 iterations from which 30000 were considered as burn-in period, were used. Diagnostic analysis (Gelman and Rubin, 1992) for the parameters of interest (presented in the appendix) indicates convergence for all parameters.

Table 9.1: DIC fit statistics for different overdispersion models. The smaller the value, the better the model fits the data (the "best" model based on DIC is shown in bold). BP: basic hierarchical Bayesian binomial-Poisson model.

		DIC	
	Model	Parallel	Crossover
No OD	BP	1070.96	1493.19
Multiplicative OD	BP+OD(Poisson)	1052	1445
	BP+OD(Binomial)	993.9	1495
	BP+OD(Both)	971.9	1445
Additive OD	BP+OD(Poisson)	1052	**1444**
	BP+OD(Binomial)	990.3	1496
	BP+OD(Both; independent)	967.8	**1444**
	BP+OD(Both; correlated)	**900.3**	1456

The Deviance Information Criterion (DIC, Spiegelhalter *et al.*, 2002) for assessing the goodness-of-fit for the models is shown in Table 9.1. In both case studies, models that accounted for overdispersion performed better than the basic hierarchical Bayesian binomial-Poisson model (BP) without overdispersion. The model with additive correlated overdispersion parameters (denoted as BP + OD(Both; correlated)) is the one with the best goodness-of-fit for the parallel design. For the crossover study, the additive model that adjust for overdispersion in the Poisson outcome only (denoted as BP + OD(Poisson)) and the additive model that assumes independent overdispersion random effects in both outcomes (denoted as BP + OD(Both; independent)) provide better model fit compared to the BP model ignoring overdispersion. In what follows, we present results of the best fitting overdispersion models and models without overdispersion for comparison purposes (additional output for alternative overdispersion models is presented and discussed in the appendix).

9.3.1 DRL-72 Experiment with Parallel Design

Posterior Means for Log(Odds Ratio) and Log(Relative Intensity)

For the binomial outcome, the posterior means for the log odds ratio, denoted as log(OR), are presented in Table 9.2 for the BP and BP + OD(Both; correlated) models. The posterior log(OR) estimates for the two models are similar, although the standard errors for the overdispersion model were larger. This is expected due to the additional variability introduced in the model in estimating the overdispersion parameters. Based on the 95% credible intervals for the BP model, we can conclude that 5.0 mg/kg was the effective dose, while 10.0 mg/kg was the only effective dose using the additive overdispersion model.

For the Poisson outcome, the posterior means of the log relative intensity, denoted as log(RI), are presented in Table 9.3. The posterior mean estimates for both models were found to be relatively similar although standard errors for the overdispersion model were

Table 9.2: Binomial outcome-parallel design. Posterior estimates of log(OR) for the contrasts of each dose level against the zero dose. The estimate shown for the zero dose is the log odd of success versus failure for the zero dose, while for the other doses, the estimate is a contrast of the log odds for the given dose versus the zero dose. SE: Posterior mean of the standard error of the mean. CI: credible interval.

Model	Dose	Estimate (SE)	95% CI
BP	0	-1.446 (0.135)	(-1.711, -1.18)
DIC=1070.96	2.5	-0.078 (0.108)	(-0.29, 0.132)
	5	0.276 (0.105)	(0.07, 0.482)
	10	0.763 (0.103)	(0.563, 0.964)
BP+OD(Add;Correlated)	0	-1.471 (0.178)	(-1.823, -1.124)
DIC=900.30	2.5	-0.09 (0.229)	(-0.539, 0.363)
	5	0.333 (0.231)	(-0.124, 0.79)
	10	0.851 (0.232)	(0.395, 1.31)

Table 9.3: Poison outcome-parallel design. Posterior mean estimates for the contrasts of the log(RI) with the zero dose. The estimate for zero dose is the log of the risk for zero dose while for other doses, the estimate denotes the contrast of log risk of the given dose versus the zero dose. 95% credible intervals (CI) are also included.

Model	Dose	Estimate (SE)	95% CI
BP	0	4.171 (0.042)	(4.086, 4.254)
DIC=1070.96	2.5	0.005 (0.041)	(-0.076, 0.086)
	5	-0.071 (0.042)	(-0.153, 0.011)
	10	-0.141 (0.043)	(-0.225, -0.058)
BP+OD(Add;Correlated)	0	4.169 (0.055)	(4.06, 4.277)
DIC=900.30	2.5	0.007 (0.072)	(-0.136, 0.15)
	5	-0.075 (0.074)	(-0.219, 0.071)
	10	-0.155 (0.075)	(-0.302, -0.008)

higher. Moreover, the 95% credible intervals for both models infer an effect of only the high dose.

A graphical summary of the estimated dose effects is presented in Figure 10.10, which shows that log(OR) increases with dose, while log(RI) decreases with dose. This implies that as the dose administered increases, the odds of successfully obtaining a reward increases since the risk of pressing the lever before 72 seconds have elapsed decreases.

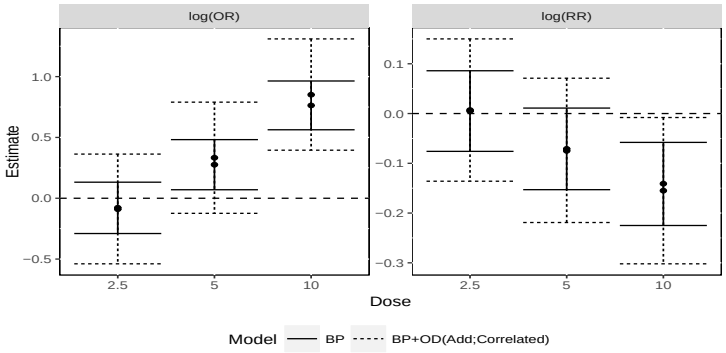


Figure 9.1: Parallel DRL-72 experiment. Posterior mean estimates with the 95% credible intervals (CI) for the response and reward models with no overdispersion (BP) and correlated additive overdispersion model.

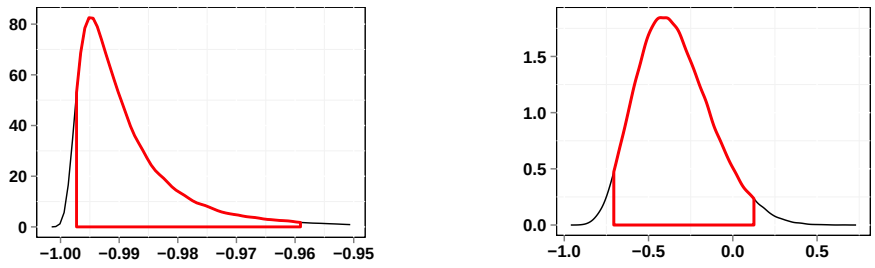
Posterior Means for the Correlation Between the Number of Responses and Rewards

As mentioned in Section 9.2, the correlation between the subject-specific random effects, given in (9.6) is of primary interest since it quantifies the correlation between the number of responses and rewards at an individual level. For example, the case in which $\rho_{ab} < 0$ implies that the number of rewards increases as the number of responses decreases. That is, if the rat become more relaxed, it has a tendency to wait for at least 72 seconds, which results in a lower number of times it presses the lever (responses). Subsequently, the rat will be rewarded more times.

Table 9.4 (upper panel) presents the posterior mean for the clustering correlation coefficient while Figure 9.2 shows the density estimates for the posterior distributions obtained for the BP and BP + OD(Both; correlated) models. We notice that there is a substantial difference between the posterior means of the correlation coefficients for the two models. For the BP model, $\rho_{ab} = -0.987(-0.998, -0.958)$, while for the model with correlated additive overdispersion parameters, $\rho_{ab} = -0.365(-0.712, 0.122)$. The latter implies that the correlation between a_i and b_i diminishes when overdispersion parameters are included in the model. DIC for the model that assumes that the two random effects are independent is equal to 967.8 (see Table 9.1) and therefore this model will not be considered further. For the BP model, the two random effects are highly correlated, suggesting that a shared random effect can be considered. However, for a shared random effects model (not shown here), DIC$=1240.77$. Note that for the model with correlated additive overdispersion parameters, the posterior mean estimates (shown in the lower panel of Table 9.4) denote the components of the covariance matrix (9.13).

Table 9.4: Parallel design. Posterior mean (median) for clustering random effects variance and correlation ρ_{ab}, overdispersion parameters and overdispersion correlation ρ_{OD} for the binomial-Poisson (BP) models under different assumptions.

Clustering random effects			
	σ_a^2	σ_b^2	ρ_{ab}
BP	0.017 (0.016)	0.227 (0.209)	-0.988 (-0.987)
BP+OD(Both; Correlated)	0.084 (0.077)	0.298 (0.268)	-0.345 (-0.365)
Overdispersion random effects			
	θ_1	θ_2	ρ_{OD}
BP+OD(Both; Correlated)	0.027 (0.026)	0.265 (0.256)	-0.990 (-0.993)



(a) BP model **(b)** BP+OD (Both; Correlated)

Figure 9.2: Parallel DRL-72 experiment: Kernel density estimate for the clustering random effects correlation. The vertical lines marks the 95% credible intervals. This correlation captures the dependency between the number of responses and number of rewards.

The posterior means for the number of response and rewards are shown in Figure 9.3, which reveals the same inverse relationship between the two outcomes for both models. For the 5.0 mg/kg and 10 mg/kg doses, the number of responses decrease while the number of rewards increases.

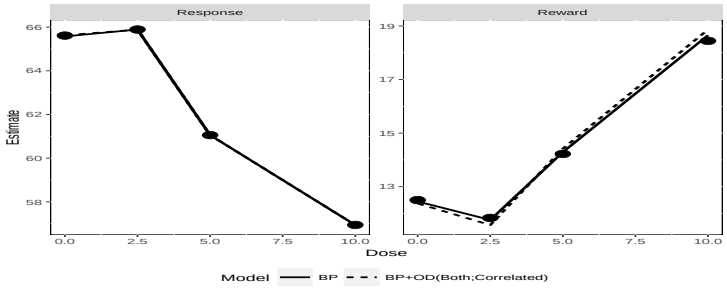


Figure 9.3: Parallel DRL-72 experiment. Posterior estimate of the average profiles for the response and reward models with no overdispersion and with correlated additive overdispersion parameters. The solid symbols denotes the observed data.

9.3.2 DRL-72 Experiment with a Crossover Design

For the crossover study, models accounting for overdispersion in both the Poisson and binomial outcomes had a smaller DIC compared to a model without overdispersion parameters. In addition, for the additive overdispersion model, independent overdispersion parameters were preferred over correlated overdispersion parameters. Note that a model with only additive overdispersion random effect in the Poisson outcome had the same DIC as the independent additive overdispersion model. Moreover, the multiplicative overdispersion counterparts of these two models had similar DIC values (DIC=1445, see Table 9.1). In what follows, we discuss the results of the hierarchical Bayesian Binomial-Poisson model (BP), BP with an additive overdipersion parameter in the Poisson outcome (BP+OD(Add, Poisson)) and BP with independent additive overdipersion parameters in both outcomes (BP + OD(Both; add_ind)).

Posterior Estimates of the Dose Effects for Resposes and Rewards

For the Poisson outcome, the posterior mean of log(RI) with the corresponding 95% credible intervals for the BP, BP+OD(Poisson) and BP + OD(Both; add_ind) models are presented in Table 9.5. Posterior mean estimates for the dose effect revealed the same pattern in the three models. The independent additive overdispersion model had higher standard error estimates compared to the BP model. Inference based on the 95% credible intervals for the dose effect was consistent across the two models. In particular, as can be seen in Figure 9.4 (right panel), the 95% credible intervals of all contrasts of the active doses with the control dose do not include the zero line, indicating a reduction in the log(RI) with increasing dose levels. This implies that rats had a tendency to calm down and press the lever fewer times as the administered dose of the treatment increased.

The posterior estimates for the log(OR) presented in Table 9.6 and Figure 9.4 (left panel), reveal the same pattern for both models. Standard errors for the parameters were larger for the model with overdispersion compared to those of the basic model. The 95% credible intervals for the two models suggested a non-zero effect of the 5.0 mg/kg and 10.0 mg/kg doses compared to the control dose. In particular, the odds of obtaining a reward were not different from those obtained under the control dose for 1.25 mg/kg and

Table 9.5: Poisson outcome. Posterior mean estimates for the contrasts of the log(RI) with the placebo.

Model	Dose	Estimate (SE)	95% CI
BP	1.25	-0.187 (0.044)	(-0.274, -0.103)
	2.5	-0.167 (0.043)	(-0.251, -0.084)
	5	-0.331 (0.046)	(-0.424, -0.242)
	10	-0.329 (0.043)	(-0.414, -0.246)
BP+OD(Add, Poisson)	1.25	-0.19 (0.065)	(-0.316, -0.062)
	2.5	-0.146 (0.067)	(-0.277, -0.012)
	5	-0.322 (0.068)	(-0.449, -0.185)
	10	-0.333 (0.063)	(-0.454, -0.205)
BP+OD(Both; add_ind)	1.25	-0.19 (0.064)	(-0.316, -0.064)
	2.5	-0.148 (0.065)	(-0.277, -0.02)
	5	-0.321 (0.067)	(-0.451, -0.189)
	10	-0.334 (0.062)	(-0.456, -0.211)

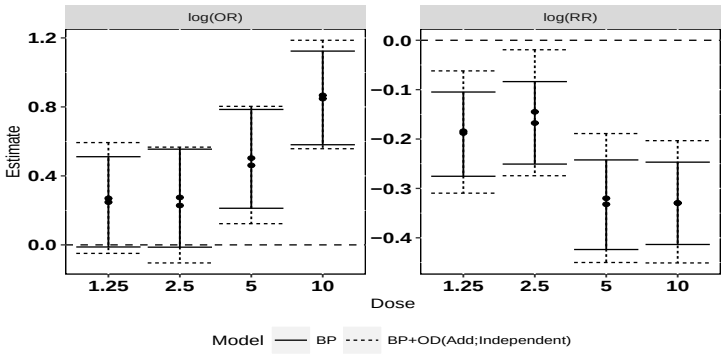


Figure 9.4: Crossover DRL-72 experiment. Posterior mean estimates with the 95% credible intervals for log(RI) and log(OR).

Table 9.6: Binomial outcome. Posterior estimates of the log(OR) for the contrasts of each dose level against the zero dose.

Model	Dose	Estimate (SE)	95% CI
BP	1.25	0.250 (0.134)	(-0.008, 0.518)
	2.5	0.276 (0.146)	(-0.017, 0.556)
	5	0.503 (0.145)	(0.216, 0.786)
	10	0.847 (0.140)	(0.575, 1.125)
BP+OD(Add,Poisson)	1.25	0.26 (0.159)	(-0.05, 0.575)
	2.5	0.217 (0.167)	(-0.121, 0.54)
	5	0.457 (0.166)	(0.125, 0.782)
	10	0.862 (0.146)	(0.576, 1.143)
BP+OD(Both; add_ind)	1.25	0.276 (0.166)	(-0.049, 0.604)
	2.5	0.228 (0.174)	(-0.113, 0.568)
	5	0.464 (0.173)	(0.124, 0.804)
	10	0.869 (0.158)	(0.562, 1.181)

2.5 mg/kg while for higher doses, the odds increased compared to those under the control dose.

Posterior Means of the Random Effects

Table 9.7 present the posterior estimates of the variance components and the correlation for the subject-specific random effects and the variance of the overdispersion parameters. The posterior correlation obtained for the two models (BP: -0.48, BP + OD(Both; add_ind): -0.462) and the density estimate for the posterior distributions presented in Figure 9.5 indicates a negative association between the number of rewards and responses at an individual level. The average predictions for the number of times rats press the lever and the number of rewards obtained are presented in Figure 9.6 together with a plot of the observed means. The posterior means obtained for the two models are similar.

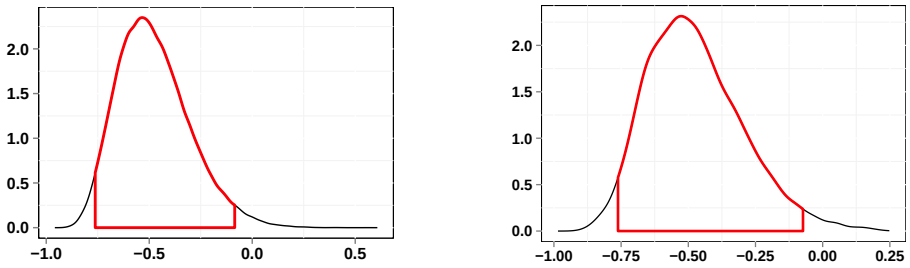
9.4 Discussion

The DRL-72 is a standard drug development behavioural experiment in which the two endpoints of interest, the number of responses and rewards are assumed to be Poisson and binomial random variables, respectively. In this chapter, a parallel and a crossover design DRL-72 experiment were discussed. The issues posed by these case studies were addressed using hierarchical Bayesian joint binomial-Poisson models. The random effects in the model capture the heterogeneity between rats in their reaction towards the administered dose. Moreover, by assuming a multivariate normal distribution on the random effects, the two outcomes were jointly modelled and the association between them captured by the correlation between the random effects.

Table 9.7: Crossover design. Posterior mean (median) for clustering and overdispersion random effects covariance estimates for the binomial-Poisson (BP) models under different assumptions.

	Clustering random effects		
	σ_a^2	σ_b^2	ρ_{ab}
BP	0.112 (0.105)	0.292 (0.270)	-0.480 (-0.500)
BP+OD(Add; Poisson)	0.112 (0.102)	0.296 (0.273)	-0.476 (-0.482)
BP+OD(Both; Add_ind)	0.110 (0.102)	0.284 (0.261)	-0.463 (-0.482)

	Overdispersion random effects	
	$\sigma_{\theta_1}^2$	$\sigma_{\theta_2}^2$
BP+OD(Add, Poisson)	0.010 (0.010)	
BP+OD(Both; Add_ind)	0.010 (0.010)	0.011 (0.007)



(a) BP model **(b)** BP+OD (Both: independent)

Figure 9.5: Crossover DRL-72 experiment. Kernel density estimate for the clustering random effects correlation. The vertical lines marks the 95% credible intervals. This correlation captures the dependency between the number of responses and number of rewards at an individual level.

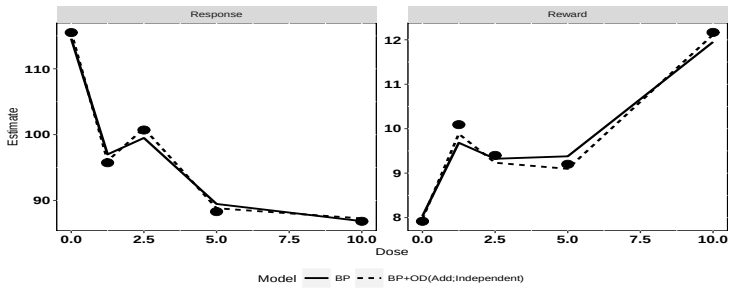


Figure 9.6: Crossover DRL-72 experiment. Observed and predicted average number of responses and rewards for the BP and independent additive overdispersion random effects models. The solid symbol denotes the observed data.

In many practical situations, the variability in the data is larger than that which could be expected in a standard binomial or Poisson distribution. To adequately address these potential pitfalls, we extended the classical hierarchical Bayesian binomial-Poisson models for the analysis of DRL-72 experiments, to the combined modelling framework. Specifically, we proposed to introduce observation-specific random effects, either at the mean parameters (multiplicative) or in the linear (additive) part of the Bayesian models, in order to capture the extra Poisson or binomial variability in both case studies.

In the presence of overdispersion, the variability attributed to clustering is different than that observed in a model without overdispersion random effects. For instance, in the parallel design experiment, the clustering correlation coefficient $\rho_{ab} = -0.988$ in a model not accounting for overdispersion, while $\rho_{ab} = -0.345$ in a model with overdispersion random effects. Therefore, by incorporating the overdispersion random effects, the association is decomposed into subject-level (ρ_{ab}) and observation-level ρ_{OD} correlation. In case there is no extra-Poisson or extra-binomial variability, the impact on the subject-level association (ρ_{ab}) is minimal as was seen in the crossover DRL-72 experiment. Note that the impact of overdispersion on the posterior means is minimal, although inference based on models ignoring overdispersion may be invalid since presence of overdispersion has an impact on the standard errors of the parameter estimates.

In conclusion, combined models are a flexible tool for accounting for over/under-dispersion in analysis involving Poisson or binomial outcomes. In the joint modelling context, the flexibility allows us to specify overdispersion random effects in either one or both outcomes. In cases where there is no overdispersion, the inclusion of the overdispersion parameters may still improve the individual-level predictions.

9.5 Appendix: Model Diagnostics

9.5.1 DRL-72 Experiment with a Parallel Design

The trace plots shown in Figure 9.7 for the basic hierarchical Bayesian binomial- Poisson model indicates proper mixing of the three chains of the dose parameters in the Poisson (Panel a) and binomial (Panel b) outcomes. Formal diagnostic measures presented in Table 9.8 for the BP model and Table 9.9 for the model with additive correlated overdispersion parameters all indicated attainment of acceptable convergence. Trace plots for the model with additive overdispersion parameters for the parallel design experiment are presented in Figure 9.8.

9.5.2 DRL-72 Experiment with Crossover Design

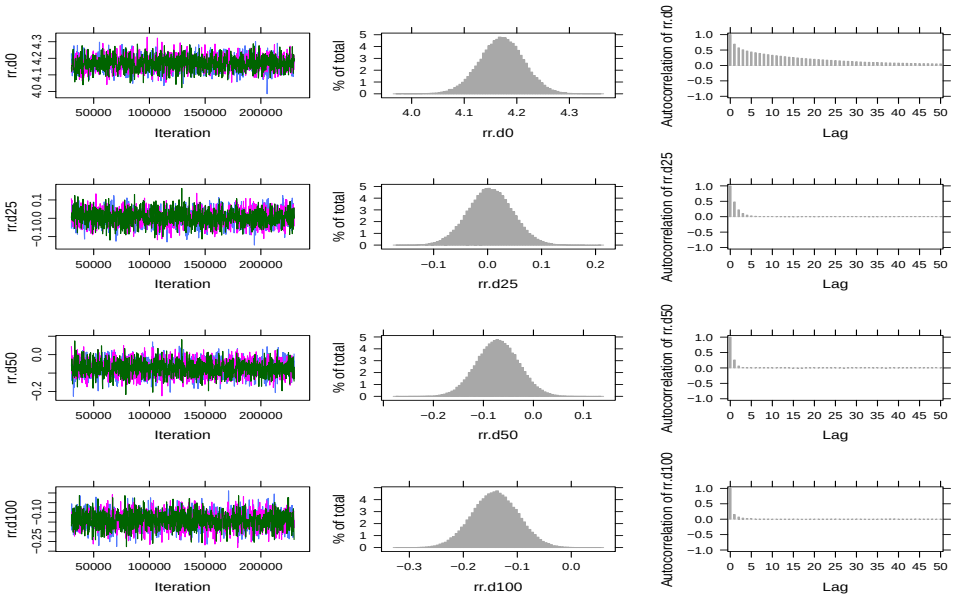
Figure 9.9 and 9.10 presents the diagnostic plots for dose parameters in the final crossover models discussed in the chapter. All the parameters of interest attained good chain mixing and the autocorrelation was negligible by lag 10. Formal diagnostic checks including the Gelman and Rubin test were performed, results of which are presented in

Table 9.8: BP model. Diagnostic measures of goodness of fit for the dose contrast parameters. MCerr: Monte Carlo error. AC.200: autocorrelation at lag 200.

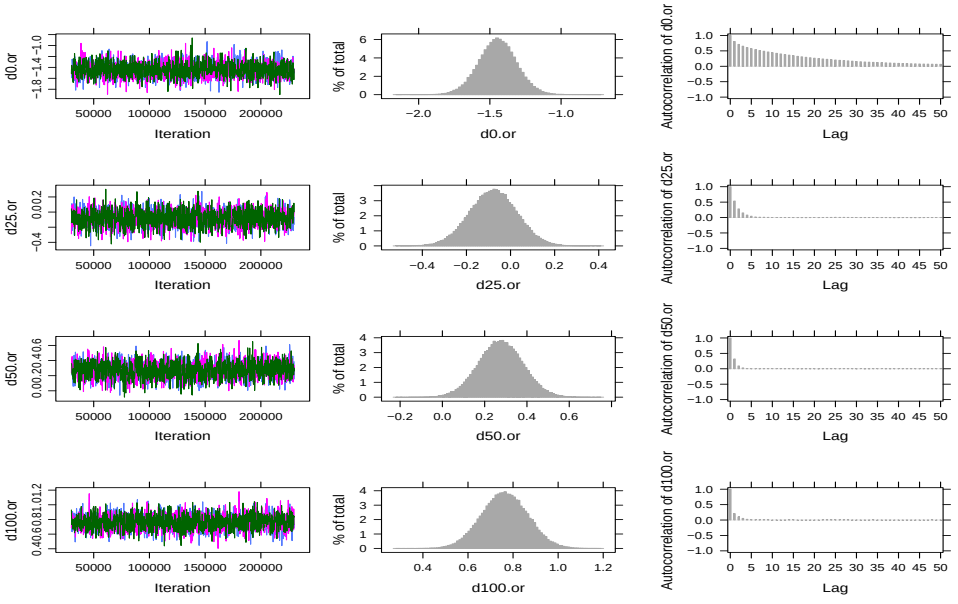
	Dose	MCerr	AC.200	Gelman-Rubin statistic
log(OR)	0	0.002	0.092	1.003
	2.5	0.001	-0.01	1
	5	0.001	-0.001	1
	10	0.001	-0.003	1
log(RI)	0	0.001	0.073	1.002
	2.5	0	-0.001	1
	5	0	-0.01	1
	10	0	-0.008	1
Clustering	σ_a^2	0	0.033	1
	ρ_{ab}	0	0.004	1.001
	σ_b^2	0.001	0.013	1

Table 9.9: BP +OD(Both; correlated) additive model. Diagnostic measures of goodness of fit for the dose contrast parameters. MCerr: Monte Carlo error. AC.200: autocorrelation at lag 200.

	Dose	MCerr	AC.200	Gelman-Rubin statistic
log(OR)	0	0.003	0.109	1
	2.5	0.004	0.115	1
	5	0.004	0.093	1.001
	10	0.004	0.082	1.001
log(RI)	0	0.001	0.087	1
	2.5	0.001	0.094	1
	5	0.001	0.073	1
	10	0.001	0.066	1.001
Clustering	σ_a^2	0	0.699	1.015
	ρ_{ab}	0.026	0.551	1.013
	σ_b^2	0.006	0.828	1.019
OD	ρ_{OD}	0	0.103	1.005
	$\sigma^2_{\theta_2}$	0.006	0.481	1.014
	$\sigma^2_{\theta_1}$	0	0.381	1.005

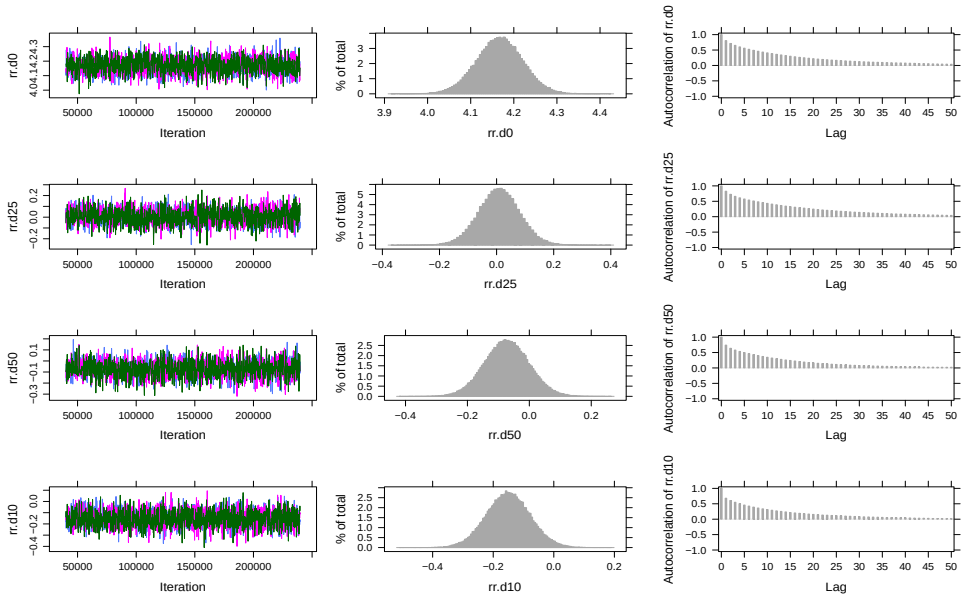


(a) Poisson outcome: log(relative intensity).

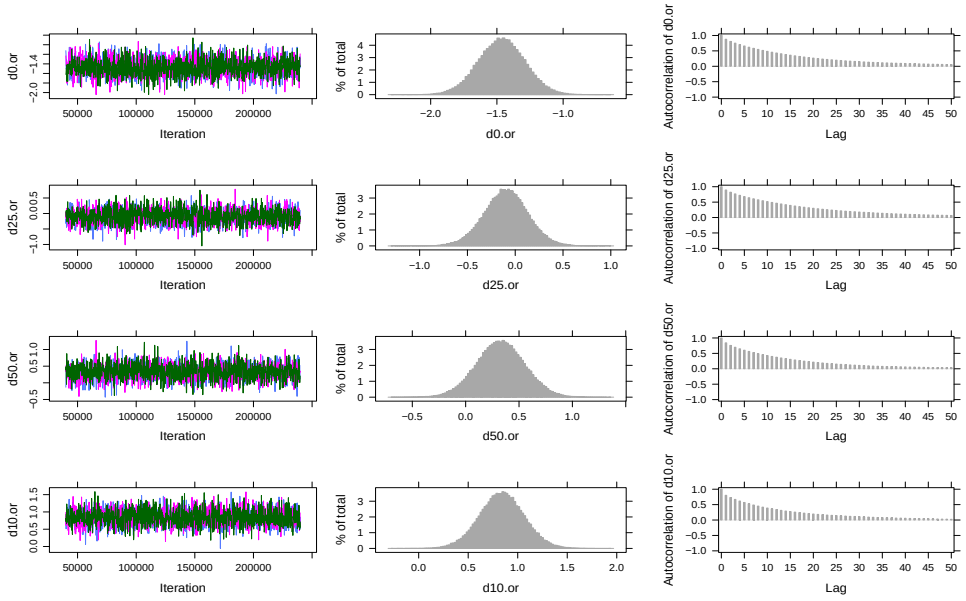


(b) Binomial outcome: log(odds ratio).

Figure 9.7: Parallel design DRL-72 experiment. Diagnostic plots for dose parameters in the basic model without overdispersion.



(a) Poisson outcome: log(relative intensity).



(b) Binomial outcome: log(odds ratio).

Figure 9.8: Parallel design with additive overdispersion. BP+OD (Both; correlated) diagnostic plots for dose parameters.

Table 9.10: BP model. Diagnostic measures of goodness of fit for the dose contrast parameters. MCerr: Monte Carlo error. AC.200: autocorrelation at lag 200.

	Dose	MCerr	AC.200	Gelman-Rubin statistic
log(RI)	1.25	0	-0.007	1
	2.5	0	0.001	1
	5.0	0	-0.003	1
	10	0	0	1
log(OR)	1.25	0.001	0.003	1
	2.5	0.001	0.004	1
	5.0	0.001	-0.001	1
	10	0.001	-0.005	1
Clustering	σ_a^2	0	0.011	1
	σ_b^2	0.001	0.004	1
	ρ	0.001	0.002	1

Table 9.10 and 9.11, all of which indicate good convergence for the parameters of interest.

In the crossover case study, a model with overdispersion in only the Poisson outcome had the same DIC as a model with independent additive overdispersion parameters. Moreover, the models with multiplicative overdispersion parameters in either the Poisson outcome or both outcomes did not significantly differ in DIC from these two additive overdispersion models. In Table 9.12 and 9.13

Table 9.11: BP +OD(Both; independent) additive model. Diagnostic measures of goodness of fit for the dose contrast parameters. MCerr: Monte Carlo error. AC.160: autocorrelation at lag 160.

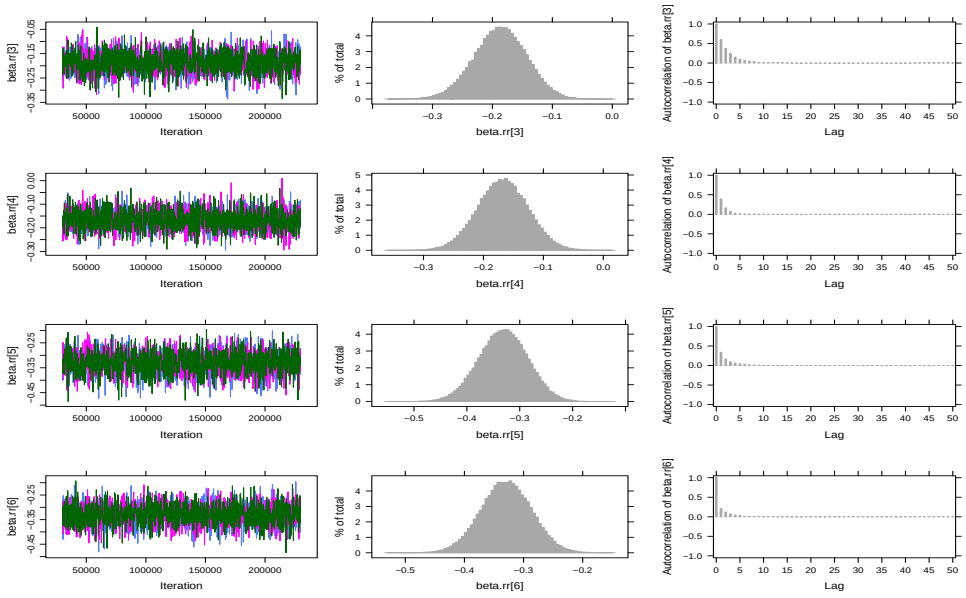
	Dose	MCerr	AC.160	Gelman-Rubin statistics
log(RI)	1.25	0	0.004	1
	2.5	0	0	1
	5	0	0.005	1
	10	0	0.005	1
log(OR)	1.25	0.001	-0.002	1
	2.5	0.001	-0.005	1
	5	0.001	-0.003	1
	10	0.001	0	1
Clustering	σ_a^2	0	0.028	1
	σ_b^2	0.001	0.012	1
	ρ	0.001	0.007	1
OD	$\sigma_{\theta_1}^2$	0	-0.002	1
	$\sigma_{\theta_2}^2$	0	0.307	1.003

Table 9.12: Multiplicative model for the binomial outcome. Posterior mean estimates of log(OR).

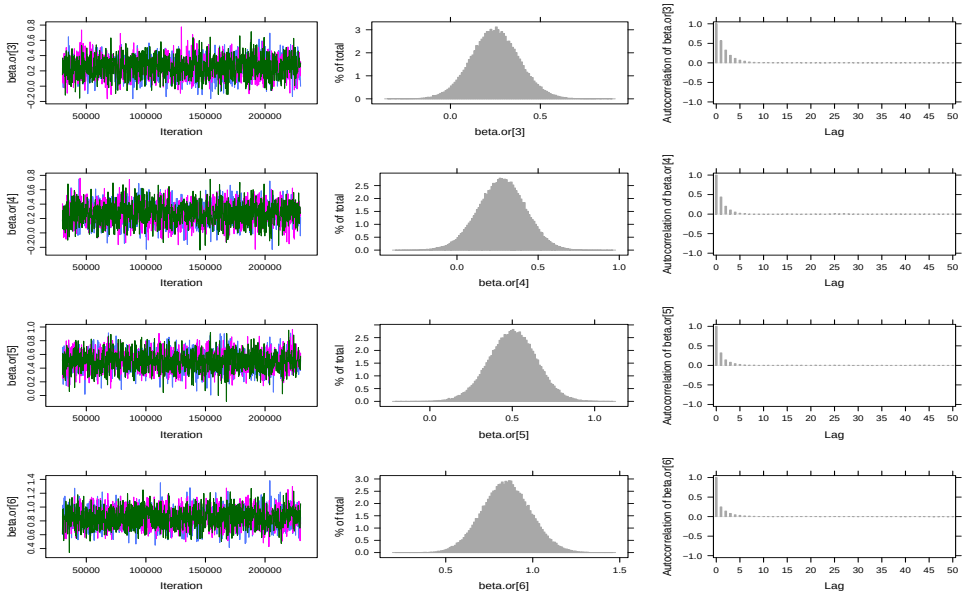
Model	Dose	Estimate (SE)	95% CI
BP+OD(Mult,Poisson)	1.25	0.269 (0.159)	(-0.04, 0.579)
DIC=1445	2.5	0.227 (0.166)	(-0.101, 0.552)
	5	0.466 (0.165)	(0.143, 0.789)
	10	0.87 (0.15)	(0.576, 1.164)
BP+OD(Mult,Both)	1.25	0.322 (0.178)	(-0.027, 0.672)
DIC=1445	2.5	0.276 (0.186)	(-0.088, 0.64)
	5	0.519 (0.186)	(0.156, 0.885)
	10	0.991 (0.171)	(0.658, 1.327)

Table 9.13: Multiplicative model Poisson outcome. Posterior mean estimates of log(RI).

Model	Dose	Estimate (SE)	95% CI
BP+OD(Mult,Poisson)	1.25	-0.186 (0.07)	(-0.323, -0.047)
DIC=1445	2.5	-0.149 (0.071)	(-0.288, -0.01)
	5	-0.315 (0.073)	(-0.457, -0.17)
	10	-0.331 (0.068)	(-0.464, -0.197)
BP+OD(Mult,Both)	1.25	-0.192 (0.063)	(-0.314, -0.067)
DIC=1445	2.5	-0.151 (0.064)	(-0.275, -0.025)
	5	-0.325 (0.066)	(-0.453, -0.195)
	10	-0.334 (0.061)	(-0.453, -0.213)

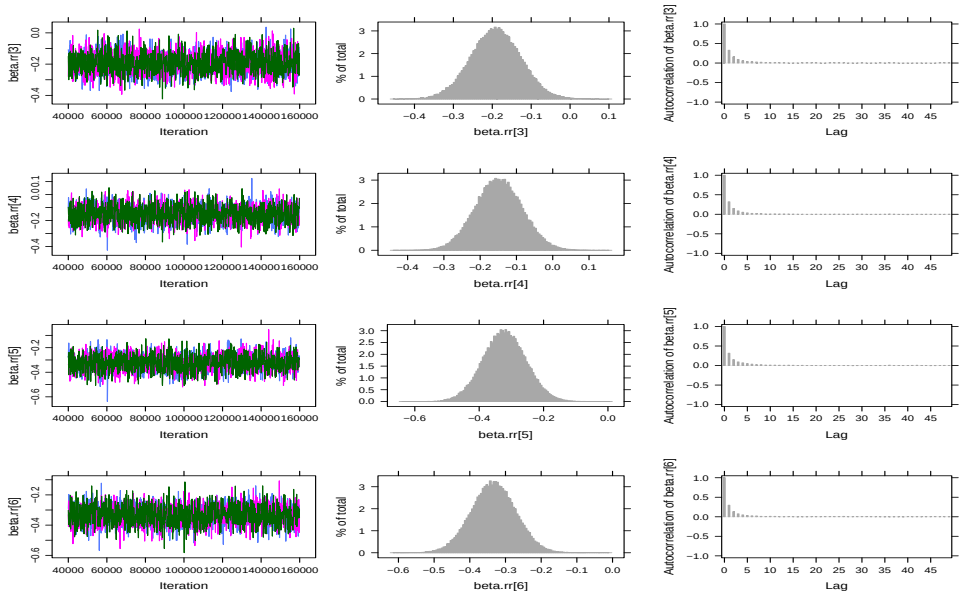


(a) Poisson outcome: log(relative intensity).

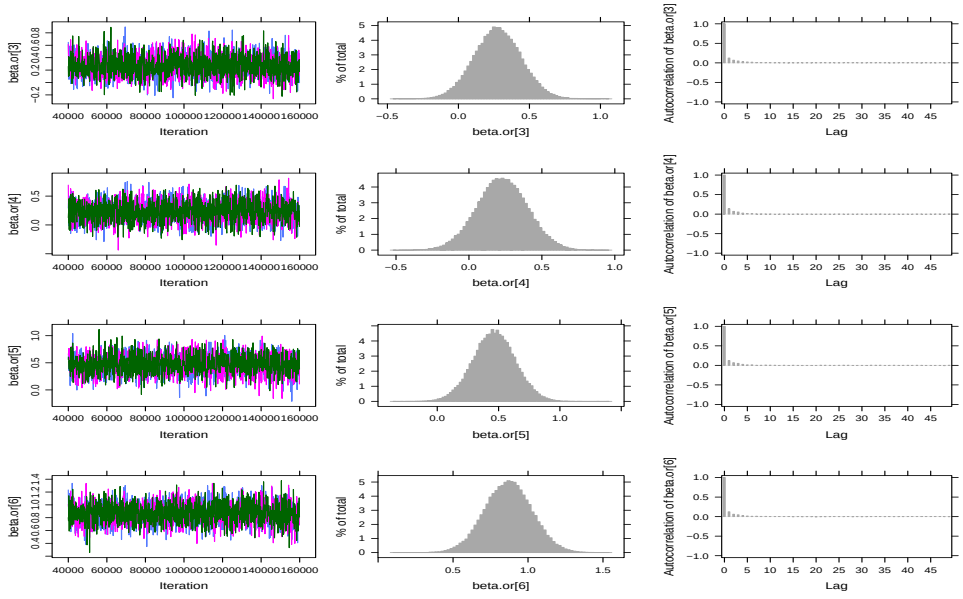


(b) Binomial outcome: log(odds ratio)

Figure 9.9: Crossover design DRL-72 experiment. BP model diagnostic plots for dose parameters (1.25, 2.5, 5.0 10.0) mg/kg, respectively, for row 1-4, in each panel.



(a) Poisson outcome: log(relative intensity).



(b) Binomial outcome: log(odds ratio).

Figure 9.10: Crossover design DRL-72 experiment with additive overdispersion. BP+OD (Both; ind) diagnostic plots for dose parameters.

Chapter 10

Bayesian Variable Selection: Unrestricted Dose-Response Models

10.1 Introduction

In Chapter 9, we presented an analysis of two dose-response experiments, using models which took into account possible overdesperion but did not account for model uncertainty. This chapter extends the models presented in Chapter 9 by including the Bayesian variable selection component in the models. In particular, for both case studies, we define a set of plausible dose-response profiles where no prior assumption on the monotonicity of the profiles is taken (hereby referred to unrestricted dose-response analysis). This results in a large set of models from which the most plausible model given the data is to be selected.

When the competing models are nested, the choice between the models reduces to a choice of a subset of predictors. Bayesian Variable Selection (BVS: George and McCulloch, 1993, O'Hara and Sillanpaa, 2009) allows for a probabilistic selection of parameters into the model using Gibbs sampling (Dellaportas *et al.*, 2002) whereby, parameters with higher posterior inclusion probability are incorporated in the model. Lin *et al.* (2012) and Otava *et al.* (2013) applied the BVS methodology to dose-response microarray data, while Whitney and Ryan (2009) used the methodology to account for uncertainty due to covariate adjustment in a dose-response bioassay experiment.

In this chapter, we apply the BVS model mentioned above to the two case studies consisting of a behavioural experiment aimed at testing for clinically active anti-depressant compounds. The chapter is arranged as follows. The statistical methodology is presented in Section 10.2, while results of the analysis are presented in Section 10.3. Finally, Section 10.4 presents concluding remarks.

10.2 Bayesian Variable Selection Models for the DRL-72 Experiment

10.2.1 Joint Model for the Responses and Rewards in the DRL-72 Experiment

In this section, we briefly present the joint model for the DRL-72 study. Let n_{ij} be the number of times that the ith rat presses the lever under dose j (i.e. the number of responses) assumed to follow a Poisson distribution, while the number of rewards (the number of times that the rat waits 72 seconds before pressing the lever), Y_{ij}, obtained under dose j follows a binomial distribution. That is,

$$\begin{aligned} Y_{ij} &\sim Binomial(n_{ij}, \pi_{ij}), \\ n_{ij} &\sim Poisson(\lambda_{ij}). \end{aligned} \tag{10.1}$$

The linear predictors for the mean response, γ_{ij}, and reward π_{ij} are, respectively, given by

$$\begin{aligned} \text{log}(\lambda_{ij}) &= \mathbf{X}_i\boldsymbol{\alpha} + a_i, \\ \text{logit}(\pi_{ij}) &= \mathbf{X}_i\boldsymbol{\beta} + b_i. \end{aligned} \tag{10.2}$$

Here, $\mathbf{X}_i$ is an $n_i \times p$ design matrix, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $p \times 1$ vectors of coefficients, while a_i and b_i are the rat-specific intercepts for response and rewards, respectively.

In order to take into account the differences in design between the two case studies, different specifications for the right hand side of (10.2) is needed. The complete model formulation is given in Section 9.2. Compared to the mean structure discussed in Chapter 9, the BVS implementation requires reparametrization of the dose parameters in the model. The configuration of the design matrix is discussed in Section 10.2.2 and 10.2.3 for the mean structure in the parallel and crossover DRL-72 case studies, respectively.

10.2.2 A BVS Model for a DRL-72 Experiment with Parallel Design: Mean Structure

Let $\mathbf{X}_i$ be an $a \times a$ design matrix and let $(\gamma_0, \gamma_1, \gamma_2, \gamma_3)$ be the parameter vector representing the dose effects on the number of responses, while $(\delta_0, \delta_1, \delta_2, \delta_3)$ are the dose effects on the number of rewards. Note that, δ_j and γ_j are defined as the additional effects of dose j, $\quad j = 1, 2, 3$, in the model for response and reward, respectively, while γ_0 and δ_0 are the placebo effects. The design matrix $\mathbf{X}$ is given by

$$\mathbf{X} = \left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \\ \hline 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ \hline 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 0 \\ \hline 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 \end{array}\right). \tag{10.3}$$

Hence, the linear predictor for the responses is given by

$$\log\left(\lambda_{ij}\right) = \gamma_0 + \sum_{j=1}^{3} \gamma_j X_{ij} + a_i. \tag{10.4}$$

Which implies that, the dose effects are given by

$$\begin{array}{ll} \alpha_0 = \gamma_0, & \text{if dose} = 0.0\text{mg/kg}, \\ \alpha_1 = \gamma_0 + \gamma_1, & \text{if dose} = 2.5\text{mg/kg}, \\ \alpha_2 = \gamma_0 + \gamma_1 + \gamma_2, & \text{if dose} = 5.0\text{mg/kg}, \\ \alpha_3 = \gamma_0 + \gamma_1 + \gamma_2 + \gamma_3, & \text{if dose} = 10.0\text{mg/kg}. \end{array} \tag{10.5}$$

The linear predictor for the number of rewards can be formulated in the same way, from which, the effect of the jth dose is given by

$$\beta_j = \sum_{\ell=0}^{j} \delta_\ell, \;\; j = 0, 1, 2, 3. \tag{10.6}$$

In order to formulate the DRL-72 model as a Bayesian variable selection model, two random variables are added to the linear predictor. Let $\mathbf{W}$ and $\mathbf{Z}$ be vectors of indicator variables such that,

$$\begin{array}{l} w_j = \left\{ \begin{array}{ll} 1, \text{ if } \gamma_j & \text{is included in the model,} \\ 0, \text{ if } \gamma_j & \text{is not included in the model.} \end{array} \right. \\ \text{and} \\ z_j = \left\{ \begin{array}{ll} 1, \text{ if } \delta_j & \text{is included in the model,} \\ 0, \text{ if } \delta_j & \text{is not included in the model.} \end{array} \right. \end{array} \tag{10.7}$$

Table 10.1: Overview of the model configuration, number of parameters in each model and the number of models for each configuration in the BVS model for the number of responses. Similar output for the rewards model can be obtained.

$\mathbf{W}$	Parameters	Number of possible models
0 0 0	γ_0	1
1 0 0	γ_0, γ_1	2
0 1 0	γ_0, γ_2	2
0 0 1	γ_0, γ_3	2
1 1 0	$\gamma_0, \gamma_1, \gamma_2$	4
1 0 1	$\gamma_0, \gamma_1, \gamma_3$	4
0 1 1	$\gamma_0, \gamma_2, \gamma_3$	4
1 1 1	$\gamma_0, \gamma_1, \gamma_2, \gamma_3$	8

The linear predictor for the number of responses and reward, respectively, is given by,

$$\begin{aligned} \text{log}(\lambda_{ij}) &= \gamma_0 + \sum_{j=1}^{3} w_j \gamma_j X_{ij} + a_i, \\ \text{logit}(\pi_{ij}) &= \delta_0 + \sum_{j=1}^{3} z_j \delta_j X_{ij} + b_i. \end{aligned} \tag{10.8}$$

For the number of responses, in a study with four dose levels,

$$\mathbf{W} = (w_1, w_2, w_3) \text{ and } \alpha_j = \gamma_0 + \sum_{j=1}^{3} w_j \gamma_j. \tag{10.9}$$

Which implies that,

$$\begin{array}{ll} \alpha_0 = \gamma_0, & \text{if dose} = 0.0\text{mg/kg}, \\ \alpha_1 = \gamma_0 + w_1\gamma_1, & \text{if dose} = 2.5\text{mg/kg}, \\ \alpha_2 = \gamma_0 + w_1\gamma_1 + w_2\gamma_2, & \text{if dose} = 5.0\text{mg/kg}, \\ \alpha_3 = \gamma_0 + w_1\gamma_1 + w_2\gamma_2 + w_3\gamma_3, & \text{if dose} = 10.0\text{mg/kg}. \end{array} \tag{10.10}$$

The mean structure in the model for the number of rewards can be formulated in the same way. Note that, each configuration of the indicator variables $\mathbf{W} = (w_1, w_2, w_3)$, implies different number of parameters. Figure 10.1 shows an illustration of the null model that assumes no dose effect. For this model, $(\mathbf{W} = (0, 0, 0))$. Figure 10.2 provides an illustration for the case where $\mathbf{W} = (1, 0, 0)$. For this configuration of $\mathbf{W}$, there are two possible models. For the configuration $\mathbf{W} = (1, 1, 0)$, the four models are shown in Figure 10.3. For the four-doses study, an illustration of the complete set of 27 models is presented in the appendix. Moreover, as shown in Table 10.1, for a given configuration of $\mathbf{W}$, it is possible to have different models since the dose parameters are unrestricted.

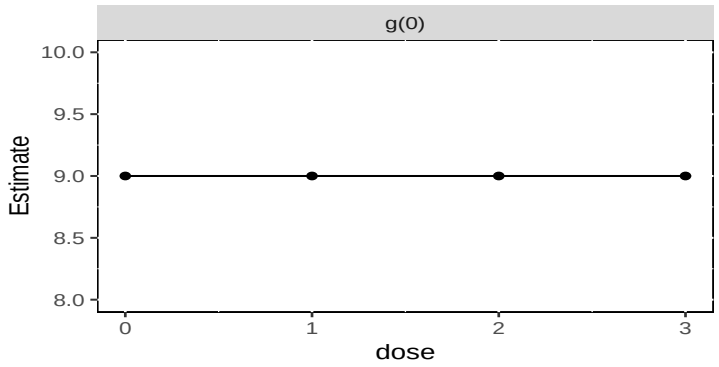


Figure 10.1: Illustration of the null model with configuration $\mathbf{W} = (0, 0, 0)$.

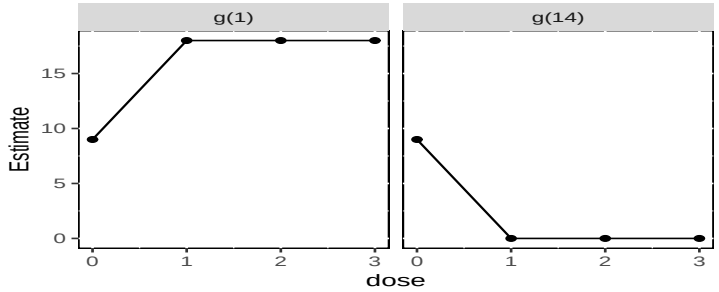


Figure 10.2: Illustration of the models with configuration $\mathbf{W} = (1, 0, 0)$.

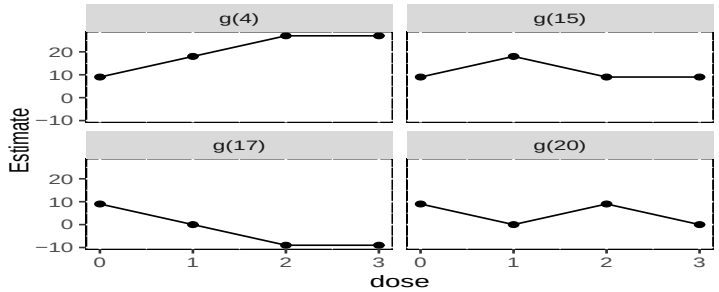


Figure 10.3: Illustration of the models with configuration $\mathbf{W} = (1, 1, 0)$.

10.2.3 A BVS Model for a DRL-72 Experiment with Crossover Design: Mean Structure

For the crossover design, the model includes parameters to adjust for possible carryover and period effects. Similar to Muchene *et al.* (2016a), the mean structure is given by

$$\begin{aligned} log(\lambda_{ij}) &= \alpha_0 + \alpha_{1j} + \alpha_{2k} + \alpha_{3l} + a_i, \\ logit(\pi_{ij}) &= \beta_0 + \beta_{1j} + \beta_{2k} + \beta_{3\ell} + b_i. \end{aligned} \tag{10.11}$$

Here, α_0 and β_0 are the placebo effects in period 1, α_{ij} and β_{1j}, are the effects of dose $j, j = 1, 2, \ldots 5$ for $0, 1.25, 2.5, 5, 10$ mg/kg, respectively, α_{2k} and β_{2k} are the period effects ($k = 1, 2, 3$), while $\alpha_{3\ell}$ and $\beta_{3\ell}$ are the carryover effects of the administered doses ($\ell = 0, 1.25, 2.5, 5, 10$ mg/kg). Note that in addition to a placebo which was administered during the training phase, a control dose (0.0 mg/kg) was one of the doses administered during the treatment day. Moreover, this model specification allows to distinguish between the first study period in which there is no carryover effect and the later periods in which a treatment effect from the previous period could be carried over onto the following period (Jones and Kenward, 1989).

The primary interest is placed on the dose effects. Thus, the BVS component is defined only on the dose parameters α_{1j} and β_{1j} in (10.11), which are parametrized as follows:

$$\begin{aligned} \alpha_{1j} &= \sum_{j=1}^{5} w_j \gamma_j X_{ij}, \\ \beta_{1j} &= \sum_{j=1}^{5} z_j \delta_j X_{ij}. \end{aligned} \tag{10.12}$$

Similar to the mean structure of the parallel design, the dose effects γ_j and δ_j, are defined respectively, for response and rewards model, while the BVS indicator variables w_j and z_j are as defined in (10.7). Further, the change from the control dose can be estimated from contrasts between each dose and the control dose. That is, $\alpha_{1j} - \alpha_{11}$ and $\beta_{1j} - \beta_{11}$, for $j = 2, 3, 4, 5$.

10.2.4 Hierarchical Bayesian Binomial-Poisson Joint Model with a BVS Component

Parallel Design DRL-72 Experiment

Non-informative Gaussian priors for the regression parameters in (10.8) are given by,

$$\begin{aligned} \gamma_0 \sim Normal(0.0, \tau_{a_0}),\ \delta_0 \sim Normal(0.0, \tau_{b_0}), \\ \gamma_j \sim Normal(0.0, \tau_g),\ \delta_j \sim Normal(0.0, \tau_d). \end{aligned} \tag{10.13}$$

For all parameters, the inverse of the variance is assumed to follow a $Gamma(0.001, 0.001)$.

As mentioned in Section 8.1.2, a BVS model requires specifying prior models for the inclusion parameters. Following the methodology of O'Hara and Sillanpaa (2009), we assume that the prior for w_j and z_j are independent. That is,

$$\begin{aligned} w_j &\sim Bernoulli(\psi_j), \\ z_j &\sim Bernoulli(\upsilon_j). \end{aligned} \tag{10.14}$$

and

$$\begin{aligned} \psi_j &\sim Uniform(0,1), \\ \upsilon_j &\sim Uniform(0,1). \end{aligned} \tag{10.15}$$

In order to capture the association between the number of responses and number of rewards, the prior model for random effects for response and rewards models, is a bivariate normal distribution given by,

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim Normal\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{D} \right), \ \mathbf{D} = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}. \tag{10.16}$$

Subsequently, the correlation between the random effects, having corrected for covariate effect is given by,

$$\rho_{ab} = \frac{\sigma_{ab}}{\sqrt{\sigma_a^2 \times \sigma_b^2}}. \tag{10.17}$$

An Wishart prior distribution is specified for $\mathbf{D}^{-1}$ given by,

$$\mathbf{D}^{-1} \sim Wishart(\mathbf{R}_D, k) \tag{10.18}$$

The degrees of freedom *k* is set to be at least equal to the rank of $\mathbf{D}^{-1}$. The scale matrix $\mathbf{R}_D$ is a 2×2 diagonal matrix with off-diagonal elements set to 0.001 (Shkedy *et al.*, 2005, Gelman and Hill, 2007, Lesaffre and Lawson, 2012, Huang and Wand, 2013).

Crossover Design DRL-72 Experiment

The priors and hyper priors for the dose parameters, random effects, w_j and z_j are the same as defined in Section 10.2.4. For the remaining fixed effects parameters in (10.11), the priors and hyper priors are given by,

$$\begin{aligned} &\alpha_{2k} \sim Normal(0.0, \tau_{a_2}), \ \beta_{2k} \sim Normal(0.0, \tau_{b_2}), \\ &\alpha_{3\ell} \sim Normal(0.0, \tau_{a_3}), \ \beta_{3\ell} \sim Normal(0.0, \tau_{b_3}). \end{aligned} \tag{10.19}$$

10.2.5 Posterior Model Probability and Posterior Means

Posterior Model Probability

Consider a dose-response study with d dose levels. Let $\mathbf{Z} = (z_1, z_2, \ldots, z_{d-1})$ and $\mathbf{W} = (w_1, w_2, \ldots, w_{d-1})$, be the inclusion vectors. As explained in Section 10.2.1, the mean structure depends on a specific configuration of $\mathbf{W}$ and $\mathbf{Z}$. For an experiment with d dose levels, there are 2^{d-1} different configurations of $\mathbf{W}$ and $\mathbf{Z}$. The total number of

unique models without an order constraint on the dose parameters is equal to 3^{d-1}. This is the reason why the configuration of $\mathbf{W}$ and $\mathbf{Z}$ does not uniquely define a model. For example, as shown in Section 10.2.2, for $\mathbf{W} = (1, 0, 0)$, there are two possible models. For a four-dose experiment ($d = 4$), there are 27 unique models as shown in Table 10.1, denoted by $g_0, g_1, \ldots, g_{26}$. Note that g_0 corresponds to the configuration $\mathbf{W} = (0, 0, 0)$, hence for the no-dose-effect (the null) model, there is only one configuration of $\mathbf{W}$. For each model, let $P(g_r|\text{data}, g_0, g_1, \ldots, g_R)$ be the posterior model probability. In contrast with the order-restricted models for which the configuration of $\mathbf{W}$ uniquely defines a model, i.e., $P(g_r|\text{data}, g_0, g_1, \ldots, g_R) = P(\mathbf{W} = (w_w, w_2, w_3)|\text{data}, g_0, g_1, \ldots, g_R)$, in the unrestricted setting, this is not the case, since as shown in Table 10.1, Figure 10.2 and 10.3, several models correspond to the same configuration.

In order to estimate the posterior model probability, we need to define a transformation, U_r which uniquely identifies a model. In this case,

$$P(U_r|\text{data}, g_0, \ldots\ g_R) = P(g_r|\text{data}, g_0, \ldots, g_R). \tag{10.20}$$

We define a transformation $U_r = \tilde{\mathbf{W}}\mathbf{c}^T$ with $\mathbf{c} = (3^0, 3^1, \ldots, 3^{[d-2]})$ and $\tilde{\mathbf{W}}$ is a $3^{d-1} \times (d-1)$ matrix . For the four-dose experiment, $\mathbf{c} = (1, 3, 9)$ and $\tilde{\mathbf{W}}$ is given by (10.22).

Table 10.2 presents the model transformation $U_r = \tilde{\mathbf{W}}\mathbf{c}^T$, with the configuration of $\mathbf{W} = (w_1, w_2, w_3)$.

Using the transformation $\tilde{\mathbf{W}}\mathbf{c}^T$ we can uniquely identify each model in the set of all possible models. For example, for the two models shown in Figure 10.2, model g_1 and g_{13}, the configuration of $\mathbf{W}$ is the same, $\mathbf{W} = (1, 0, 0)$ but $U_1 = (1, 0, 0)(1, 3, 9)^T = 1$ and $U_{14} = (-1, 0, 0)(1, 3, 9)^T = -1$. Hence,

$$P(U_r|\text{data}, g_0, \ldots, g_R) = P(g_r|\text{data}, g_0, \ldots, g_R). \tag{10.21}$$

This implies that we can use the posterior distribution of U_r to estimate the posterior model probability.

$$\tilde{\mathbf{W}} = \tilde{\mathbf{Z}} = \left(\begin{array}{rrr} 0 & 0 & 0 \\ \hline 1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \\ \hline 0 & -1 & -1 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & -1 \\ -1 & 0 & 1 \\ -1 & 1 & 0 \\ -1 & -1 & 0 \\ \hline 1 & -1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & -1 \\ 1 & 1 & 1 \\ -1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & -1 & -1 \end{array}\right). \tag{10.22}$$

Posterior Mean: Bayesian Model Averaging

As pointed out by Lin *et al.* (2012), Kasim *et al.* (2012), and Otava *et al.* (2013) the posterior means for all model parameters are a weighted average of the posterior mean estimates, $\hat{\boldsymbol{\mu}}_r$, obtained under each model where the weights are the posterior model probabilities $P(g_r|\text{data})$. That is,

$$\hat{\boldsymbol{\mu}}_{BVS} = \sum_{r=0}^{R} P(g_r|\text{data})\hat{\boldsymbol{\mu}}_r. \tag{10.23}$$

10.2.6 Accounting for Overdispersion in the BVS Model

The BVS model discussed above takes into account potential clustering within a subject using random effects. In addition to the variability due to clustering, Poisson and binomial outcomes potentially exhibit extra variability than would be expected of these distributions. The extra variability is termed as overdispersion and it occurs when the population mean is larger than the population variance. In Chapter 9, following the methodology

Table 10.2: Parallel design experiment. Overview of the model configuration for a four-dose experiment. Parameters: the number of fixed effects parameters in the model.

Model	Notation	w_1	w_2	w_3	$\tilde{w}_1$	$\tilde{w}_2$	$\tilde{w}_3$	U_r	Parameters
1	g(0)	0	0	0	0	0	0	0	1
2	g(1)	1	0	0	1	0	0	1	2
3	g(2)	1	1	0	-1	1	0	2	3
4	g(3)	0	1	0	0	1	0	3	2
5	g(4)	1	1	0	1	1	0	4	3
6	g(5)	1	1	1	-1	-1	1	5	4
7	g(6)	0	1	1	0	-1	1	6	3
8	g(7)	1	1	1	1	-1	1	7	4
9	g(8)	1	0	1	-1	0	1	8	3
10	g(9)	0	0	1	0	0	1	9	2
11	g(10)	1	0	1	1	0	1	10	3
12	g(11)	1	1	1	-1	1	1	11	4
13	g(12)	0	1	1	0	1	1	12	3
14	g(13)	1	1	1	1	1	1	13	4
15	g(14)	1	0	0	-1	0	0	-1	2
16	g(15)	1	1	0	1	-1	0	-2	3
17	g(16)	0	1	0	0	-1	0	-3	2
18	g(17)	1	1	0	-1	-1	0	-4	3
19	g(18)	1	1	1	1	1	-1	-5	4
20	g(19)	0	1	1	0	1	-1	-6	3
21	g(20)	1	1	1	-1	1	-1	-7	4
22	g(21)	1	0	1	1	0	-1	-8	3
23	g(22)	0	0	1	0	0	-1	-9	2
24	g(23)	1	0	1	-1	0	-1	-10	3
25	g(24)	1	1	1	1	-1	-1	-11	4
26	g(25)	0	1	1	0	-1	-1	-12	3
27	g(26)	1	1	1	-1	-1	-1	-13	4

proposed by Molenberghs *et al.* (2007, 2010) and Aregay *et al.* (2015), we proposed a hierarchical Bayesian binomial-Poisson joint model for the two case studies presented above.

Formulation of overdispersed BVS model is straightforward. The BVS formulation for the $\mathbf{X}_i\boldsymbol{\alpha}$ and $\mathbf{X}_i\boldsymbol{\beta}$ in (10.8) can be combined with the model formulation for the overdispersed models in (9.10) and (9.11) for multiplicative and additive models, respectively.

10.3 Application of the Bayesian Variable Selection Methodology

10.3.1 DRL-72 Experiment with Parallel Design

The hierarchical models discussed in Section 10.2 were fitted using Markov Chain Monte Carlo (MCMC) algorithm with 60000 iterations of which 30000 were considered as burn-in period. Three parallel chains with different starting values were initiated, whose mixing and convergence was monitored using trace plots and the potential scale reduction factor (Lesaffre and Lawson, 2012, Gelman and Rubin, 1992). As mentioned in Section 10.2, for this case study, the set of models to be evaluated consists of 27 models for each outcome (a graphical illustration of the models is presented in Figure 10.13 in the appendix).

Model Selection

The corresponding posterior probabilities for a subset of the 27 models with non-zero probability for the expected number of responses and rewards respectively are shown in Figure 10.4. For the expected number of rewards, model g_{12} had the highest posterior probability ($P_{g_{12}} = 0.86$). For this model, $\mathbf{Z} = (0, 1, 1)$ which implies that the posterior mean of the dose parameters are given by

$$\begin{array}{ll} \beta_0 = \delta_0, & \text{if dose} = 0.0\text{mg/kg}, \\ \beta_1 = \delta_0, & \text{if dose} = 2.5\text{mg/kg}, \\ \beta_2 = \delta_0 + \delta_2, & \text{if dose} = 5.0\text{mg/kg}, \\ \beta_3 = \delta_0 + \delta_2 + \delta_3, & \text{if dose} = 10.0\text{mg/kg}. \end{array} \tag{10.24}$$

Hence, for this model, the minimum effective dose (i.e., the dose for which a first response is observed) for the number of rewards is 5.0mg/kg. Furthermore, an additional change is observed in 10.0mg/kg. Note that the configuration of $\mathbf{Z}$ for this model implies that the number of parameters is equal to three. There are 4 possible models satisfying this configuration. Using the transformation function, we conclude that g_{12} is a model for which $\delta_2 > 0$ and $\delta_3 > 0$. Note that for this model, $\tilde{\mathbf{W}}\mathbf{c}^T = (0, 1, 1)(1, 3, 9)^T = 12$. For all other models, the posterior model probability is less than 10%.

For the number of responses, the model with the highest posterior probability is g_{16} with $P(g_{16}|\text{data}, g_0, g_1, \ldots, g_{26}) = 0.55$. The configuration of this model, $\mathbf{W} = (0, 1, 0)$ implies a two parameters model, $\tilde{\mathbf{W}}\mathbf{c}^T = (0, -1, 0)(1, 3, 9)^T = -3 = U_{16}$. Hence, for

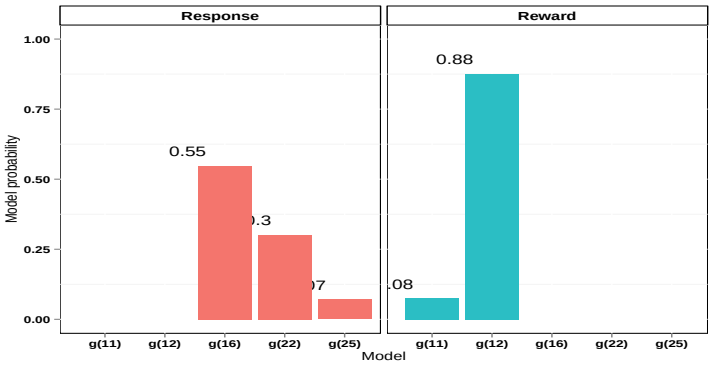


Figure 10.4: Parallel DRL-72. Posterior model probability for the number of responses (left panel) and number of rewards (right panel), respectively.

this model, $\gamma_2 < 0$ and the minimum effective dose is 5mg/kg with no additional effect at the highest dose level. The mean parameter at each dose level for this model is given by

$$\begin{array}{ll} \alpha_0 = \gamma_0, & \text{if dose} = 0.0\text{mg/kg}, \\ \alpha_1 = \gamma_0, & \text{if dose} = 2.5\text{mg/kg}, \\ \alpha_2 = \gamma_0 + \gamma_2, & \text{if dose} = 5.0\text{mg/kg}, \\ \alpha_3 = \gamma_0 + \gamma_2, & \text{if dose} = 10.0\text{mg/kg}. \end{array} \tag{10.25}$$

Note that model g_{22}, corresponding to a configuration of $\mathbf{W} = (0, 0, 1)$ and $\tilde{\mathbf{W}}\mathbf{c}^T = (0, 0, -1)(1, 3, 9)^T = -9 = U_{22}$, has a posterior probability of $P(g_{22}|\text{data}, g_0, g_1, \ldots, g_{26}) = 0.31$. This model has two parameters which implies that the highest dose (10mg/kg) is the minimum effective dose. For each dose, the posterior mean is given by

$$\begin{array}{ll} \alpha_0 = \gamma_0, & \text{if dose} = 0.0\text{mg/kg}, \\ \alpha_1 = \gamma_0, & \text{if dose} = 2.5\text{mg/kg}, \\ \alpha_2 = \gamma_0, & \text{if dose} = 5.0\text{mg/kg}, \\ \alpha_3 = \gamma_0 + \gamma_2, & \text{if dose} = 10.0\text{mg/kg}. \end{array} \tag{10.26}$$

The estimates for the posterior mean of the number of responses and number of rewards were obtained as a weighted average of the estimates per model during the MCMC simulation as described in Section 10.2.5 and are shown in Figure 10.5. Note that BVS shrinks the posterior estimates of the dose effects to an average of the dose effects for the competing models with non zero probability.

Posterior Means for the Dose Effect

Posterior means for the contrasts between the dose effect at each dose level to the control dose are presented in Table 10.3. Note that, similar to the posterior mean of the

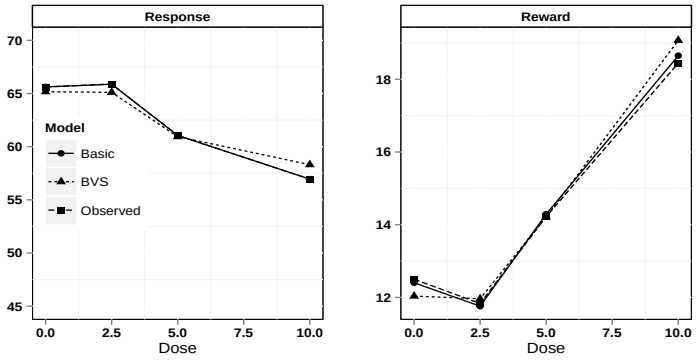


Figure 10.5: BVS model average of the dose effect in the response and reward models. Dashed line: the observed data.

Table 10.3: BVS model-averaged posterior mean (95% credible intervals) estimates for dose effect.

Parameter	Dose	Estimate (CI)
α_1	0.0mg	4.169 (4.092, 4.245)
γ_1	2.5mg vs. 0.0mg	-0.001 (-0.008, 0)
$\gamma_1 + \gamma_2$	5.0mg vs. 0.0mg	-0.069 (-0.161, 0)
$\gamma_1 + \gamma_2 + \gamma_3$	10.0mg vs. 0.0mg	-0.111 (-0.186, 0)
β_1	0.0mg	-1.482 (-1.745, -1.247)
δ_1	2.5mg vs. 0.0mg	-0.007 (-0.138, 0.017)
$\delta_1 + \delta_2$	5.0mg vs. 0.0mg	0.307 (0.000, 0.491)
$\delta_1 + \delta_2 + \delta_3$	10.0mg vs. 0.0mg	0.794 (0.612, 0.969)

expected values, these are the model-averaged values. Interestingly, although the prior distributions for all parameters in the model are unconstrained, the posterior distributions shown in Figure 10.6 (and credible intervals) are bounded by zero. For the number of responses, the posterior distributions for the parameters γ_1, γ_2 and γ_3 are bounded on the right side by zero implying that during the MCMC simulation, these parameters were negative and whenever a positive value was drawn, the parameter was not included in the model. The same pattern, on the positive scale, is observed for δ_2 for the number of rewards.

Inclusion Probabilities

The posterior inclusion probabilities are shown in Table 10.4. Note that the posterior inclusion probability is the average of the inclusion indices, where T is the number of MCMC simulations. That is,

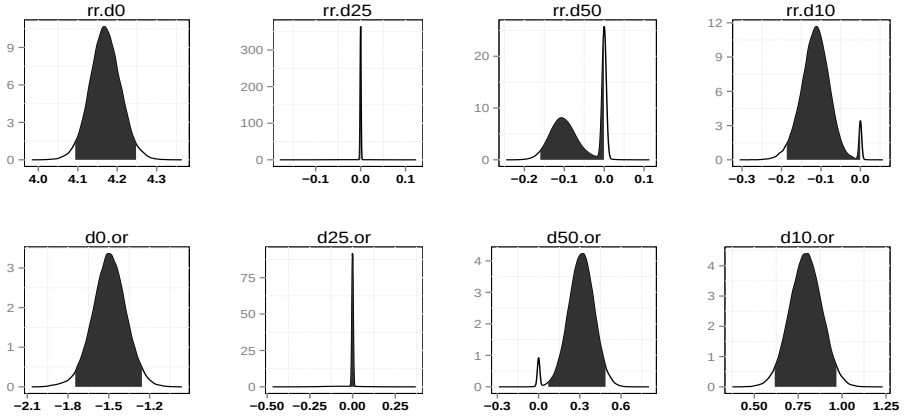


Figure 10.6: Kernel density estimate for the the posterior distribution of dose parameters. The solid fill denotes area under the 95% credible interval.

$$E\left(\tilde{w}_j\right) = \frac{1}{T}\sum_{b=1}^{T}\tilde{w}_{bj} \text{ and } E\left(\tilde{z}_j\right) = \frac{1}{T}\sum_{b=1}^{T}\tilde{z}_{bj}. \tag{10.27}$$

For the number of rewards, the inclusion probability for δ_1 is 0.1 which implies that, in 90% of the models in the MCMC simulation, there was no effect between dose 2.5 and the zero dose. The inclusion probability of δ_2 is 0.971, this explains why the 95% credible interval for the parameter is bounded by zero; 2.9% of the values on the left side of the posterior distribution are equal to zero while the non-zero value are always positive. Note that δ_3, the effect of the last dose level was included in 100% of the models within the MCMC simulation.

For the number of responses, the inclusion probabilities are 0.045, 0.653 and 0.393 for γ_1, γ_2 and γ_3, respectively. The density estimate for the posterior distribution shown in Figure 10.6 and the credible intervals presented in Table 10.3 indicate that these parameters were either not included in the models fitted within the MCMC simulation and whenever they were included, their values were negative; an indication of a decreasing trend in expected number of responses with increasing dose.

Correlation Between Rewards and Responses

The joint model formulated on Section 10.2 allows for the estimation of the association between the number of times the rats press the lever and the number of rewards obtained. Parameter estimates for the element of the covariance matrix $\mathbf{D}$ in (10.16) are shown in Table 10.5, from which, the posterior mean of the correlation between the random effects, $\bar{\rho} = -0.998$, implying that, as the number of lever presses decreases, the number of rewards obtained increases. A similar result was obtained for the models fitted in

Table 10.4: Posterior mean of δ_j and γ_j and the resulting posterior inclusion probabilities for the dose components.

Model	Parameter	Estimate	Inclusion Probability
Poisson	γ_1	0.0043	0.045
	γ_2	-0.0389	0.653
	γ_3	0.0048	0.393
Binomial	δ_1	-0.0074	0.1
	δ_2	0.3227	0.971
	δ_3	0.4871	1

Table 10.5: Estimated parameters for the covariance matrix $\mathbf{D}$ of random effects for response and rewards models.

Parameter	Estimate
σ_a^2	0.017
σ_b^2	0.227
ρ_{ab}	-0.988

Chapter 9, (see Table 9.4).

10.3.2 DRL-72 Experiment with Crossover Design

Model Selection

The set of plausible models comprised of 243 models with varying dose-response profiles (an illustration of the models is presented in Figure 10.14-10.16 in the appendix). The posterior probabilities of a subset of the models are shown in Figure 10.7. The most probable model for the number of rewards had a posterior model probability $P(g_{31}|\text{data}, g_0, g_1, \ldots, g_{243}) = 0.20$. For this model, the configuration $\mathbf{Z} = (1,1,0,1,0)$ and $\tilde{\mathbf{W}}\mathbf{c}^T = (1,1,0,1,0)(1,3,9,27,81)^T = 31 = U_{31}$, which implies that the minimum effective dose for the rewards is 0.0 mg/kg, with an additional effect of the 1.25 and 5.0 mg/kg doses. The mean number of rewards at each dose level is given by,

$$\begin{array}{ll} \beta_0 = \delta_0, & \text{if dose} = \text{Placebo}, \\ \beta_1 = \delta_0 + \delta_1, & \text{if dose} = 0.0 \text{ mg/kg}, \\ \beta_2 = \delta_0 + \delta_1 + \delta_2, & \text{if dose} = 1.25 \text{ mg/kg}, \\ \beta_3 = \delta_0 + \delta_1 + \delta_2, & \text{if dose} = 2.50 \text{ mg/kg}, \\ \beta_4 = \delta_0 + \delta_1 + \delta_2 + \delta_4, & \text{if dose} = 5.0 \text{ mg/kg}. \\ \beta_5 = \delta_0 + \delta_1 + \delta_2 + \delta_4, & \text{if dose} = 10.0 \text{ mg/kg}. \end{array} \tag{10.28}$$

For the number of responses, the model with the highest posterior model probability is g_{151} with $P(g_{151}|\text{data}, g_0, \ldots g_{243}) = 0.51$. The configuration of this model, $\mathbf{W} = (0,1,0,1,0)$ and $\tilde{\mathbf{W}}\mathbf{c}^T = (0,-1,0,-1,0)(1,3,9,27,81)^T = -30 = U_{151}$, implies that 1.25 mg/kg is the minimum effective dose, while only the 5.0 mg/kg dose has an additional

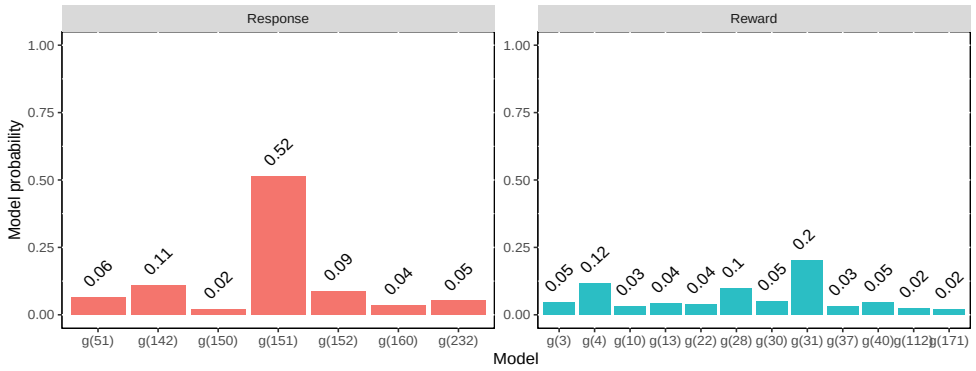


Figure 10.7: Posterior model probabilities for the number of responses (left panel) and number of rewards (right panel), respectively.

effect on the number of responses. The mean dose effect for the number of responses is given by,

$$\begin{array}{ll} \alpha_0 = \gamma_0, & \text{if dose} = \text{Placebo}, \\ \alpha_1 = \gamma_0, & \text{if dose} = 0.0 \text{ mg/kg}, \\ \alpha_2 = \gamma_0 + \gamma_2, & \text{if dose} = 1.25 \text{ mg/kg}, \\ \alpha_3 = \gamma_0 + \gamma_2, & \text{if dose} = 2.5 \text{ mg/kg}, \\ \alpha_4 = \gamma_0 + \gamma_2 + \gamma_4, & \text{if dose} = 5.0 \text{ mg/kg}, \\ \alpha_4 = \gamma_0 + \gamma_2 + \gamma_4, & \text{if dose} = 10.00 \text{ mg/kg}. \end{array} \tag{10.29}$$

Posterior Means for the Dose Effect

The posterior mean estimates of the contrast between the dose effect at each dose level and the 0.0 mg/kg effect are shown in Table 10.6. Note that, as explained in Section 10.3.1, although the prior distributions for all the parameters in the model are unconstrained, the posterior distribution of the dose effects for the number of rewards shown in Figure 10.8, have a lower bound of zero.

Inclusion Probabilities

The posterior inclusion probabilities are shown in Table 10.7. For the response model, all the models in the MCMC simulation contained γ_2 and 99.9% of the models contained γ_4. The inclusion probabilities for γ_1, γ_3 and γ_5 were $0.186, 0.231$ and 0.201, respectively. Note that, although $\gamma_1, \gamma_3, \gamma_4$ and γ_5 were not included in the model in some of the MCMC simulations, the posterior distribution of these parameters contained both positive and negative values as shown in Figure 10.8. This is different in the model for the number of rewards whereby, the 95% credible intervals for $\delta_1, \delta_2, \ldots, \delta_5$ had a lower bound of zero.

Table 10.6: BVS model-averaged posterior mean (95% credible intervals) estimates for dose effect.

Parameter	Dose	Estimate (CI)
$\alpha_{12} - \alpha_{11}$	1.25mg vs. 0.0mg	-0.201 (-0.266, -0.136)
$\alpha_{13} - \alpha_{11}$	2.5mg vs. 0.0mg	-0.194 (-0.254, -0.124)
$\alpha_{14} - \alpha_{11}$	5.0mg vs. 0.0mg	-0.354 (-0.416, -0.285)
$\alpha_{15} - \alpha_{11}$	10.0mg vs. 0.0mg	-0.353 (-0.412, -0.285)
$\beta_{12} - \beta_{11}$	1.25mg vs. 0.0mg	0.247 (0, 0.61)
$\beta_{13} - \beta_{11}$	2.5mg vs. 0.0mg	0.275 (0, 0.629)
$\beta_{14} - \beta_{11}$	5.0mg vs. 0.0mg	0.458 (0, 0.887)
$\beta_{15} - \beta_{11}$	10.0mg vs. 0.0mg	0.842 (0.512, 1.201)

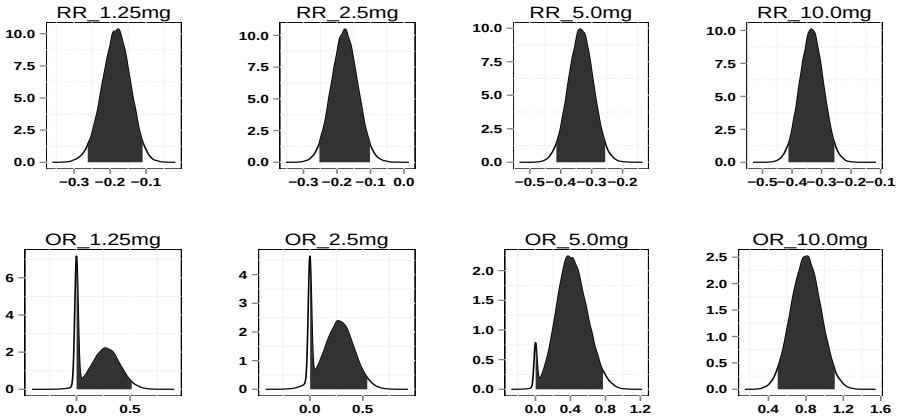


Figure 10.8: Posterior kernel density plots for the BVS dose contrasts. The solid fill denotes the 95% credible intervals.

Table 10.7: The crossover DRL-72 experiment. Posterior inclusion probabilities for the dose parameters.

Model	Parameter	Inclusion Probability
Poisson	γ_1	0.186
	γ_2	1.000
	γ_3	0.231
	γ_4	0.999
	γ_5	0.201
Binomial	δ_1	0.842
	δ_2	0.768
	δ_3	0.310
	δ_4	0.653
	δ_5	0.925

Table 10.8: Estimated parameters for the covariance matrix $\mathbf{D}$ of random effects for response and rewards models.

Parameter	Estimate
σ_a^2	0.113
σ_b^2	0.291
ρ_{ab}	-0.479

Correlation Between Rewards and Responses

Parameter estimates for the posterior mean estimates of the element of the covariance matrix $\mathbf{D}$ is shown in Table 10.8. The parameter estimate for the posterior mean of the correlation between the random effects is equal to -0.479, implying that, the number of rewards increase as the number of responses decrease. A similar pattern was reported in Section 9.3.2 (see Table 9.7).

10.3.3 Overdispersion in the BVS Models

As mentioned in Section 10.2.6, the BVS mean structure can be included in a model which accounts for overdispersion. In this section, we present the results obtained from eight different BVS models (four for the number of responses and four for the number of rewards):

- Basic BVS model (Section 10.2, model (10.8) and (10.11)).
- BVS model + overdispersion on the Binomial endpoint,

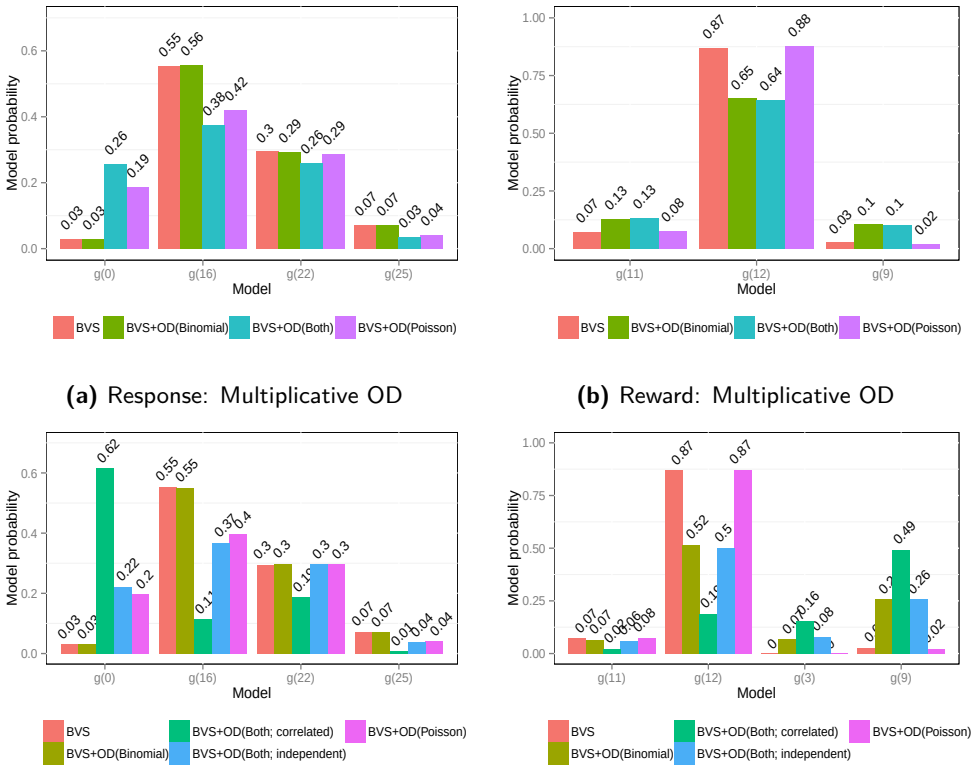
$$\begin{array}{l} Y_{ij} \sim Binomial(n_{ij}, \theta_{2ij}\pi_{ij}), \\ n_{ij} \sim Poisson(\gamma_{ij}). \end{array} \tag{10.30}$$

- BVS model + overdispersion on the Poisson endpoint,

$$\begin{array}{l} Y_{ij} \sim Binomial(n_{ij}, \pi_{ij}), \\ n_{ij} \sim Poisson(\theta_{1ij}\gamma_{ij}). \end{array} \tag{10.31}$$

- BVS model + overdispersion on both endpoints (Section 9.2.2, model (9.10) and (9.11)).

For each type of model, both multiplicative and additive models, discussed in Section 9.2.2, were considered. Note that for the additive models, two BVS models with overdispersion on both endpoints were fitted. The first with independent overdispersion parameters (Section 9.2.2, model (9.12)) and the second with correlated overdispersion parameters (Section 9.2.2, model (9.13)).



(a) Response: Multiplicative OD

(b) Reward: Multiplicative OD

(c) Response: Additive OD

(d) Rewards: Additive OD

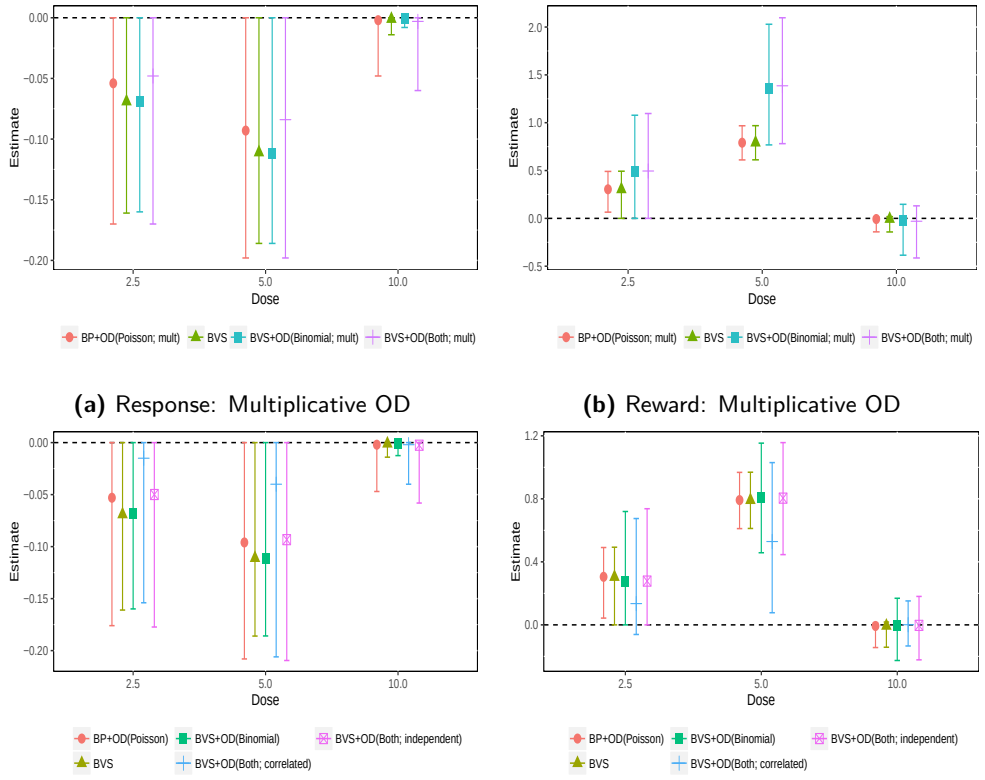
Figure 10.9: DRL-72 with parallel design. Posterior model probability for several BVS models with and without overdispersion, for competing models with large $P(g_r|\text{data}, g_0, g_1, \ldots, g_{27}) > 0.001$.

DRL-72 Experiment with Parallel Design

Figure 10.9 shows the posterior probabilities for the eight BVS models by endpoint. For multiplicative overdispersion models, for both endpoints, all models indicate the same model, g_{16}, for the number of responses and g_{12} for the number of rewards.

For the models with additive overdispersion parameters, the model with correlated overdispersion parameters (Section 9.2.2, model (9.13)) estimates the highest posterior probability for g_0 for the number of responses and g_9 for the number of rewards. All other models estimate the highest posterior probability for g_{16} (number of responses) and g_{12} (for the number of rewards). This is in agreement with the results reported in the previous section, when the BVS model was fitted without overdispersion.

The estimates for the posterior mean for the contrasts between the control dose (0.0 mg/kg) and the treatment doses are shown In Figure 10.10. Although the posterior mean



(a) Response: Multiplicative OD

(b) Reward: Multiplicative OD

(c) Response: Additive OD

(d) Reward: Additive OD

Figure 10.10: DRL-72 with parallel design. Posterior mean estimates with the 95% credible intervals for the contrasts between treatment doses and the control dose $(0.0$ mg/kg). The error bars are jittered for clarity.

estimates for both additive and multiplicative overdispersion models are similar, the 95% credible intervals for the multiplicative model were wider.

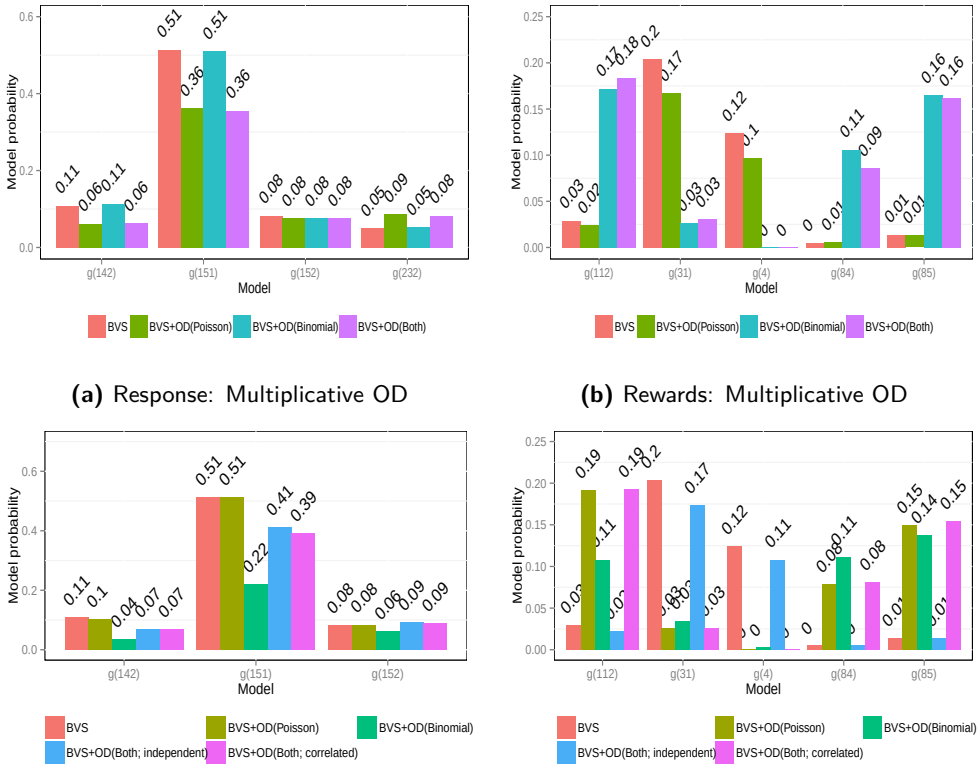
DRL-72 Experiment with a Crossover Design

Figure 10.11 shows the posterior probabilities for the eight BVS models by endpoint. For multiplicative overdispersion models, for both endpoints, all BVS models indicate that model g_{151} was the most probable model for the number of responses. For the number of rewards, the models with the highest probability are shown in Table 10.9.

For the models with additive overdispersion parameters, all the BVS models estimate the highest posterior probability for g_{151} for the number of responses. For the number fo rewards, the model g_{31} which is a BVS model without overdisperion, has the highest posterior probability.

Table 10.9: Crossover DRL-72 experiment. Overview of the additive overdispersion models with the highest posterior probability.

BVS model	Dose-response model	$P(g_r\|\text{data}, g_0, \ldots, g_{243})$
Basic BVS	g_{31}	0.20
BVS+OD(Poisson)	g_{112}	0.19
BVS+OD(Binomial)	g_{85}	0.14
BVS+OD(Both, independent)	g_{31}	0.17
BVS+OD(Both, correlated)	g_{112}	0.19



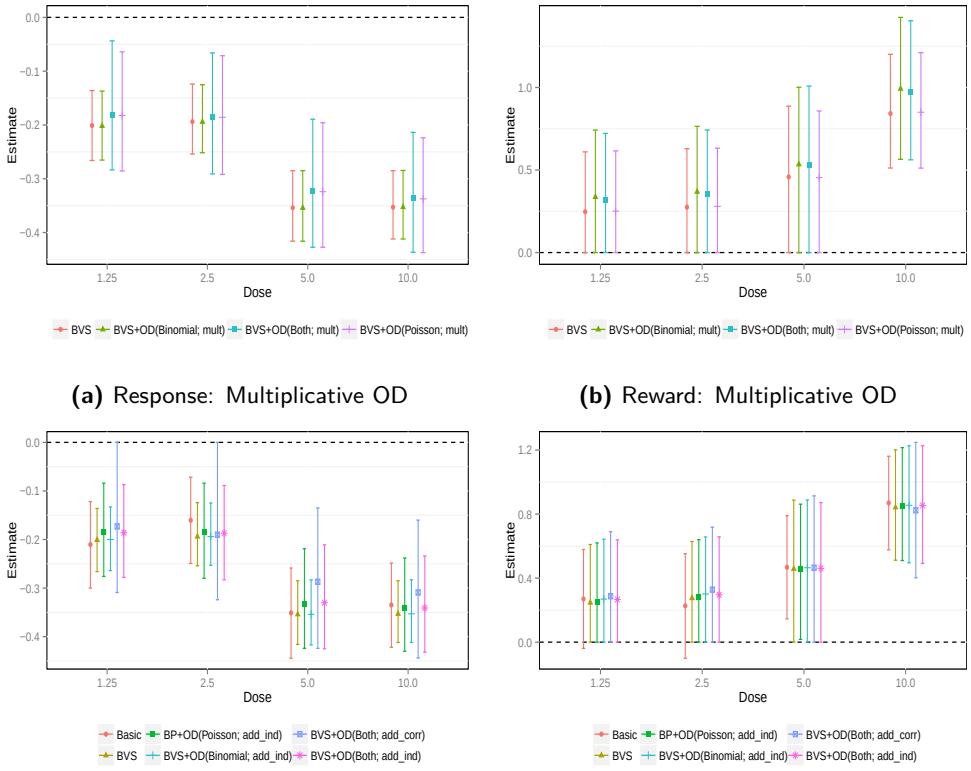
(a) Response: Multiplicative OD

(b) Rewards: Multiplicative OD

(c) Response: Additive OD

(d) Rewards: Additive OD

Figure 10.11: DRL-72 experiment with crossover design. Posterior model probability for several BVS models with and without overdispersion, for competing models with $P(g_r|\text{data}, g_0, \ldots, g_{243} > 0.001)$.



(a) Response: Multiplicative OD

(b) Reward: Multiplicative OD

(c) Response: Additive OD

(d) Reward: Additive OD

Figure 10.12: DRL-72 experiment with Crossover design. Posterior mean estimates with the 95% credible intervals for the contrasts between treatment doses and the control dose $(0.0$ mg/kg$)$. The error bars are jittered for clarity.

The posterior mean estimates for the contrast between treatment doses and the control dose $(0.0$ mg/kg$)$ are presented in Figure 10.12 for all the BVS models. Note that these estimates are model-averaged estimates with the weights corresponding to the posterior model probabilities. Only the highest dose was effective for the number of rewards obtained, while for the response models, the 95% credible intervals did not include zero for all models apart from the correlated additive overdispersion model.

10.4 Discussion

The analysis presented in this chapter focused on model selection and estimation under uncertainty. In contrast with the post selection estimation presented in Chapter 9, the estimation in this chapter was done when multiple models are taken into account.

The Bayesian variable selection model discussed in the chapter allows us to estimate the parameters of interest when a set of candidate models is taken into account. For the analysis we presented in this chapter, we assume that the prior of the parameters of interest and the prior of the inclusion parameters are independent. The configuration of the inclusion parameters correspond to the mean structure and in order to uniquely identify a model from the set of candidate models, we defined a transformation function that was used to estimate the posterior model probability. Using the posterior model probability, we can select a model, if model selection is of interest, and calculate model-averaged estimates for the parameters of interest.

Few issues were not addressed in this chapter. We use non-informative prior probabilities $P(g_r) = 1/R$ and the influence of this choice on the posterior probably was not investigated. Independent priors were used for the parameters of interest and the inclusion parameters. Other approaches are used in the literature (O'Hara and Sillanpaa, 2009). These approaches specify a joint distribution for the two parameters. Several BVS models were used (with and without overdsperion); a question that remains open is how to compare between these models. For example, how can we select between a basic BVS model and a BVS model with overdispersion? All these issues are topics for future investigation.

10.5 Appendix

10.5.1 Parallel Design DRL-72 Experiment

To uniquely identify the model being fitted, in each iteration, a row of elements of $\tilde{\mathbf{Z}} = (\tilde{z}_1, \tilde{z}_2, \tilde{z}_3)$ and $\tilde{\mathbf{W}} = (\tilde{w}_1, \tilde{w}_2, \tilde{w}_3)$ are randomly selected (independent of each other). For the parallel design case study, let $\mathbf{c}^T = (1, 3, 9)$ and $\mathbf{U} = \tilde{\mathbf{Z}}\mathbf{c}^{\mathbf{T}}$, where $\mathbf{U}$ is a $r \times 1$ vector. Further, let $Q_r = 1$ if $U_r \geq 0$ and zero otherwise. The expression $I_m = 1 + |U_r| + (|\max(U_1, U_2, \ldots U_r)| \times (1 - Q_r))$ provides the unique identifiers for all the models, $I_1, \ldots, I_{27}$, which are used as indices to uniquely identify the models in the `RJAGS` package, (Plummer, 2015). Similar output is obtained for $\tilde{\mathbf{W}}$. Figure 10.13 shows an illustration of the 27 models corresponding to a dose-response experiment with four doses, and unconstrained parameters for the dose effects.

10.5.2 Crossover Design DRI-72 Experiment

In this section, we present an illustration of the 243 models corresponding to a dose-response model with six doses (Figure 10.14-10.16).

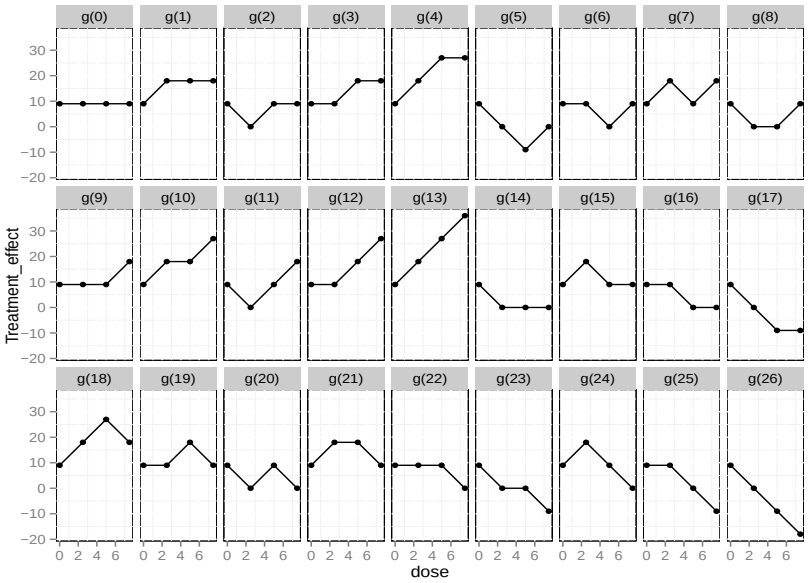


Figure 10.13: Illustration of the set of all plausible dose response profiles in a four dose experiment. In this case, three active doses are compared with a control dose resulting in 27 possible set of models.

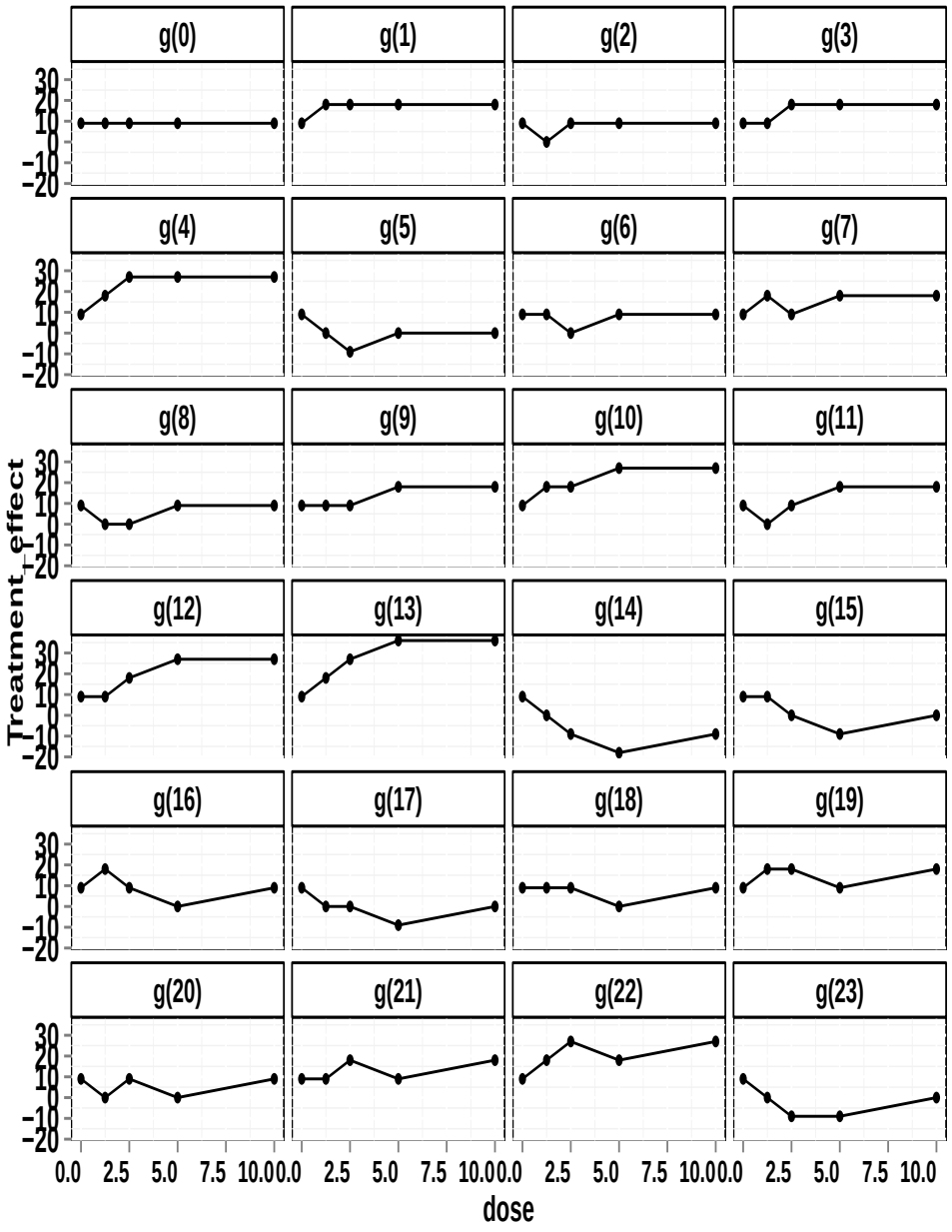


Figure 10.14: Illustration of models $g_0 - g_{23}$ from a crossover study with five treatment doses and a placebo (negative offset in the illustration).

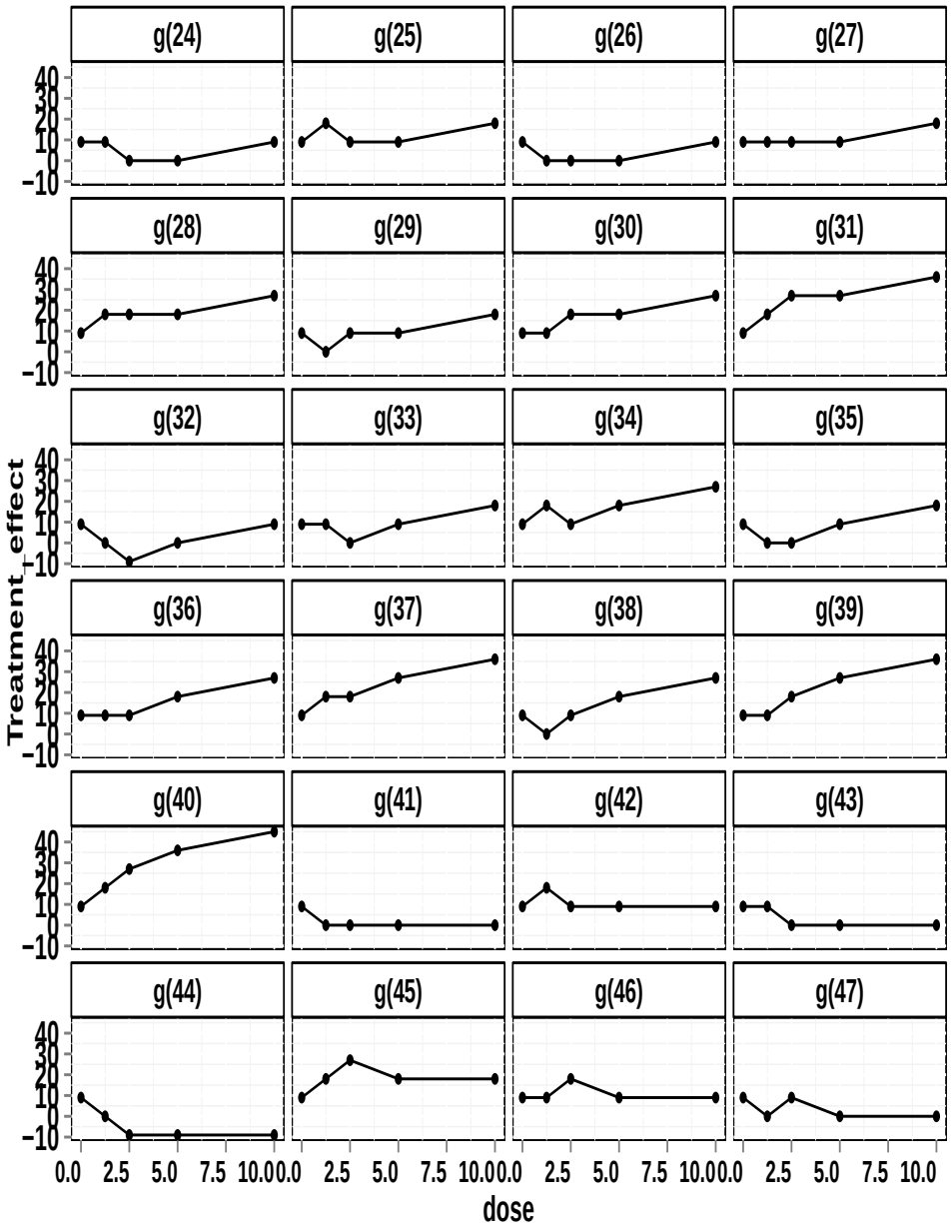


Figure 10.15: Illustration of models $g_{24} - g_{47}$ from a crossover study with five treatment doses and a placebo (negative offset in the illustration).

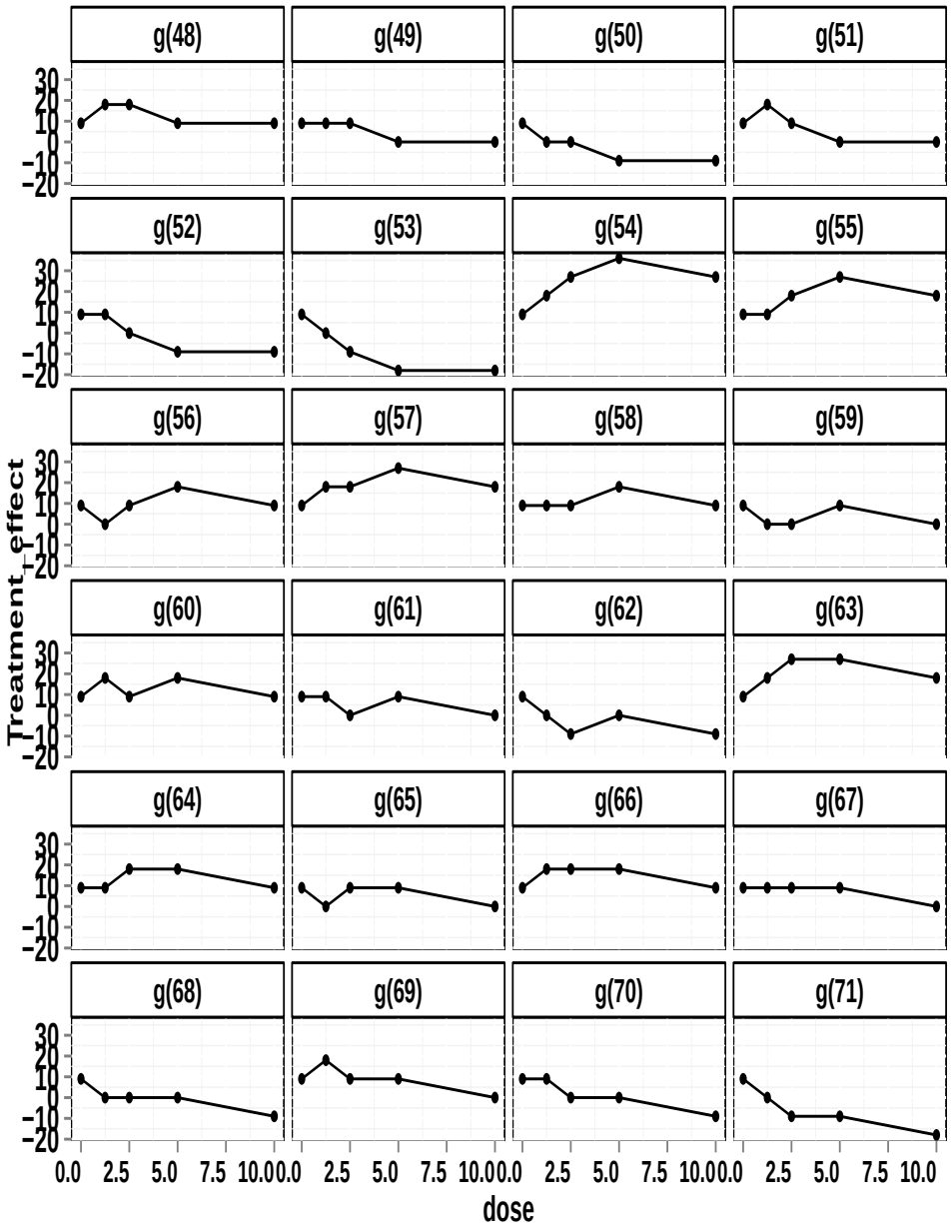


Figure 10.16: Illustration of models $g_{48} - g_{71}$ from a crossover study with five treatment doses and a placebo (negative offset in the illustration).

Chapter 11

Bayesian Variable Selection for Order Restricted Parameters

11.1 Introduction

In the previous chapters, the parameters in the dose-response models were unconstrained. Dose-response studies may postulate a trend in the dose-response relationship such as an increase (decrease) of the effect with increasing dose, up to a point where toxicity presents. Within the Bayesian framework, Klugkist *et al.* (2005), Klugkist and Hoijtink (2007) proposed the use of encompassing priors in modelling order-restricted dose-response models. For the rat growth data presented in Chapter 8), they formulated either an unconstrained model in which there is no order-restriction on the dose effect, or alternatively, a model whose slope coefficients reduce with dose (implying that the growth rate decreases as the dose increase). Appropriate encompassing priors are defined from which alternative hypotheses are then defined by restricting the parameter space according to the constraints imposed by the alternative model. The posterior probability of each model is then derived using the Bayes factor as shown in Chapter 8, hence taking into account the presence of other competing hypotheses.

Within the frequentist framework, Pinheiro *et al.* (2014) proposed a method, implemented in the `dosefinding` package in `R`, in which several parametric order-restricted dose-response models including the Emax, linear and quadratic models are evaluated and the best model selected using Akaike Information Criterion (AIC). Procedures to control for multiple test are then applied on the final model. Model averaging (Hoeting *et al.*, 1999, Cooke, 2009, Claeskens and Hjort, 2008, Burnham and Anderson, 2003) can also be performed by computing the model weights based on information criteria such as AIC and BIC.

Pinheiro *et al.* (2014) suggested that in situations where there is large uncertainty about the most plausible dose-response shape, the candidate models set should contain

a sufficiently diverse set of plausible models to be reviewed. In this chapter, we present Bayesian Variable Selection (BVS) methodology in the context of order-restricted parameters (Lin *et al.*, 2012, Otava *et al.*, 2013). For an order-restricted dose-response model with d dose levels, there are 2^{d-1} plausible monotone dose-response profiles which can be evaluated. The BVS methodology allows fitting all the plausible monotone dose-response profiles. As shown in Chapter 10, the posterior probability of each model is simultaneously computed in one model fitting step and subsequently, the final parameters represents a model-average of estimates from all the models fitted.

In contrast with the modelling approach of Klugkist *et al.* (2005) that focuses only on inequality alternatives, the analysis presented in this chapter introduces **strict equality** constraints in defining the alternative models of interest. In addition, in all their examples (Klugkist *et al.*, 2005, Kato and Hoijtink, 2006), the null model comprised of an unrestricted model, while for all the examples considered in this chapter, the null model comprises of a "no-dose-effect" hypothesis. In other words, for an experiment with $d+1$ dose levels in which the first dose level is a reference and μ_j is the mean response at the jth dose level, the null model defined in (Klugkist *et al.*, 2005, Kato and Hoijtink, 2006) corresponds to the null hypothesis,

$$H_0 : \mu_0 \neq, \ldots, \neq \mu_d, \tag{11.1}$$

while the null model used in this chapter corresponds to the null hypothesis,

$$H_0 : \mu_0 =, \ldots, = \mu_d. \tag{11.2}$$

Furthermore, Klugkist *et al.* (2005) and Kato and Hoijtink (2006) considered a set of predefined alternative models while the analysis presented in this chapter focuses on decomposition of the alternative hypothesis to all plausible order-restricted models fitted under the alternative.

It is important to mention that, although we use the terminology of null hypothesis and alternative hypothesis, we do not aim, in this chapter, at proposing an inference procedure that can be used to test the null hypothesis against an orderred alternative but rather, we focus on model selection and our aim is to develop a selection procedure in which all models that are formulated under the null and the order-restricted alternative are fitted and estimation of the model parameters and selection is done taking into account all possible models.

This chapter is structured as follows; Section 11.2 presents the methodology and model formulation for the case studies, while the results of applying the methodology to the case studies is presented in Section 11.3. Finally, a discussion of is presented in Section 11.4.

11.2 Model Formulation

11.2.1 The Rat Data

Linear Mixed-Effects Model with Unrestricted Slopes

For the rat growth data, Verbeke and Molenberghs (2000) proposed a linear mixed-effects model of the form,

$$\begin{aligned} Y_{jk} &= (\beta_0 + u_{1j}) + (\beta_1 C_j + \beta_2 L_j + \beta_3 H_j + u_{2j}) t_{jk} + \varepsilon_{jk}, \\ t_{jk} &= \log(1 + (\text{Age}_{jk} - 45)/10). \end{aligned} \tag{11.3}$$

Here, Y_{jk} is the cranial length measured between pre-determined points on the skull for rat j at age k, β_0 is the average response at the start of the treatment and $\beta_1 \ldots \beta_3$ are the average time effects (slopes) for the control, low and high doses, respectively. Note that C_j, L_j and H_j are all indicator variables that identify the dose group, i.e.,

$$C_j = \begin{cases} 1, \text{control}, \\ 0, \text{otherwise}. \end{cases} , L_j = \begin{cases} 1, \text{low}, \\ 0, \text{otherwise}. \end{cases} , H_j = \begin{cases} 1, \text{high}, \\ 0, \text{otherwise}. \end{cases} \tag{11.4}$$

The subject-specific parameters u_{1j} and u_{2j}, are the random intercept and slope coefficients, respectively, while t_{jk} is a transformation of the follow-up time such that $t = 0$ corresponds to the start of the experiment. The measurement error, ε_{jk} is assumed to follow a normal distribution, $\varepsilon_{jk} \sim Normal(0, \sigma^2)$.

The model specified in (11.3) implies that the mean response at each dose level is given by

$$E(Y_{jk}) = \begin{cases} (\beta_0 + u_{1j}) + (\beta_1 + u_{2j}) t_{jk}, & \text{for control dose}, \\ (\beta_0 + u_{1j}) + (\beta_2 + u_{2j}) t_{jk}, & \text{for low dose}, \\ (\beta_0 + u_{1j}) + (\beta_3 + u_{2j}) t_{jk}, & \text{for high dose}. \end{cases} \tag{11.5}$$

The model is implemented as a hierarchical Bayesian model by specifying the following priors and hyper priors, for $r = 0, 1, 2, 3$, given by

$$\begin{aligned} \beta_r &\sim Normal\,(0, \tau)\,, \\ \tau &\sim Gamma\,(0.001, 0.001)\,. \end{aligned} \tag{11.6}$$

The random effects follow a joint bivariate normal distribution given by,

$$\begin{pmatrix} u_{1j} \\ u_{2j} \end{pmatrix} \sim Normal\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma \right). \tag{11.7}$$

Here, Σ is a 2×2 covariance matrix, whose inverse follows a Wishart distribution $\Sigma^{-1} \sim Wishart(R, f)$, where R is a 2×2 diagonal matrix and $f = 3$ are the degrees of freedom.

Order Restriction in the Slope Parameters

Let us consider the following order-restricted hypotheses:

$$\begin{array}{l} H_0 : \beta_1 = \beta_2 = \beta_3, \\ H_1 : \beta_1 \geqslant \beta_2 \geqslant \beta_3. \end{array} \tag{11.8}$$

The alternative hypothesis can be decompose into three sub alternatives. That is,

$$\begin{array}{l} H_{11} : \beta_2 > \beta_3 = \beta 4, \\ H_{12} : \beta_2 = \beta_3 > \beta 4, \\ H_{13} : \beta_2 > \beta_3 > \beta 4. \end{array} \tag{11.9}$$

Thus, four different models can be formulated. The first is formulated under the null hypothesis, assuming no dose effect and the later three models can be formulated under the alternative. Often, in such cases with several competing models, model selection is performed using information criteria such as AIC and BIC. The "best" model is then used for hypothesis testing and inference (inference post-selection), while completely ignoring the presence of competing models. Rather, Bayesian variable selection uses information from all the competing models in both estimation and subsequent inference, thus taking into account the model uncertainty.

Next, we reparametrize the model by introducing δ_1 and δ_2, for the incremental effect of the low and high doses, respectively. That is,

$$Y_{jk} = (\beta_0 + u_{1j}) + (\beta_1 X_{1j} + \delta_1 X_{2j} + \delta_2 X_{3j} + u_{2j})\, t_{jk} + \varepsilon_{jk}. \tag{11.10}$$

Here, β_1 is the slope for the control dose, while X_{1j}, X_{2j} and X_{3j} are indicator variables for control, low and high doses, respectively. Note that the design matrix $\mathbf{X}$, is given by

$$\left(\begin{array}{ccc} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ \hline 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ \hline 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 1 \end{array}\right) \tag{11.11}$$

Hence, for a subject randomized to the control dose, $X_{1j} = 1, X_{2j} = 0$ and $X_{3j} = 0$. The mean dose-specific evolution profiles of model (11.10) are given by

$$E(Y_{jk}) = \left\{ \begin{array}{ll} (\beta_0 + u_{1j}) + (\beta_1 + u_{2j})t_{jk} & \text{control dose,} \\ (\beta_0 + u_{1j}) + (\beta_1 + \delta_1 + u_{2j})t_{jk}, & \text{low dose,} \\ (\beta_0 + u_{1j}) + (\beta_1 + \delta_1 + \delta_2 + u_{2j})t_{jk} & \text{high dose.} \end{array} \right. \tag{11.12}$$

The prior distributions for the model parameters are given by

$$\begin{array}{l} \beta_0 \sim Normal\left(0, \tau\right), \\ \delta_s \sim Normal\left(0, \tau_{\delta_s}\right) T\left(, 0.0\right), s = 1, 2, \\ \tau_{\delta_s} \sim Gamma\left(0.001, 0.001\right). \end{array} \tag{11.13}$$

Note that in (11.13), a truncated normal prior for δ_1 and δ_2 is specified. Truncation ensures that only negative values for delta are sampled, therefore enforcing the order constraint. Hence, in this stage, the dose-specific slope parameter is equal to β_1, $\beta_1 + \delta_1$ and $\beta_1 + \delta_1 + \delta_2$ for the control, low and high dose, respectively. This implies that the slope is **strictly** increasing with dose. Hence, using the model specified in (11.10) corresponds to the one formulated under the alternative hypothesis H_{13}.

Bayesian Variable Selection for the Order-Restricted Model

In this section, we formulate a Bayesian variable selection model which allow us to fit all sub models under the alternatives in (11.9).

To implement BVS, the order-restricted model (11.10) is re-written as,

$$Y_{jk} = (\beta_0 + u_{1j}) + (\beta_1 X_{1j} + \delta_1 z_1 X_{2j} + \delta_2 z_2 X_{3j} + u_{2j})\, t_{jk} + \varepsilon_{jk}. \tag{11.14}$$

The variables z_1 and z_2 are Bernoulli random variables such that $z_k \sim bern(\pi_k)$, and $\pi_k \sim \text{Uniform}(0, 1)$, $k = 1, 2$. Note that z_1 and z_2 are the inclusion variables for the parameters, δ_1 and δ_2, respectively. That is,

$$z_k = \begin{cases} 1, \text{ if } \delta_k \text{ in the model,} \\ 0, \text{ Otherwise.} \end{cases} \tag{11.15}$$

This implies that the mean response as each dose level is given by

$$E(Y_{jk}) = \begin{cases} (\beta_0 + u_{1j}) + (\beta_1 + u_{2j}) t_{jk} + \varepsilon_{jk}, & \text{for control dose,} \\ (\beta_0 + u_{1j}) + (\beta_1 + \delta_1 z_1 + u_{2j}) t_{jk} + \varepsilon_{jk}, & \text{for low dose,} \\ (\beta_0 + u_{1j}) + (\beta_1 + \delta_1 z_1 + \delta_2 z_2 + u_{2j}) t_{jk} + \varepsilon_{jk}, & \text{for high dose.} \end{cases} \tag{11.16}$$

The priors for all other model parameters are the same as in Section 11.2.1. The BVS model formulated in (11.14) implies that the mean response depends on the inclusion variables z_1 and z_2 and in particular, it allows us to fit all possible order-restricted models as shown in Table 11.1. Under order restriction, four models can be fitted: the null model and three models for which the mean response is monotone increasing with dose. In the next section, we discuss a model selection procedure based on the BVS model.

Table 11.1: Rat data. Order-restricted models corresponding to the order-restricted hypothesis in (11.14).

Hypothesis	Configuration		Model
	z_1	z_2	
$\beta_1 = \beta_2 = \beta_3$	0	0	$g(0)$
$\beta_1 > \beta_2 = \beta_3$	1	0	$g(1)$
$\beta_1 = \beta_2 > \beta_3$	0	1	$g(2)$
$\beta_1 > \beta_2 > \beta_3$	1	1	$g(3)$

11.2.2 The Milk Protein Content Trial

An Introduction: The Order Restricted Model in Kato and Hoijtnik *et al.* (2006)

A linear mixed-effects model has been proposed for the analysis of the milk protein content trial data by several authors (Verbeke and Molenberghs, 2000, Verbyla and Cullis, 1990, Diggle, 1990). Let Y_{jk} be the milk protein content measured for the jth cow, $j = 1 \dots 79$, in week k, $k = 1 \dots K_j$. Further, let bar_j, mix_j and lup_j, be indicator variables for the diet to which cow j was assigned. That is, $bar_j = 1$, for barley only diet and zero otherwise, $mix_j = 1$, for the mixture diet (barley and lupins mixture) and zero otherwise, $lup_j = 1$, for lupins only diet and zero otherwise. Kato and Hoijtink (2006) analysed the data using a linear mixed-effects model with diet-specific intercepts, linear and quadratic parameters, given by

$$Y_{jk} = \pi_{1j} + \pi_{2j} t_{jk} + \pi_{3j} t_{jk}^2 + \varepsilon_{jk}. \tag{11.17}$$

Note that t_{jk} is the rescaled time variable such that, $t_{jk} = 0$ corresponds to a duration of 11.5 weeks (Kato and Hoijtink, 2006). Moreover,

$$\begin{aligned} \pi_{1j} &= \beta_1 bar_j + \beta_2 mix_j + \beta_3 lup_j + u_{1j}, \\ \pi_{2j} &= \beta_4 bar_j + \beta_5 mix_j + \beta_6 lup_j + u_{2j}, \\ \pi_{3j} &= \beta_7 bar_j + \beta_8 mix_j + \beta_9 lup_j. \end{aligned} \tag{11.18}$$

Here, $\beta_1, \beta_2 \dots \beta_9$ are the regression coefficients, ε_{jk} is the measurement error, u_{1j} and u_{2j} are the random intercept and random linear slope effects, respectively.

In the context of order-restricted hypotheses, Kato and Hoijtink (2006) presented an analysis of the dataset in which four models were fitted: (M1) an unconstrained model, (M2) ordered intercepts and linear slope parameters whereby, the milk protein content is highest for barley only diet, moderate in the mixture diet and lowest in the lupins only diet, (M3) only the intercepts are ordered such that the milk protein content is highest for barley only diet, moderate in the mixture diet and lowest in the lupins only diet and (M4), there is no effect of the diet on milk protein content. The four models represent four hypothesis (Kato and Hoijtink, 2006), given by

$$
\begin{array}{l}
M_1 : \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \\
M_2 : \{\beta_1 > \beta_2 > \beta_3\}, \{\beta_4 > \beta_5 > \beta_6\}, \beta_7, \beta_8, \beta_9, \\
M_3 : \{\beta_1 > \beta_2 > \beta_3\}, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \\
M_4 : \{\beta_1 \approx \beta_2 \approx \beta_3\}, \{\beta_4 \approx \beta_5 \approx \beta_6\}\{\beta_7 \approx \beta_8 \approx \beta_9\}.
\end{array} \tag{11.19}
$$

In what follow, similar to Kato and Hoijtink (2006), we propose a linear mixed-effects model for the data in which both the intercept and linear slope fixed effects are order-restricted, while the quadratic fixed effects are unconstrained. However, in contrast with Kato and Hoijtink (2006), the proposed model allows for equality constraint and similar to the models in Section 11.2.1, the null model that implies no dose effect is included in the set of all possible models.

The Order-Restricted Hierarchical Bayesian Model

As before, in the first step, for the intercept and slope parameters, we formulate two models, the first assumes equality of the parameters across the diet groups the second assumes monotonicity among the parameters, i.e. the intercept in the barley group is strictly larger than the intercept in the mixture group and the intercept in the mixture group is strictly larger than the intercept in the lupins group. A similar ordering is enforced for the slope coefficients. The parameters for the quadratic time are unrestricted. Thus,

$$
\begin{array}{l}
M_0 : \{\beta_1 = \beta_2 = \beta_3\}, \{\beta_4, \beta_5 = \beta_6\}, \beta_7, \beta_8, \beta_9, \\
M_1 : \{\beta_1 > \beta_2 > \beta_3\}, \{\beta_4 > \beta_5 > \beta_6\}, \beta_7, \beta_8, \beta_9.
\end{array} \tag{11.20}
$$

To implement order-restriction in the intercepts and linear slope parameters, we rewrite (11.18) as follows:

$$
\begin{array}{l}
\pi_{1j} = \alpha_1 X_{1j} + \delta_1 X_{2j} + \delta_2 X_{3j} + u_{1j}, \\
\pi_{2j} = \alpha_2 X_{1j} + \gamma_1 X_{2j} + \gamma_2 X_{3j} + u_{2j}, \\
\pi_{3j} = \alpha_3 X_{1j} + \phi_1 X_{2j} + \phi_2 X_{3j}.
\end{array} \tag{11.21}
$$

Here, α_1, α_2 and α_3 are the intercept, linear and quadratic effects, respectively, for cows in the barley diet and u_{1jk} and u_{2jk} are the random effects for intercept and linear slope, respectively. Further, $\varepsilon_{jk} \sim Normal(0.0, \tau)$.

Similar to Section 11.2.1, we define δ_1 and δ_2 as the additional effects of mixture and lupins diets, respectively, on the intercepts, γ_1 and γ_2 are the additional effects of mixture and lupins diets, respectively, on the linear slopes while ϕ_1 and ϕ_2 are the additional effects of the mixture and lupins diets, respectively, on the quadratic slopes. Note that X_1, X_2 and X_3 are indicator variables for the diet to which cow j is assigned. If cow j is assigned to barley, $X_{1j} = 1, X_{2j} = 0$ and $X_{3j} = 0$, if assigned to the mixture diet, $X_{1j} = 1, X_{2j} = 1$ and $X_{3j} = 0$ and if assigned to lupins diet, $X_{1j} = 1, X_{2j} = 1$ and $X_{3j} = 1$.

Note that from (11.21), the average milk protein content at 11.5 weeks, denoted by β_1, β_2 and β_3, for barley, mixture and lupins diet, respectively, is given by,

$$
\begin{array}{l}
\beta_1 = \alpha_1, \\
\beta_2 = \alpha_1 + \delta_1, \\
\beta_3 = \alpha_1 + \delta_1 + \delta_2.
\end{array} \tag{11.22}
$$

Similarly, the average linear $(\beta_4, \beta_5, \beta_6)$ and quadratic $\beta_7, \beta_8, \beta_9$ effects of barley only, mixture and lupins only, respectively, are given by,

$$\begin{aligned}
\beta_4 &= \alpha_2, \\
\beta_5 &= \alpha_2 + \gamma_1, \\
\beta_6 &= \alpha_2 + \gamma_1 + \gamma_2, \\
\beta_7 &= \alpha_3, \\
\beta_8 &= \alpha_3 + \phi_1, \\
\beta_9 &= \alpha_3 + \phi_1 + \phi_2.
\end{aligned} \tag{11.23}$$

Order constraints for the fixed intercept and linear slope parameters are implemented by restricting the priors for δ_k and γ_k, $k = 1, 2$, to be negative as follows,

$$\begin{aligned}
\delta_k &\sim Normal\,(0.0, \tau_\delta)\, T\,(, 0.0)\,, \\
\gamma_k &\sim Normal\,(0.0, \tau_\gamma)\, T\,(, 0.0)\,, \\
\tau_\delta &\sim Gamma\,(0.001, 0.001)\,, \\
\tau_\phi &\sim Gamma\,(0.001, 0.001)\,.
\end{aligned} \tag{11.24}$$

The truncated priors in (11.24) implies that $\beta_1 > \beta_2 > \beta_3$ and $\beta_4 > \beta_5 > \beta_6$. Furthermore, the priors and hyperpriors for $\alpha_r, r = 1, 2, 3$, are given by, $\alpha_r \sim Normal\,(0.0, \tau_\alpha)$, where $\tau_\alpha \sim Gamma\,(0.001, 0.001)$, while $\phi_s \sim Normal\,(0.0, \tau_\phi)$ and $\tau_\phi \sim Gamma\,(0.001, 0.001)$. A bivariate normal distribution is specified for the random effects with covariance matrix $\mathbf{D}$, whose inverse follows a Wishart distribution given by $\mathbf{D}^{-1} \sim Wishart\,(R, df)$. The scale matrix R is a 2×2 diagonal matrix with elements equal to 0.001 and $df = 3$.

Bayesian Variable Selection in the Order-Restricted Model

Although Kato and Hoijtink (2006) explored only four hypotheses for the milk protein content data, the alternative hypothesis in (11.20) can be decomposed further to 15 monotone sub hypothesis, which are evaluated within the BVS setting. Let w_k and z_k, $k = 1, 2$ be inclusion indicator variables such that,

$$\begin{aligned}
w_k &= \begin{cases} 1, & \text{if } \delta_k \text{ included in the model,} \\ 0, & \text{otherwise.} \end{cases} \\
&\text{and} \\
z_k &= \begin{cases} 1, & \text{if } \gamma_k \text{ included in the model,} \\ 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{11.25}$$

The BVS model is obtained from (11.21) by including these indicator variables in the mean structure of the model as follows:

$$\begin{aligned}
\pi_{1j} &= \alpha_1 X_{1j} + w_1 \delta_1 X_{2j} + w_2 \delta_2 X_{3j} + u_{1j}, \\
\pi_{2j} &= \alpha_2 X_{1j} + z_1 \gamma_1 X_{2j} + z_2 \gamma_2 X_{3j} + u_{2j}, \\
\pi_{3j} &= \alpha_3 X_{1j} + \phi_1 X_{2j} + \phi_2 X_{3j}.
\end{aligned} \tag{11.26}$$

The specification in (11.26) allows to fit all the sub models formulated under the null hypothesis:

Table 11.2: Milk protein content trial data: Order-restricted models configuration.

Hypothesis		Configuration				
Intercepts	Slopes	w_1	w_2	z_1	z_2	Model
$\beta_1 = \beta_2 = \beta_3$	$\beta_4 = \beta_5 = \beta_6$	0	0	0	0	g(0)
$\beta_1 > \beta_2 = \beta_3$	$\beta_4 = \beta_5 = \beta_6$	1	0	0	0	g(1)
$\beta_1 = \beta_2 > \beta_3$	$\beta_4 = \beta_5 = \beta_6$	0	1	0	0	g(2)
$\beta_1 > \beta_2 > \beta_3$	$\beta_4 = \beta_5 = \beta_6$	1	1	0	0	g(3)
$\beta_1 = \beta_2 = \beta_3$	$\beta_4 > \beta_5 = \beta_6$	0	0	1	0	g(4)
$\beta_1 > \beta_2 = \beta_3$	$\beta_4 > \beta_5 = \beta_6$	1	0	1	0	g(5)
$\beta_1 = \beta_2 > \beta_3$	$\beta_4 > \beta_5 = \beta_6$	0	1	1	0	g(6)
$\beta_1 > \beta_2 > \beta_3$	$\beta_4 > \beta_5 = \beta_6$	1	1	1	0	g(7)
$\beta_1 = \beta_2 = \beta_3$	$\beta_4 = \beta_5 > \beta_6$	0	0	0	1	g(8)
$\beta_1 > \beta_2 = \beta_3$	$\beta_4 = \beta_5 > \beta_6$	1	0	0	1	g(9)
$\beta_1 = \beta_2 > \beta_3$	$\beta_4 = \beta_5 > \beta_6$	0	1	0	1	g(10)
$\beta_1 > \beta_2 > \beta_3$	$\beta_4 = \beta_5 > \beta_6$	1	1	0	1	g(11)
$\beta_1 = \beta_2 = \beta_3$	$\beta_4 > \beta_5 > \beta_6$	0	0	1	1	g(12)
$\beta_1 > \beta_2 = \beta_3$	$\beta_4 > \beta_5 > \beta_6$	1	0	1	1	g(13)
$\beta_1 = \beta_2 > \beta_3$	$\beta_4 > \beta_5 > \beta_6$	0	1	1	1	g(14)
$\beta_1 > \beta_2 > \beta_3$	$\beta_4 > \beta_5 > \beta_6$	1	1	1	1	g(15)

$$H_1 : \{\beta_1 \geq \beta_2 \geq \beta_3\}, \{\beta_4 \geq \beta_5 \geq \beta_6\}, \beta_7, \beta_8, \beta_9. \tag{11.27}$$

Table 11.2 gives an overview of the configuration of w_1, w_2, z_1 and z_2 for the null model (g_0) and all the 15 order-restricted models formulated under the alternative hypothesis.

To complete the BVS model specification, priors and hyperpriors for the indicator variables are given by

$$\begin{aligned} w_s &\sim Bernoulli\,(\varphi_s)\,, \\ z_s &\sim Bernoulli\,(\upsilon_s)\,, \\ \varphi_s &\sim Uniform\,(0,1)\,, \\ \upsilon_s &\sim Uniform\,(0,1)\,. \end{aligned} \tag{11.28}$$

Note that all the remaining priors and hyperpriors are as defined in Section 11.2.1.

11.3 Application to the Data

11.3.1 Rat Data

Figure 11.1 presents the posterior model probabilities and Table 11.3 present the posterior means for the inclusion indicator variables, z_1 and z_2. The model with the highest posterior probability is g_2 with $P(g_2|\text{data}, g_0, g_1, g_2, g_3) = 0.58$. This model assumes that $\beta_1 = \beta_2 > \beta_3$, i.e., the slope for the control and low dose group are equal and both are larger than the slope in the high dose group. Note that the posterior

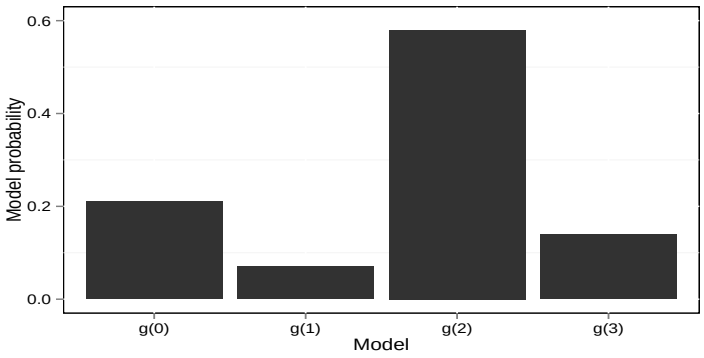


Figure 11.1: Rat data. Posterior model probability.

Table 11.3: Rat data. Posterior mean of the inclusion indicator variables.

Inclusion parameter	Inclusion probability
z_1	0.20
z_2	0.73

inclusion probability of δ_1 is equal to 0.21 while the posterior probability of δ_2 being in the model is 0.71. The kernel density plots for the slope coefficients shown in Figure 11.2 indicate that the posterior mean of the slopes for the low and high doses are bimodal (with mass point at zero for the iterations in which z_1 or z_2 were equal to zero).

Interestingly, the posterior probability of the null model, which assumes no dose effect, is $P(g_0|\text{data}) = 0.22$.

Table 11.4 presents the posterior mean estimates for the hierarchical Bayesian model without order-restriction (i.e, a model with mean structure specified in (11.12) without the order restriction specified in (11.13)), the order-restricted model (11.12)

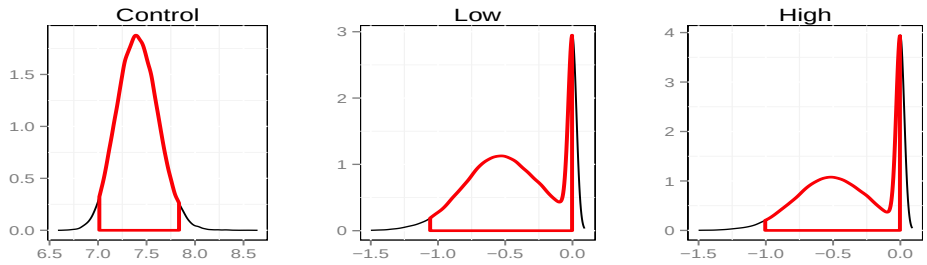


Figure 11.2: Rat data: Kernel density plot for the slope coefficients.

Table 11.4: Rat data. Posterior mean estimates (95% credible intervals) for the Bayesian hierarchical model (HM), order-restricted model (OR) and BVS model with order-restriction (BVS+OR).

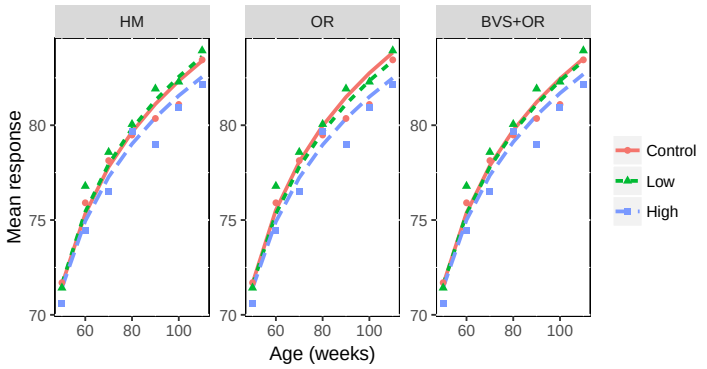
	HM	OR	BVS+OR
β_0	68.61(68, 69.24)	68.6(67.98, 69.22)	68.59(67.97, 69.22)
β_1	7.34(6.79, 7.89)	7.56(7.15, 8.02)	7.4(7.01, 7.84)
β_2	7.45(7, 7.9)	7.34(6.97, 7.73)	7.35(6.96, 7.75)
β_3	6.92(6.46, 7.39)	6.89(6.44, 7.31)	6.99(6.49, 7.45)
$\sigma^2_{u_1}$	3.1(1.51, 5.5)	3.06(1.66, 5.29)	3.18(1.61, 5.39)
$\sigma u_1, u_2$	0.17(-0.49, 0.71)	0.2(-0.47, 0.63)	0.17(-0.34, 0.68)
$\sigma^2_{u_2}$	0.05(0, 0.22)	0.04(0, 0.16)	0.04(0, 0.22)
σ^2_ϵ	1.47(1.21, 1.79)	1.47(1.21, 1.79)	1.46(1.2, 1.78)

and (11.13) and the BVS model with order-restriction (11.14) . Contrasts performed on the slope parameters (and 95% credible intervals) are presented in Table 11.5. The hierarchical model without order-restriction suggests no difference between slopes, while the order-restricted model postulates a difference in slopes across all three contrasts. The contrast estimates for the model-averaged order-restricted estimates obtained from the BVS model are all bounded at zero.

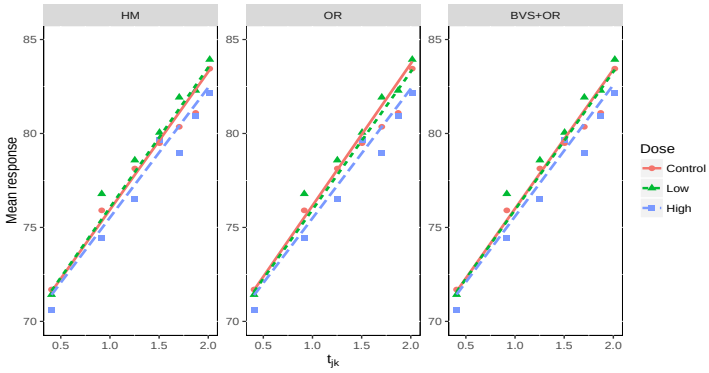
Table 11.5: Rat data. Contrasts (95% credible intervals) for the slope coefficients in the hierarchical Bayesian model (HM), the order-restricted model (OR) and the BVS model with order-restriction (BVS+OR).

Contrast	Parameters	HM	OR	BVS+OR
$\beta_3 - \beta_1$	$\delta_1 + \delta_2$	-0.42 (-1.108,0.256)	-0.67 (-1.244,-0.178)	-0.43 (-1.06,0)
$\beta_2 - \beta_1$	δ_1	0.11 (-0.537,0.755)	-0.22 (-0.651,-0.008)	-0.05 (-0.475,0)
$\beta_3 - \beta_2$	δ_2	-0.53 (-1.137,0.075)	-0.45 (-0.972,-0.041)	-0.38 (-1.005,0)

Figure 11.3 shows the fitted profiles for the three models. For all the three models, the slope for the high dose was the smallest, for the order-restricted models, the slope for the low dose was smaller than that of the control dose.



(a) Average profiles plotted against the original time scale.



(b) Average profiles plotted against the transformed time scale.

Figure 11.3: Rat data. Fitted profiles for the three models. Panel 11.3a shows the original time scale in weeks, while panel 11.3b shows the transformed time scale. The symbols denotes the observed average response.

11.3.2 Milk Protein Content Trial

Figure 11.4 shows the posterior model probability of the competing order-restricted models. Model g_2 was the most probable model with $P(g_2|\text{data}, \text{g}_0, \ldots, \text{g}_{15} = 0.63)$. Note that the configuration of g_2 is $w_1 = 0, w_2 = 1, z_1 = 0, z_2 = 0$, corresponding to the hypothesis of a difference in milk protein content between a lupins only diet and a diet containing Barley (purely or in mixture with lupins) at 11.5 weeks (intercepts). For this model, the barley only diet does not differ from the mixture diet in the mean milk protein content. However, the diet composition does not influence the evolution of the milk protein content over time (no linear slope effects). Other competing models included g_3 with $P(g_3|\text{data}, g_o, \ldots, g_{15}) = 0.2$, with a configuration of, $w_1 = 1, w_2 = 1, z_1 = 0, z_2 = 0$ and g_1 with $P(g_1|\text{data}, \text{g}_0, \ldots, \text{g}_{15}) = 0.12$, with a configuration of, $w_1 = 1, w_2 = 0, z_1 = 0, z_2 = 0$.

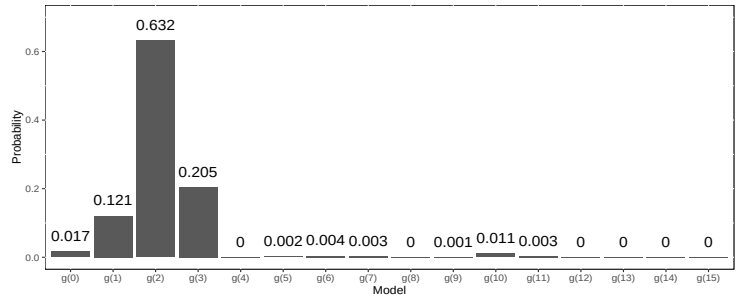


Figure 11.4: Milk protein content trial. Posterior model probability for competing order-restricted models. Only models with $P(g_r|\text{data}, g_o, \ldots, g_{15}) > 0.01$ are shown.

Table 11.6 presents the posterior inclusion probability of $\delta_1, \delta_2, \gamma_1$ and γ_2. The posterior inclusion probability was relatively high for the lupins diet for the intercept coefficient, where it was included in 85.7% of the models. The covariance matrix estimates (see Table 11.7) were similar in the three models, with the correlation between the random intercept and slopes, $\hat{\rho}_{ab} = 0.384$.

Table 11.6: Milk protein content trial. Posterior inclusion probabilities ($\bar{w}$ and $\bar{z}$).

Inclusion parameter	Posterior Mean
w_1	0.344
w_2	0.857
z_1	0.009
z_2	0.019

Table 11.7: Milk protein content trial. Posterior mean of the variance and correlation estimates. HM: unrestricted hierarchical Bayesian model. OR: order-restricted model. BVS+OR: OR model with Bayesian variable selection.

Parameter	HM	OR	BVS+OR
σ_a^2	0.036	0.036	0.037
σ_b^2	0.001	0.001	0.001
ρ_{ab}	0.384	0.384	0.389
σ_ε^2	0.04	0.04	0.04

In Table 11.8, the posterior mean estimates of the fixed effects parameters in (11.18), (11.21) and (11.26) are presented. Although the posterior mean estimates for the intercept and quadratic slope parameters were similar across the three models, the linear slope parameters for the BVS model were different and had wider 95% credible intervals.

Table 11.8: Milk protein content trial. Posterior mean estimates for the fixed effects regression coefficients in the order-restricted (OR) and BVS model with order-restriction (BVS+OR) model. CI: Credible interval.

Parameter	HM: Mean (95% CI)	OR: Mean (95% CI)	BVS+OR: Mean (95% CI
β_1	3.485 (3.405, 3.566)	3.494 (3.417, 3.57)	3.461 (3.392, 3.547)
β_2	3.405 (3.335, 3.476)	3.407 (3.343, 3.471)	3.425 (3.321, 3.496)
β_3	3.291 (3.214, 3.37)	3.285 (3.211, 3.361)	3.294 (3.213, 3.386)
β_4	-0.004 (-0.02, 0.011)	0.001 (-0.012, 0.015)	-0.008 (-0.017, 0.001)
β_5	-0.008 (-0.022, 0.008)	-0.008 (-0.019, 0.004)	-0.008 (-0.018, 0.001)
β_6	-0.013 (-0.028, 0.003)	-0.017 (-0.03, -0.005)	-0.008 (-0.018, 0.001)
β_7	-0.002 (-0.004, -0.001)	-0.002 (-0.004, -0.001)	-0.002 (-0.003, -0.001)
β_8	-0.003 (-0.004, -0.002)	-0.003 (-0.004, -0.002)	-0.003 (-0.004, -0.002)
β_9	-0.002 (-0.004, -0.001)	-0.002 (-0.004, -0.001)	-0.002 (-0.004, -0.001)

11.4 Discussion

In this chapter, we have shown that the Bayesian variable selection framework allows for the estimation of order-restricted parameters in dose-response studies. Order restrictions are incorporated in the model using constrained prior distribution for the order restricted parameters.

When the shape of the dose-response profile is hypothesized to be monotone, but there are several sets of competing monotone profiles to be evaluated, the posterior probability of each of the models can be computed. With this, either model selection can be performed by selecting the dose-response model with the highest posterior model probability, or a model-average of the model parameters can be obtained by averaging over the complete model set.

Several issues were not addressed in this chapter:

- We focus on **model selection** and estimation rather on **inference**. The main questions that remain open is how to use the posterior probability of the model fitted under the null hypothesis, $P(g_0|\text{data}, g_0, \ldots, g_R)$ to conduct inference and equally important, how Type *I* error and power can be controlled.
- Within the hierarchical Bayesian framework, the equivalent number of parameters can be calculated when a single model is fitted to the data (for example, when the DIC is calculated for a fitted model, the equivalent number of parameters is a part of the statistic). The answer to the question-what is the numbers of parameters in the model when a BVS model is fitted?- is at this stage, not clear. One can argue that the complexity of the BVS model depends on the model posterior probabilities and should be averaged over all models in the set of the candidate models. This point should be investigated further.
- Closely related to the previous item are the following questions: how can we select between two non-nested BVS models? Which type of information criterion can we

use?

- It is obvious that model selection based on BVS models provides an attractive alternative to the use of the DIC simply due to the fact that within the BVS framework, model selection can be done in one run of the model while for a model selection based on DIC, all possible models should be fitted. The question is, which method performs better?

Chapter 12

Finite Mixture Models for Dose-Response Data

12.1 Introduction

Throughout Chapter 8 to 11, a joint model for the number of responses and number of rewards, was fitted for the DRL-72 experiment. The two outcomes of interest were obtained by summarizing the observed time lag between subsequent lever presses during the study. For each rat, the number of responses is the count of the time lags within the 60 minutes that the study was performed, while the number of rewards obtained was obtained by evaluating the number of time lags which were at least 72 seconds. Both outcomes are a simplification of the actual observed outcome: the Inter-Response Time (IRT) between subsequent lever presses in the 60 minutes interval. In this chapter, we propose to use a mixture modelling approach to directly model the IRT instead of the Poisson and binomial outcomes used in Chapter 8 to 11. Figure 12.1 illustrates the difference between the variables used in Chapter 8 to 11 and the current Chapter. The response variable, Y_{ijk}, denotes the kth IRT for the ith rat while the jth dose is administered. From Figure 12.1, for rat 1, the number of responses is 6. The number of rewards is the sum of $\mathrm{Y}_{ijk} \geq 72$ for the ith rat and jth dose. In this chapter, we model the length of the IRT (Y_{ijk}) directly.

In behavioural studies, several researchers observed that the distribution of the inter-response time was multi-modal, but proposed simpler analyses which ignore the multi-modal aspect. For example, Jackson *et al.* (1995) excluded observations from the first and last time-bins of their distribution in order to obtain a uni-modal distribution for which they subsequently performed two-way analysis of variance using treatment and bin as factors. Jackson *et al.* (1995) found that there was a shift in the inter-response time peak (mode) for amphetamines. Richards *et al.* (1993) performed a peak deviation analysis of a DRL-72 anti-depressant study from which, they concluded that peak deviation analysis of the inter-response time distribution may provide a useful tool for

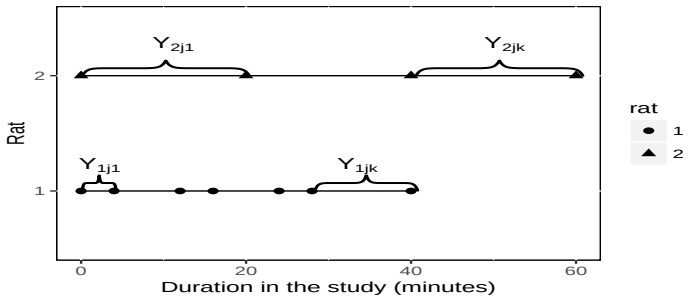


Figure 12.1: An illustration of the inter-response time (IRT) for two rats.

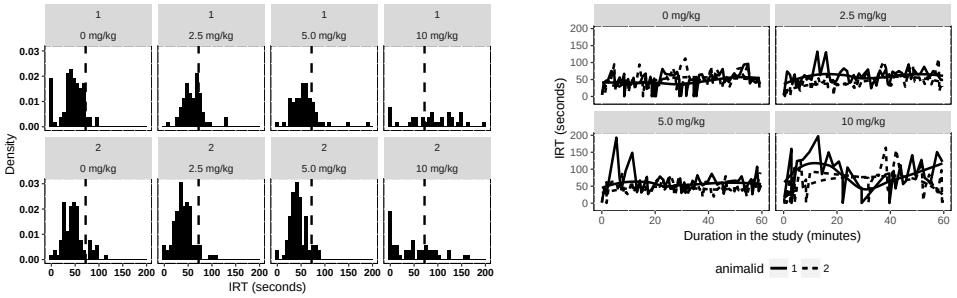
characterising the effect of the drug of interest.

In this chapter, we apply a finite mixture model to the observed IRT data resulting from the DRL-72 experiment with parallel design described in Section 8.3.1, (impying that for each rat, all four doses of the treatment were administered following the same sequence). The inter-response time distribution consists of two (or more) latent components whose underlying parameters are possibly dose-dependent. Similar to the models presented in Chapter 8 to 11, the proposed mixture model accounts for between-rat variability as well. The IRT data is presented in Section 12.2. The general framework of finite mixture models as well as the specific formulation for the application to the DRL-72 experiment is presented in Section 12.3. Subsequently, Section 12.4 provides the results of the final finite mixture model, while a discussion is provided in Section 12.5.

12.2 The DRL-72 Experiment: Inter-Response Time Data

As mentioned in Chapter 8, in total, each of the 18 rats received four doses of an active drug. The number of IRT intervals ranged between 17 and 96. Figure 12.2 shows data for two randomly selected rats. Panel a shows the histogram of IRT in the four doses administered. The dotted line marks 72 seconds beyond which, the rats obtain a reward. For the same rats, Panel b shows the relationship between IRT and the duration in the study for the four administered doses.

Figure 12.3 presents the histogram of the IRT for each of the doses for all animals. For all doses, the distribution is clearly bimodal, with the first component corresponding to very short time intervals (less than approximately 18 seconds) while the second component has a peak for the IRT of about 56 seconds. This suggests different distributions for the IRT which are probably driven by different biological processes. The IRT distribution for individual rats, shown in Figure 12.7 to 12.9 (presented in the appendix of this chapter), indicate presence of variability between rats which ought to be accounted for in statistical analysis.



(a) Histogram of the observed IRT for two rats. **(b)** IRT versus duration in the study.

Figure 12.2: Observed IRT data for two randomly selected rats.

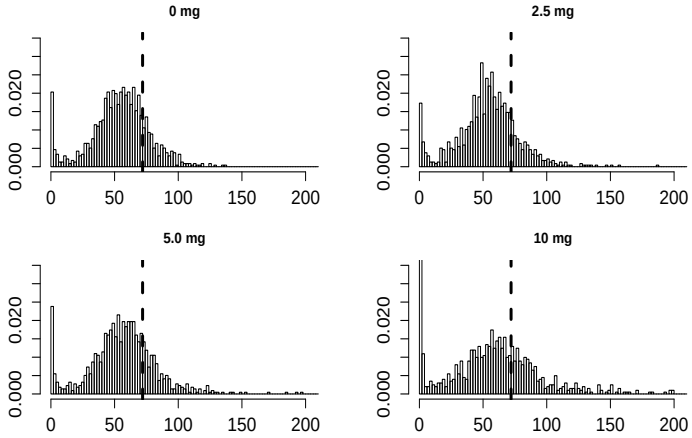
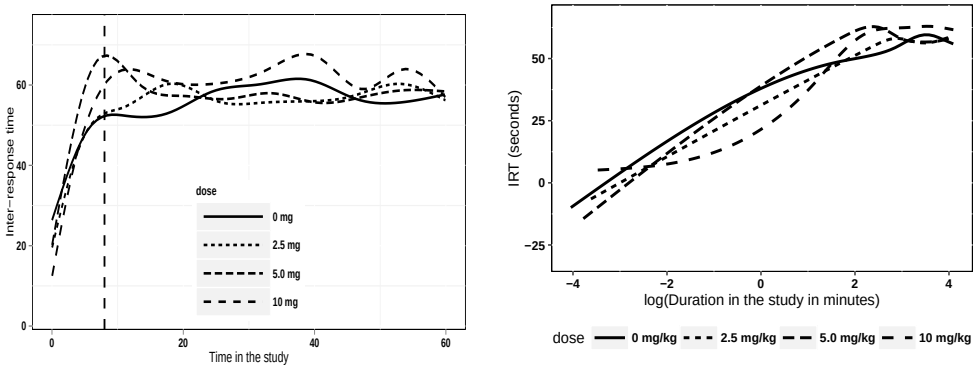


Figure 12.3: Histogram of the time to pressing the lever (IRT) across different doses for all animals. The dashed vertical line marks 72 seconds of IRT, while the area under the curve to the right of the 72 seconds indicates the proportion of lever presses for which a reward is obtained.



(a) IRT versus the duration in the study.

(b) IRT versus the logarithm of the duration in the study.

Figure 12.4: Relationship between IRT and the duration in the study. The vertical dotted line denotes the first 8 minutes in the study.

The relationship between time spent in the study and the IRT is shown in Figure 12.4. Inference for the slope coefficients across the four treatments allows to detect the presence (absence) of a treatment effect. Note that for the analysis presented below, the logarithm of the duration in the study is included as possible covariate. In Figure 12.5, the observed data is superimposed on the plot for the IRT versus time in study, from which, it is clear that even for the first 8 minutes stage of the study, some rats obtain rewards (IRT greater than 72 seconds). In what follows, we describe a finite mixture model which allow to estimate the parameters for each component of the IRT distribution.

12.3 Statistical Methodology

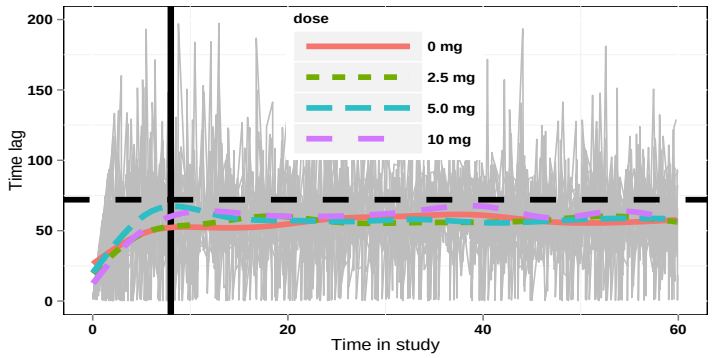
12.3.1 Finite Mixture Models

Finite mixture models (Lindsay, 1995, McLachlan and Peel, 2000) are a way of quantifying latent components of a distribution in a heterogeneous population. Let $\mathbf{Y}_i$ be the IRT for the ith rat, $i = 1, \ldots, 18$, and let $f(y)$ be the density function of Y, assumed to follow a finite mixture distribution given by

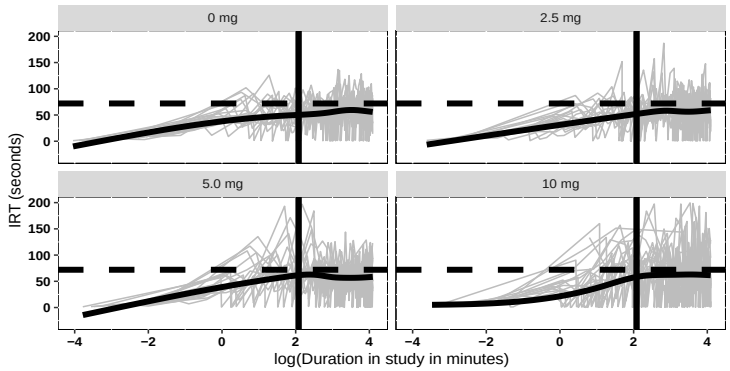
$$f(y) = \pi_1 f_1(y) + \ldots + \pi_g f_g(y). \tag{12.1}$$

Here, $f_r(y)$, $r = 1, \ldots, g$, are g density functions which can be the same distribution with different parameters, $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_g$, or a mixture of different discrete and/or continuous distributions. Note that π_r are the mixing weights (mixing proportions or mixing probabilities) for the finite mixture density satisfying

$$0 < \pi_r < 1 \text{ and } \sum_{r=1}^{g} \pi_r = 1. \tag{12.2}$$



(a) Scatter plot of IRT versus duration in the study with the smoothed dose-specific mean IRT superimposed.



(b) Scatter plot of IRT versus the logarithm of the duration in the study with the smoothed dose-specific mean IRT superimposed.

Figure 12.5: Scatter plot of IRT versus duration in the study with the smoothed dose-specific mean IRT superimposed. The dashed horizontal line corresponds to 72 seconds beyond which rats obtain a reward, while the solid vertical line marks the first eight minutes in the study. The top two quadrants corresponds to time lags of at least 72 seconds for which the rats obtained a reward.

The term finite mixture refers to the fact that the number of density components in (12.2) are fixed. The number of components can either be fixed *a priori* or estimated. For the analysis presented in this chapter, the number of components g was fixed *a priori* hence allowing for classical maximum likelihood estimation to be applied. Sensitivity analysis was performed by varying the choice of the distribution family for $f_r(y)$. Moreover, the proposed model assumes that some of the density-specific parameters, $\boldsymbol{\theta}_g$, depend on both dose and the duration in the study. In order to account for subject heterogeneity, random effects are included in the model.

We limit the choice of $f_r(y)$ to exponential, lognormal and Weibull distributions, which are commonly used in parametric modelling of time-to-event data. Different combinations of two-component finite mixtures including lognormal-lognormal, Weibull-lognormal, Weibull-Weibull, exponential-Weibull and exponential-lognormal finite mixture were explored.

12.3.2 Model Formulation: The Two-Component Lognormal-Lognormal

In this section, we formulate a two-component mixture model in which, the IRT follows a lognormal distribution. The density function for a lognormal distribution (Johnson *et al.*) is given by

$$f(y) = \frac{1}{yS\sqrt{2\pi}} e^{-0.5\left(\frac{log(y)-M}{S}\right)^2}. \tag{12.3}$$

Here, M is the location parameter while S is the scale parameter (other distributions are presented in the appendix for this chapter). Let $y_{ijk} > 0$ be the kth IRT, $k = 1, \ldots, K_{ij}$, for the ith rat, $i = 1, \ldots, n$, under dose j, $j = 1, 2, 3, 4$ for 0.0 mg/kg (control), 2.5 mg/kg (low), 5.0 mg/kg (medium) and 10.0 mg/kg (high), respectively. A two-component lognormal distribution for the IRT is given by

$$y_{ijk} \sim \pi_j \times LogN\left(M_{1ijk}, S_{1j}\right) + (1 - \pi_j) \times LogN\left(M_{2ijk}, S_{2j}\right). \tag{12.4}$$

Here, $LogN()$ denotes a lognormal distribution with location and scale parameters M_{1ijk} and S_{1j}, respectively. Similarly, M_{2ijk} and S_{2j} are the location and scale parameters, respectively, for a lognormal distribution in the second component. Note that the scale parameters (S_{1j} and S_{2j}) in the mixture model specified in (12.4) are assumed to be dose dependent. Let X_{ij} be an indicator variable such that

$$X_{ij} = \begin{cases} 1, \text{ if dose } j \text{ is administered,} \\ 0, \text{ otherwise,} \end{cases} \quad \text{for } j = 2, 3, 4. \tag{12.5}$$

Table 12.1: Parametrization for the dose and subject-specific location parameters, M_{1ijk}, in the first component. A similar parametrization is used for M_{2ijk}.

Dose	Location parameter
0	$\alpha_{11} + \beta_{11}\text{logTstudy}_{ijk} + a_i$
2.5	$(\alpha_{11} + \alpha_{12}) + (\beta_{11} + \beta_{12})\,\text{logTstudy}_{ijk} + a_i$
5.0	$(\alpha_{11} + \alpha_{13}) + (\beta_{11} + \beta_{13})\,\text{logTstudy}_{ijk} + a_i$
10	$(\alpha_{11} + \alpha_{14}) + (\beta_{11} + \beta_{14})\,\text{logTstudy}_{ijk} + a_i$

We consider the following mean structure for the location parameters:

$$M_{1ijk} = \left(\alpha_{11} + \sum_{j=2}^{4} \alpha_{1j} X_{ij}\right) + \left(\beta_{11} + \sum_{j=2}^{4} \beta_{1j} X_{ij}\right) \text{logTstudy}_{ijk} + a_i,$$
$$\text{and} \tag{12.6}$$
$$M_{2ijk} = \left(\alpha_{21} + \sum_{j=2}^{4} \alpha_{2j} X_{ij}\right) + \left(\beta_{21} + \sum_{j=2}^{4} \beta_{2j} X_{ij}\right) \text{logTstudy}_{ijk} + b_i.$$

The variable $logTstudy_{ijk}$ is the logarithm of the duration the rat has been in the study, for the corresponding IRT (Y_{ijk}). For the first component, The parameters α_{11} and β_{11}, are the intercept and slope (i.e. the effect of the logarithm of the duration in the study), respectively, in the control dose. The additional effects of dose j, $j = 2, 3, 4$, on the intercept and slope, are given by α_{1j} and β_{1j}, respectively. The corresponding parameters in the mean structure of M_{2ijk} have the same interpretation. Table 12.1 presents the parametrization of the location parameter at each dose for the first component. In addition, a_i and b_i are subject-specific random intercepts for each component capturing the rat-specific variability of the location parameter from the overall mean. It is assumed that the random intercepts follow a bivariate distribution given by

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim Normal\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix}\right). \tag{12.7}$$

12.4 Application to the DRL-72 Experiment

In the first stage of the analysis, we compare different combinations of distributions for the two components. The models were compared based on the Akaike Information Criterion (AIC, Akaike, 1973) and Bayesian Information Criterion (BIC, Schwarz, 1978) fit statistics. Maximum likelihood estimation for all parameters of interest was performed with SAS `PROC NLMIXED`.

Table 12.2 presents the AIC and BIC values for the two-components finite mixture models considered in this analysis. The lognormal-lognormal finite mixture has the smallest values for both AIC and BIC, hence the most preferred of the five mixture

Table 12.2: AIC and BIC statistics for different two-component finite mixture models. Smaller values of AIC or BIC are preferred.

Model	AIC	BIC
lognormal-lognormal	**40113**	**40141**
Weibull-lognormal	40819	40844
Weibull-Weibull	41084	41106
Exponential-Weibull	41376	41395
Exponential-lognormal	41494	41513

Table 12.3: Mixing proportions π_j and $1 - \pi_j$ for the lognormal densities in the two component lognormal-lognormal finite mixture.

Dose	Component 1	Component 2
0 mg	0.095	0.905
2.5 mg	0.108	0.892
5.0 mg	0.121	0.879
10 mg	0.172	0.828

models considered in this analysis. In the remaining sections of this chapter, detailed results of the lognormal-lognormal finite mixture are presented. The observed and fitted densities of the IRT are presented in Figure 12.6.

12.4.1 Finite Mixture Parameter Estimates

The two-components mixture model specified in (12.3)-(12.7) was applied to the data and leads to the following results:

- **Mixing proportions**:
  The mixing proportions for the two lognormal densities constituting the finite mixture are presented in Table 12.3. The proportion of IRT falling in the first component increases with dose and subsequently, in the second component, the proportion decreases. One possible interpretation for this could be the fact that higher doses of the drug result in an adverse event in the rats which manifests in the form of reduced vigilance, hence the short time intervals between lever presses.

- **Location parameters**:
  The parameter estimates for the location parameters of the two-components are presented in Table 12.4. For the first component, only the high dose has a significant effect on the intercept of the location parameter. For the second component, confidence intervals for the 5.0 mg/kg dose do not cover the value of zero, in both the intercept and slope coefficients, indicating that the location

Table 12.4: Parameter estimates for the regression coefficients for the location parameters in the lognormal mixture.

	Component 1		Component 2	
Dose	Parameter	Estimate (95% CI)	Parameter	Estimate (95% CI)
0.0	α_{11}	1.533 (0.92, 2.145)	α_{21}	3.737 (3.651, 3.824)
2.5	α_{12}	0.108 (-0.518, 0.735)	α_{22}	-0.057 (-0.164, 0.05)
5.0	α_{13}	0.252 (-0.357, 0.861)	α_{23}	0.262 (0.14, 0.385)
10	α_{14}	-0.764 (-1.32, -0.208)	α_{24}	0.125 (-0.03, 0.28)
0.0	β_{11}	0.045 (-0.092, 0.181)	β_{21}	0.095 (0.073, 0.117)
2.5	β_{12}	0.062 (-0.14, 0.265)	β_{22}	0.016 (-0.017, 0.048)
5.0	β_{13}	-0.072 (-0.276, 0.133)	β_{23}	-0.072 (-0.109, -0.035)
10	β_{14}	-0.03 (-0.224, 0.164)	β_{24}	-0.011 (-0.056, 0.035)

parameter of the medium dose is different from the control dose.

- **The expected values of the location parameters**:
  The covariance matrix for the random effects (12.7) is estimated to be equal to

$$\hat{D} = \begin{pmatrix} 0.01 & 0.01 \\ 0.01 & 0.72 \end{pmatrix}. \tag{12.8}$$

  The estimated location and scale parameters for the lognormal-lognormal mixture densities, after 8 minutes and 30 minutes in the study, are presented in Table 12.5. The first component has a larger scale parameter for each dose, compared to the second component. Based on the parameters shown in Table 12.5, we estimated the probability of a rat obtaining a reward, in the first and second component, respectively. That is,

$$P_{1ijk}\left(IRT \geq 72\right) = \int_{72}^{\infty} f_1\left(y\right) dy, \text{ and } P_{2ijk}\left(IRT \geq 72\right) = \int_{72}^{\infty} f_2\left(y\right) dy. \tag{12.9}$$

  The probability to obtain a reward in the first component is between 0.2% and 5.6%, for a rat which had been in the study for 8 minutes and 30 minutes, respectively. In the second component, the probability to obtain a reward increases with dose for the first 8 minutes in the study, while after 30 minutes in the study, the probability was larger but relatively constant.

- **Median IRT**:
  Table 12.5 presents the median time lags for both components. In the first component, the highest dose level results in the smallest median IRT at both 8 minutes and 30 minutes in the study, an indication that the highest dose induces reduced vigilance in the rats resulting in very short IRT. In the second component, the median IRT for the active doses increases for the active

Table 12.5: Parameter estimates for the two-components lognormal finite mixture for a typical rat after 8 minutes and 30 minutes in the study.

Duration	Dose	Component 1		Component 2		Median		$P(IRT \geq 72)$	
		M_{1j}	S_{1j}	M_{2j}	S_{2j}	Comp. 1	Comp. 2	Comp.1	Comp. 2
8	0	1.626	1.329	3.935	0.297	5.084	51.163	0.023	0.125
8	2.5	1.864	1.43	3.911	0.308	6.451	49.926	0.046	0.117
8	5	1.729	1.589	4.048	0.316	5.633	57.26	0.054	0.234
8	10	0.799	1.176	4.038	0.44	2.224	56.686	0.002	0.294
30	0	1.686	1.329	4.061	0.297	5.395	58.017	0.026	0.234
30	2.5	2.006	1.43	4.057	0.308	7.434	57.795	0.056	0.238
30	5	1.693	1.589	4.078	0.316	5.436	59.036	0.052	0.265
30	10	0.819	1.176	4.149	0.44	2.268	63.357	0.002	0.386

doses although by a small margin, and the median IRT is larger after 30 minutes in the study compared to that at 8 minutes in the study for the corresponding doses.

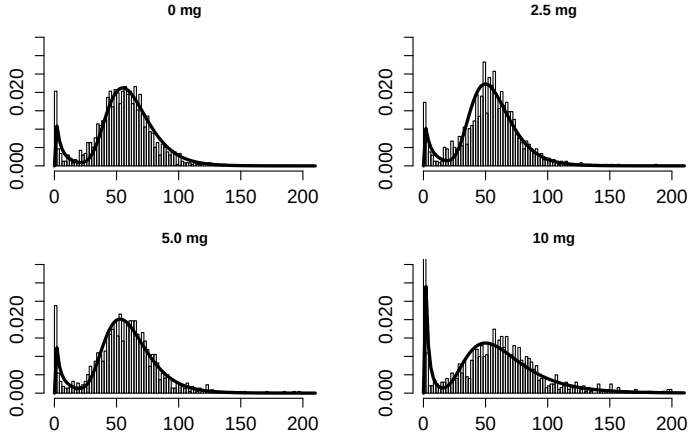


Figure 12.6: Observed inter-response time distribution with the two-component lognormal-lognormal finite mixture superimposed.

12.5 Discussion

In this chapter, the inter-response time distribution was observed to be bimodal. Generally, this is an indicator of the presence of two differentiated behavioural processes, characterized by latent classes. In order to identify the composition of these latent classes, finite mixture models were implemented. Another issue that the models discussed here sought to address was the effect the duration in the study has on the inter-response time.

This was adjusted for in estimating the parameters of the inter-response time distribution.

A two-component lognormal-lognormal finite mixture provided the best fit for the data based on information criteria such as AIC and BIC. In the first component, the median IRT was similar in the control, low and medium doses although the highest dose resulted in substantially shorter IRT. In the second component however, the high dose resulted in slightly longer time lags as compared to the low and medium doses. A formal test for the difference in these median times can be performed following the guidelines discussed by Stoline (1993).

One of the main goals of this analysis was to assess the gain(loss) resulting from summarising the inter-response time outcome into a Poisson count of the number of times the rats press the lever and a binomial outcome (the number of rewards obtained). In Chapter 9, we presented a Bayesian hierarchical binomial-Poisson joint model (BP) for the parallel design DRL-72 experiment, which was based on the summarized IRT. For the Poisson outcome, the log(relative risk) of pressing the lever for the low and medium doses compared to the control dose was not statistically significant although the high dose resulted in a significantly lower risk in comparison to that obtained under the control dose. The log(odds ratio) of successfully obtaining a reward were significantly higher in the medium and high doses as compared to the control dose.

The probability of successfully obtaining a reward computed from the Bayesian hierarchical binomial-Poisson model (BP) from Chapter 9 was $0.188,\ 0.177, 0.235$ and, 0.333 for $0.0,\ 2.5,\ 5.0$ and 10.0 mg/kg, respectively. Comparison with the results from the finite mixture models indicate that, after the first eight minutes in the study, the finite mixture model results in a slightly higher estimated probability of a rat obtaining a reinforcement for all doses. Moreover, the hierarchical model does not provide for a mechanism to assess the possible adverse effect of the high dose yet as was seen from the finite mixture analysis, the higher the dose, the higher the proportion of very short inter-response times.

Another issue that the current analysis addressed was adjusting for the duration in the study. The hierarchical Bayesian Poisson-binomial model summarises the data in such a way that it is no longer possible to associate each of the outcomes with the duration the rat has been in the study. In the finite mixture framework, the duration in the study was useful in adjusting for the location parameters for the two component finite mixture. However, the relationship between the scale parameter of the lognormal distribution on the duration in the study was not investigated in this chapter. The effect of the duration in the study was only significant for the 5.0 mg dose in the second component's location parameter.

On the downside, finite mixture modelling is more complex in terms of numerical computation. For this analysis, SAS `PROC NLMIXED` was used to obtain maximum likelihood estimates of the parameters of interest. This however needs specification of appropriate initial parameter values and a bit of fiddling with different options in order to attain convergence. With increasing model complexity resulting from the introduction of additional

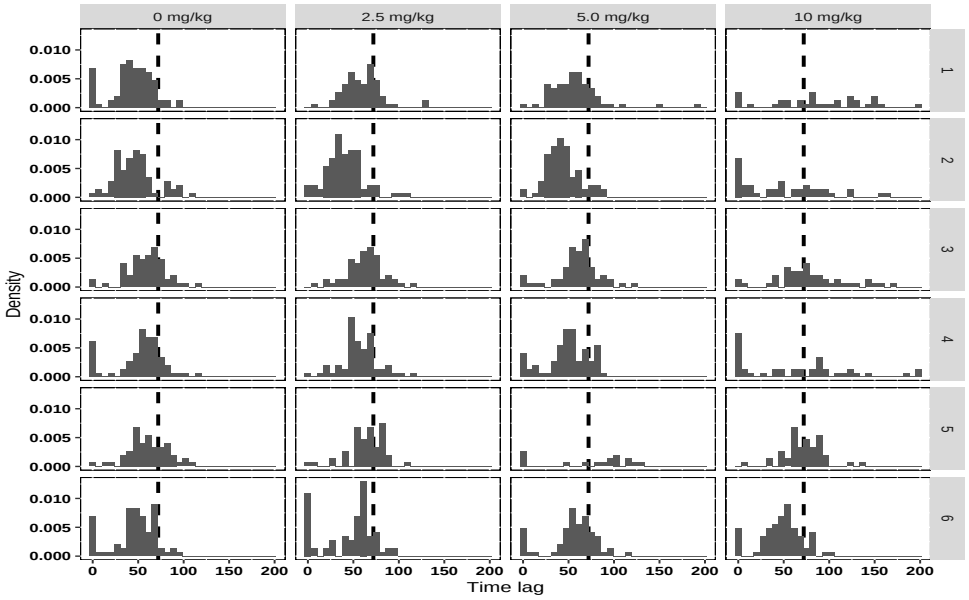


Figure 12.7: Histogram for the inter-response time distribution for the rat 1-6. The dashed vertical line denotes 72 seconds. There is clearly variability between rats in their IRT distribution and subsequently on the number of rewards obtained.

components or alternative distributions in the finite mixture, model convergence becomes a challenge. Moreover, with the finite mixture analysis, we did not obtain an association parameter such as the Pearson's correlation coefficient for the number of responses and number of rewards obtained as was done by Shkedy *et al.* (2005).

12.6 Appendix

12.6.1 Subject-Specific Density Plots for IRT

The individual density plots in the DRL-72 experiment are presented in Figure 12.7 to 12.9.

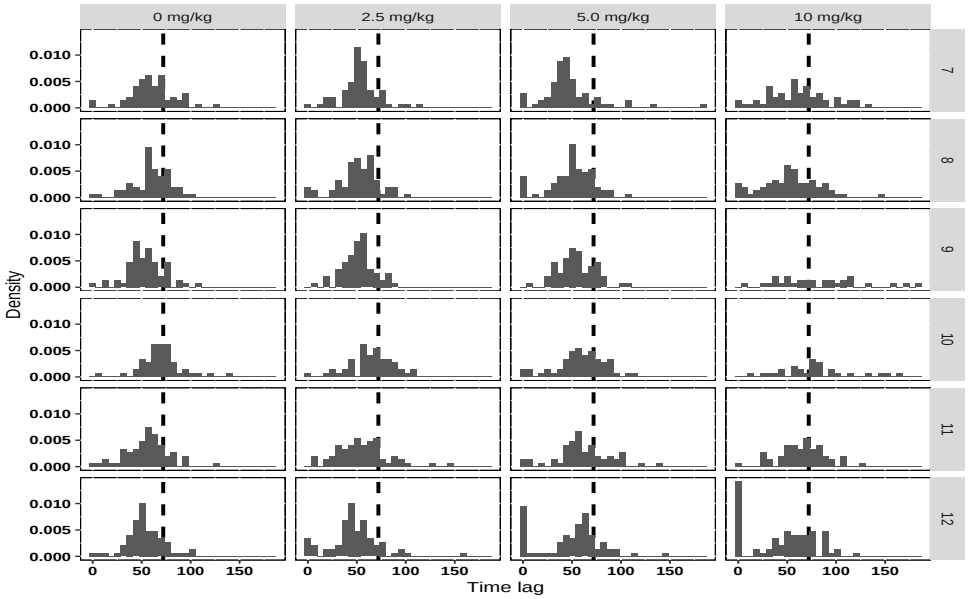


Figure 12.8: Histogram for the inter-response time distribution for the rat 7-12. The dashed vertical line denotes 72 seconds. There is clearly variability between rats in their IRT distribution and subsequently on the number of rewards obtained.

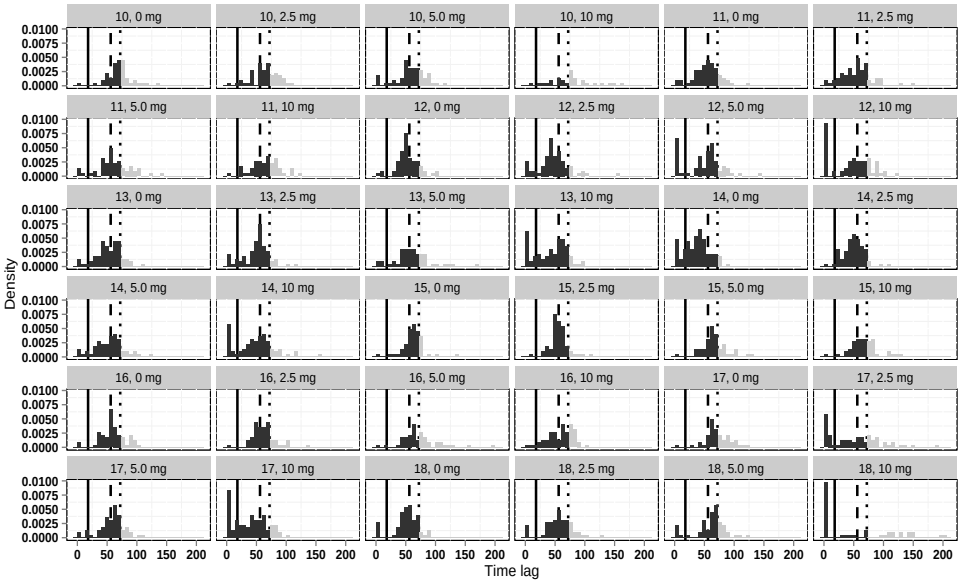


Figure 12.9: Histogram for the inter-response time distribution for the rat 13-18. The dashed vertical line denotes 72 seconds. There is clearly variability between rats in their IRT distribution and subsequently on the number of rewards obtained.

Table 12.6: Distribution functions evaluated for the finite mixture models.

Distribution (Notation)	Distribution	Median	Linear parameter
Lognormal	$f(y) = \frac{1}{yS\sqrt{2\pi}} e^{-0.5\left(\frac{log(y)-M}{S}\right)^2}$	$e^{(M)}$	M
Exponential	$f(y) = \frac{1}{\psi} e^{\left(\frac{y}{\psi}\right)}$	$\psi \log(2)$	ψ
Weibull	$f(y) = \frac{\kappa}{\lambda}\left(\frac{y}{\lambda}\right)^{\kappa-1} e^{\left(\frac{y}{\lambda}\right)^{\kappa}}$	$\lambda \log(2)^{\frac{1}{\kappa}}$	λ

12.6.2 Distribution Functions for the Finite Mixture Components

The parametrization adopted for the densities used in fitting the finite mixture model is shown in Table 12.6. As mentioned in Section 12.4.1, the location parameter depends on dose, duration in the study and subject-specific effects (see for example, Equation 12.6 for the lognormal model).

12.6.3 Software Issues

In this section, we present the SAS code used to fit the finite mixture,
$Y_{ijk} \sim \pi_j * LogN\left(M_{1ijk}, S_{1j}\right) + (1-\pi_j) * LogN\left(M_{2ijk}, S_{2j}\right).$

First, the data is sorted by the subject ID.

```
PROC SORT data=survival;BY animal dose;
RUN;
```

The finite mixture model formulated in Section 12.4.1 can be fitted using `PROC NLMIXED`. A general call of the procedure for a two-component mixture model has the form

```
PROC NLMIXED DATA = ;

Formulation of the mean structure for the location parameters (per
    component);

Definition of the mixing probability;

Definition of the scale parameters for the components;

Definition of the density for the components;

Specification of the finite mixture density;

General likelihood specification;
```

For the lognormal-lognormal model, we used:

```
    ODS OUTPUT PARAMETERESTIMATES=lognormal_LOGNORMAL_dose_Fixed_est;
 PROC NLMIXED DATA =survival GCONV=0 ;
```

The mean structure for the location parameters was specified in SAS using the following statements.

```
M1=alpha11+alpha12*dose25+alpha13*dose50+alpha14*dose100
   +beta11*logtstudy+beta12*dose25*logtstudy+beta13*dose50*logtstudy
   +beta14*dose100*logtstudy +a;

M2=alpha21+alpha22*dose25+alpha23*dose50+alpha24*dose100
   +beta21*logtstudy+beta22*dose25*logtstudy+beta23*dose50*logtstudy+
   beta24*dose100*logtstudy +b;
```

In order to obtain dose-specific mixing probability and dose-specific scale parameters for the two components respecitvely, we define,

```
*Dose specific mixing probabiity;
pi=exp(zp0+zp25*dose25+zp50*dose50+zp10*dose100)/
   (1+exp(zp0+zp25*dose25+zp50*dose50+zp10*dose100));

*Dose specific scale parameters for each of the lognormal components;
sigmares1=sigma1_0*(dose=0)+sigma1_25*(dose=2.5)+sigma1_50*(dose=5)+
sigma1_10*(dose=10);

sigmares2=sigma2_0*(dose=0)+sigma2_25*(dose=2.5)+sigma2_50*(dose=5)+
sigma2_10*(dose=10);
```

The two-component lognormal-lognormal finite mixture likelihood is defined using the `general likelihood` specification in `PROC NLMIXED`:

```
*Lognormal denstity for the first component;
dens1=( exp(-0.5*((log(time) -u1)**2)/exp(2*sigmares1))/
((time*(2*CONSTANT('PI')*exp(2*sigmares1))**0.5)));

*Lognormal density for the second component;
dens2=( exp(-0.5*((log(time) -u2)**2)/exp(2*sigmares2))/
((time*(2*CONSTANT('PI')*exp(2*sigmares2))**0.5)));

*Specify the finite mixture density;
fg1 = (pi)*dens1+(1-pi)*dens2;
ll=log(fg1);

*General likelihood specification in SAS PROC NLMIXED;
 MODEL time ~ GENERAL(ll);
```

Random effects variance-covariance matrix is specified using the `RANDOM` statement as follows:

```
RANDOM a b ~ normal([0,0],[sigmaa2,cov,sigmab2]) subject=animal;
OUT=Random_effects_est;
```

Chapter 13

Discussion and Future Research

13.1 Diffusion Kurtosis Magnetic Resonance Imaging in Neurodevelopment and Neurodegeneration

One of the growing interests in Alzheimer's Disease (AD) research is the identification and validation of appropriate diagnostic tools for early detection of AD. Considering the fact that AD is an irreversible condition, early diagnosis is paramount if interventions which can improve the patient's quality of life are to be effective. Moreover, a validated diagnostic tool for early onset of AD could allow for more targeted disease management rather than the current situation in which patients are managed for other dementia-related symptoms and only until later stages, are they diagnosed as AD patients.

Identification and validation of biomarkers for AD should be focused on two main aspects:

1. The ease of acquisition and safety profile of a biomarker. In the case of AD, although promising, invasive biomarkers such as cerebral-spinal fluid may hinder their adoption for clinical follow-up. However, if validated, non-invasive biomarkers such as the MRI parameters, evaluated in this dissertation, are more appropriate for clinical follow-up, since they can be taken repeatedly on a patient without causing excessive discomfort/risks.
2. A good biomarker should be valid in predicting a patient's disease pathology (individual-level surrogacy) and also be valid in monitoring the disease progression (disease-level surrogacy). This implies that, for a particular patient, a biomarker should not only be good at predicting the **true disease pathology at any given time**, but also, based on several biomarker measurements, the biomarker should predict the **true disease progression**.

The analysis presented in this dissertation showed that, the validity of MRI parameters as biomarkers for the true disease pathology depends on the MRI parameter of interest, the histology stain of interest and the brain region.

- The specific MRI parameter being evaluated. In general, DKI parameters (especially MRI-AK) were good surrogates for histology stains.
- The specific histology stain considered. Different histology stains quantify different aspects of the disease pathology. Therefore, not all MRI parameters are good biomarkers for all histology stains.
- The brain region of interest. The brain is a heterogeneous organ with white and grey matter regions. Some of the symptoms of AD include lack of motor coordination and difficulties with speech. These symptoms do not necessarily manifest simultaneously, which suggests that the disease pathology spreads heterogeneously across the brain. Therefore, in evaluating the validity of a given MRI parameter as a biomarker for a given histology stain, different brain regions should be evaluated.

13.1.1 Topics for Further Research

Although a comprehensive analysis of the MRI data was presented in this dissertation, several issues, listed below, were encountered but were not fully addressed in this dissertation.

Multivariate Classification for the MRI Data

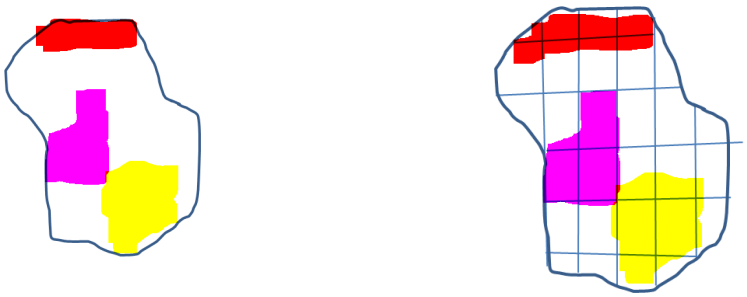
Although longitudinal data for MRI was available at 2, 4, 6 and 8 months of age, the linear discriminant analysis presented in Chapter 4 was performed for each age separately. Accounting for the evolution of the MRI parameters in a classification rule may improve the classification accuracy. In this regard, we propose the following approaches to develop a classifier with the longitudinal data.

- Functional Linear Discriminant Analysis (FLDA, James and Hastie, 2001) provides an extension of classical linear discriminant analysis to datasets where the predictor variables are curves or functions. The procedure uses spline curves and a random error term to model the individual profiles. Thus, with FLDA, new curves (evolution profiles) can be classified into one of the latent classes based on a discriminant function.
- There exists methods for classification of data from multiple sources such as multiple factor analysis (Kasim *et al.*, 2016). Adoption and extension of such ideas to longitudinal data may provide alternative tools with better classification accuracy.

Evaluation of MRI Parameters as Biomarkers for AD

- The analysis presented in Chapter 6 and 7 was performed on different brain Regions of Interest (ROI). An illustration of a hypothetical brain region is presented in Figure 13.1. Within a region of interest, the AD pathology does not necessarily spread homogeneously. Therefore, a ROI-based analysis represents an average disease effect in that region (averaged over all the voxels in the ROI) and thus ignores potential "hotspots" within the region. A voxel based analysis could potentially provide additional sensitivity in validating biomarkers for AD. The challenge

with voxel-based analyis is that, while there is a one-to-one mapping between MRI parameters and histology staining at the voxel level within an animal, there is no correspondence of voxel locations across different animals. Therefore, the modelling approach presented in this dissertation, which relies on data from several animals cannot be applied to the voxel data.



(a) Illustration of a region of interest **(b)** Illustration of the ROI divided into voxels

Figure 13.1: Illustration of the ROI and voxel based analysis. The shaded regions represents disease pathology as quantified by different histology stains

.

- The ROI based analysis presented in this dissertation also ignores the fact that MRI parameters for a given ROI are an average over all the voxels in that region. This essentially ignores the variability of the MRI scores within a region. Incorporating variability of the MRI parameters in the classification and biomarker evaluation could potentially increase the accuracy. A hierarchical Bayesian approach where the variance of MRI parameters is incorporated as informative priors or as an extra hierarchy in the model, can be considered. Alternatively, methods for statistical analysis for covariates with measurement error can be explored.

- Joint modelling of multiple MRI parameters for one histology stain and vice versa could also be of interest. The MRI parameters may provide better characterization of particular disease pathology when evaluated jointly, considering the fact that AD is a culmination of multiple changes in the disease pathology. While extension of the joint model presented in chapter 7 to multiple biomarkers is straight forward, the challenge is to obtain a single summary measure for the surrogacy for the multivariate case. Alonso *et al.* (2004) proposed the variance reduction factor for the case of validation of biomarkers in the context of repeated measures, which can form a basis for the extension of the analysis presented in this dissertation.

- The methodology presented in Chapter 7 was based on a joint model for the disease effects in order to evaluate the disease-level surrogacy. Rather than impose a joint distribution for the disease effects α_j and β_j, conditional distribution of the effect of a histology feature given the effect of an MRI parameter can be specified. In particular, let μ_{cond} be the conditional mean and σ^2_{cond} the conditional variance. The conditional distribution of the disease effect on the histology given the disease effect on the biomarker is given by,

$$\beta_j|\alpha_j \sim Normal\left(\mu_{cond}, \sigma^2_{cond}\right). \tag{13.1}$$

13.2 Joint Modelling of Correlated Data: Overdispersion, Bayesian Variable Selection and Order-Restricted Hypotheses

Dose-response modelling is often performed for inference on the parameters of interest. In most cases, a set of models is considered for a particular dose-response experiment. Subsequently, the "best" model is chosen based on information criteria and based on the selected model, inference is conducted.

In this dissertation however, focus was not on hypothesis testing and inference, but rather, the uncertainty resulting from selection of a "best model" and parameter estimation following model selection. Instead of selecting one model out of a set of plausible models, we adapted Bayesian model averaging whereby, estimation of the parameters of interest is based on a weighted average of the estimates from all the models in the set. Moreover, to ease model fitting, Bayesian variable selection was adopted, thus allowing the fitting of all nested models in one run of the model. Additional considerations in dose-response modelling which we addressed in this dissertation included methods for accounting for overdispersion in Poisson or binomial dose-response outcomes and modelling approaches for order-restricted parameters.

13.2.1 Topics for Further Research

Several issues remain unresolved with regards to the dose-response methodology presented in this dissertation. In the following section, we present an overview of the main issues that need to be take into consideration in future.

Bayesian Variable Selection Model Comparison

- Classical Bayesian analysis uses DIC (Spiegelhalter *et al.*, 2002) and other measures such as the penalized expected deviance (Plummer, 2008) for comparison of competing models. While DIC evaluates an optimal balance between a model's *goodness-of-fit* and *complexity*, it is unclear how to compute DIC in the BVS context. For BVS models, the model complexity changes within the Multiple Chains Monte Carlo Simulations, therefore making it difficult to estimate the effective number of parameters.

- Similarly, for order-restricted models, the prior distribution is constrained for some parameters. This has an impact on the posterior distribution. Note that DIC is based on the assumption of the asymptotic multivariate normal distribution of the deviance which may be violated in such models. A further investigation of the properties of DIC for order-restricted models ought to be conducted.

- In the analysis presented in this dissertation, equal prior probabilities were assigned to the models. The complete influence of the priors on BVS parameter estimation and posterior model probability should be evaluated, considering the fact that in some cases, it would be of interest to assign informative priors for some of the models. Closely related to this point, the impact of specifying correlated priors for the BVS parameters should be investigated.

- Inference within the BVS framework was out of the scope of this dissertation. We have shown that model selection can be based on the maximum posterior model probability. However, the main question that remains unsolved is how to make a decision based on the posterior probability of the null model $P(g_0|data, g_0, \ldots, g_R)$. Ideally, for a given threshold τ, we would like to "reject" the null hypothesis of no-dose-effect whenever $P(g_0|data, g_0, \ldots, g_R) < \tau$. Neither the method to choose τ, nor the property of the decision rule (in terms of power and type I error), is clearly understood at this juncture.

Bibliography

Agronin, M. E. (2007) *Alzheimer Disease and Other Dementias: A Practical Guide*. Lippincott Williams & Wilkins.

Akaike, H. (1973) Information Theory and an Extension of the Maximum Likelihood Principle. In: *Second International Symposium on Information Theory* (Eds. B. N. Petrov and F. Csaki), 267–281. Budapest: Akadémiai Kiado.

Alexander, A. L., Lee, J. E., Lazar, M. and Field, A. S. (2007) Diffusion Tensor Imaging of the Brain. *Neurotherapeutics : the Journal of the American Society for Experimental NeuroTherapeutics*, **4**, 316–329.

Alonso, A., Geys, H., Molenberghs, G., Kenward, M. G. and Vangeneugden, T. (2004) Validation of Surrogate Markers in Multiple Randomized Clinical Trials with Repeated Measurements: Canonical Correlation Approach. *Biometrics*, **60**, 845–853.

Aregay, M., Shkedy, Z. and Molenberghs, G. (2015) Comparison of Additive and Multiplicative Bayesian Models for Longitudinal Count Data with Overdispersion Parameters: A Simulation Study. *Communications in Statistics - Simulation and Computation*, **44**, 454–473.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.

Blennow, K. (2004) Cerebrospinal Fluid Protein Biomarkers for Alzheimer's Disease. *NeuroRx : the Journal of the American Society for Experimental NeuroTherapeutics*, **1**, 213–25.

Breslow, N. E. (1984) Extra-Poisson Variation in Log-Linear Models. *Applied Statistics*, **33**, 38+.

Briggs, A. H., Weinstein, M. C., Fenwick, E. A., Karnon, J., Sculpher, M. J. and Paltiel, A. D. (2012) Model Parameter Estimation and Uncertainty Analysis: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Medical Decision Making*, **32**, 722–732.

Buckland, S. T., Burnham, K. P. and Augustin, N. H. (1997) Model Selection: an Integral Part of Inference. *Biometrics*, **53**, 603–618.

Burnham, K. and Anderson, D. (2003) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer New York.

Burzykowski, T., Molenberghs, G. and Buyse, M. (2005) *The Evaluation of Surrogate Endpoints*. Statistics for Biology and Health. Springer New York.

Buyse, M. and Molenberghs, G. (1998) Criteria for the Validation of Surrogate Endpoints in Randomized Experiments. *Biometrics*, **54**, 1014–1029.

Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D. and Geys, H. (2000) The validation of Surrogate Endpoints in Meta-analyses of Randomized Experiments. *Biostatistics*, **1**, 49–67.

Chetelat, G. and Baron, J.-C. C. (2003) Early Diagnosis of Alzheimer's Disease: Contribution of Structural Neuroimaging. *NeuroImage*, **18**, 525–541.

Cheung, M. M., Hui, E. S., Chan, K. C., Helpern, J. A., Qi, L. and Wu, E. X. (2009) Does dDiffusion kKurtosis Imaging Lead to Better Neural Tissue Characterization? A Rodent Brain Maturation Study. *NeuroImage*, **45**, 386 – 392.

Claeskens, G. and Hjort, N. (2008) *Model Selection and Model Averaging*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Cooke, R. (2009) *Uncertainty Modeling in Dose Response: Bench Testing Environmental Toxicity*. Statistics in Practice. Wiley.

Cripps, E., Carter, C. and Kohn, R. (2005) Variable Selection and Covariance Selection in Multivariate Regression Models. In: *Handbook of Statistics: Bayesian Thinking, Modeling and Computation* (Eds. D. Dey and C. Rao), vol. 25 of *Handbook of Statistics*. Elsevier Science.

Del Fava, E., Shkedy, Z., Aregay, M. and Molenberghs, G. (2014) Modelling multivariate, Overdispersed Binomial Data with Additive and Multiplicative Random Effects. *Statistical Modelling*, **14**.

Dellaportas, P., Forster, J. and Ntzoufras, I. (2002) On Bayesian Model and Variable Selection Using MCMC. *Statistics and Computing*, **12**, 27–36.

Dickerson, B. C. and Sperling, R. A. (2005) Neuroimaging Biomarkers for Clinical Trials of Disease-modifying Therapies in Alzheimer's Disease. *NeuroRx*, **2**, 348–360.

Diggle, P. (1990) *Time Series: A Biostatistical Introduction*. Oxford science publications. Clarendon Press.

Doecke, J. D., Laws, S. M., Faux, N. G., Wilson, W., Burnham, S. C., Lam, C.-P., Mondal, A., Bedo, J., Bush, A. I., Brown, B., De Ruyck, K., Ellis, K. A., Fowler, C., Gupta, V. B., Head, R., Macaulay, S. L., Pertile, K., Rowe, C. C., Rembach, A., Rodrigues, M., Rumble, R., Szoeke, C., Taddei, K., Taddei, T., Trounson, B., Ames, D., Masters, C. L. and Martins, R. N. (2012) Blood-based Protein Biomarkers for Diagnosis of Alzheimer Disease. *Archives of Neurology*, **69**, 1318–25.

Duff, K. and Suleman, F. (2004) Transgenic Mouse Models of Alzheimer's Disease: How Useful Have They Been for Therapeutic Development? *Briefings in Functional Genomics & Proteomics*, **3**, 47–59.

Evenden, J., Ryan, C. and Matilla, M. (1993) Behavioral Testing of Antidrpressants: a Practical Pre-clinical Approach to Clinical Problems. In: *Behavioural neuroscience: a practical approach* (Ed. A. Sahgal), vol. 2 of *Practical Approach Series*. Oxford University Press.

Fisher, R. A. (1936) The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, **7**, 179–188. URL `http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x`.

Fletcher, L. C. B., Burke, K. E., Caine, P. L., Rinne, N. L., Braniff, C. A., Davis, H. R., Miles, K. A. and Packer, C. (2013) Diagnosing Alzheimer's Disease: Are We Any Nearer to Useful Biomarker-based, Non-invasive Tests? *GMS Health Technology Assessment*, **9**, Doc01.

Freedman, L. S., Graubard, B. I. and Schatzkin, A. (1992) Statistical Validation of Intermediate Endpoints for Chronic Diseases. *Statistics in Medicine*, **11**, 167–78.

Galvin, J. E. and Sadowsky, C. H. (2012) Practical Guidelines for the Recognition and Diagnosis of Dementia. *Journal of the American Board of Family Medicine : JABFM*, **25**, 367–82.

Gareth M. James, T. J. H. (2001) Functional Linear Discriminant Analysis for Irregularly Sampled Curves. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **63**, 533–550.

Gelman, A. and Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research. Cambridge University Press.

Gelman, A. and Rubin, Donald, B. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–511.

George, E. I. and McCulloch, R. E. (1993) Variable Selection via Gibbs Sampling. *Journal of the American Statistical Association*, **88**, 881–889.

Götz, J. and Götz, N. N. (2009) Animal Models for Alzheimer's Disease and Frontotemporal Dementia: a Perspective. *ASN NEURO*, **1**, e00019.

Hampel, H., Broich, K., Hoessler, Y. and Pantel, J. (2009) Biological Markers for Early Detection and Pharmacological Treatment of Alzheimer's Disease. *Dialogues in Clinical Neuroscience*, **11**, 141–157.

Hardy, J. and Selkoe, D. J. (2002) The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science*, **297**, 353–356.

Hastie, T., Tibshirani, R. and Friedman, J. (2003) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, Corrected edn.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999) Bayesian Model Averaging: A Tutorial. *Statistical Science*, **14**, 382–401.

Hol, E. M., Roelofs, R. F., Moraal, E., Sonnemans, M. A. F., Sluijs, J. A., Proper, E. A., de Graan, P. N. E., Fischer, D. F. and van Leeuwen, F. W. (2003) Neuronal Expression of GFAP in Patients with Alzheimer Pathology and Identification of Novel GFAP Splice Forms. *Molecular Psychiatry*, **8**, 786–796.

Huang, A. and Wand, M. P. (2013) Simple Marginally Noninformative Prior Ddistributions for Covariance Matrices. *Bayesian Analysis*, **8**, 1–14.

Hui, E. S., Cheung, M. M., Qi, L. and Wu, E. X. (2008) Towards Better MR Characterization of Neural Tissues Using Directional Diffusion Kurtosis Analysis. *NeuroImage*, **42**, 122–134.

Humpel, C. (2011) Identifying and Validating Biomarkers for Alzheimer's Disease. *Trends in Biotechnology*, **29**, 26–32.

Jackson, A., Koek, W. and Colpaert, F. (1995) Can the DRL 72s Schedule Selectively Reveal Antidepressant Drug Activity? *Psychopharmacology*, **117**, 154–161.

James, G. M. and Hastie, T. J. (2001) Functional Linear Discriminant Analysis for Irregularly Sampled Curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **63**, 533–550.

Johnson, N. L., Kotz, S. and Balakrishnan, N. () *Continuous Univariate Distributions, Vol. 1 (Wiley Series in Probability and Statistics)*. Wiley-Interscience, volume 1 edn.

Jones, B. and Kenward, M. G. (1989) *Design and Analysis of Cross-over Trials*. Chapman and Hall.

Kasim, A., Shkedy, Z., Kaiser, S., Talloen, W. and Hochreiter, S. (2016) *Applied Biclustering Methods for Big and High Dimensional Data Using R*. Chapman and Hall/CRC Biostatistics Series. Taylor & Francis Group.

Kasim, A., Shkedy, Z. and Kato, B. S. (2012) Estimation and Inference Under Simple Order Restrictions: Hierarchical Bayesian Approach. In: *Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R*, 193–214. Springer.

Kato, B. S. and Hoijtink, H. (2006) A Bayesian Approach to Inequality Constrained Linear Mixed Models: Eestimation and Model Selection. *Statistical Modelling*, **6**, 231–249.

Klingenberg, B. (2009) Proof of Concept and Dose Estimation with Binary Responses Under Model Uncertainty. *Statistics in Medicine*, **28**, 274–292.

Klohs, J., Politano, I. W., Deistung, A., Grandjean, J., Drewek, A., Dominietto, M., Keist, R., Schweser, F., Reichenbach, J. R., Nitsch, R. M., Knuesel, I. and Rudin, M. (2013) Longitudinal Assessment of Amyloid Pathology in Transgenic ArcAβ Mice Using Multi-parametric Magnetic Resonance Imaging. *PLoS ONE*, **8**, e66097.

Klugkist, I. and Hoijtink, H. (2007) The Bayes Factor for Inequality and About Equality Constrained Models. *Comput. Stat. Data Anal.*, **51**, 6367–6379.

Klugkist, I., Kato, B. and Hoijtink, H. (2005) Bayesian Model Selection Using Encompassing Priors. *Statistica Neerlandica*, **59**, 57–69.

Kohama, S. G., Goss, J. R., Finch, C. E. and McNeill, T. H. (1995) Increases of Glial Fibrillary Acidic Protein in the Aging Female Mouse Brain . *Neurobiology of Aging*, **16**, 59 – 67.

Komárek, A., Hansen, B. E., Kuiper, E. M. M., van Buuren, H. R. and Lesaffre, E. (2010) Discriminant Analysis Using a Multivariate Linear Mixed Model with a Normal Mixture in the Random Effects Distribution. *Statistics in Medicine*, **29**, 3267–3283.

Lambert, D. (1992) Zero-inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics*, **34**, 1–14.

Lesaffre, E. and Lawson, A. (2012) *Bayesian Biostatistics*. Statistics in Practice. Wiley.

Lin, D., Shkedy, Z., Yekutieli, D., Amaratunga, D. and Bijnens, L. (2012) *Modeling Dose-response Microarray Data in Early Drug Development Experiments Using R: Order Restricted Analysis of Microarray Data*. Use R! Springer.

Lindsay, B. G. (1995) *Mixture Models: Theory, Geometry and Applications*. Penn. State University: NSF-CBMS Conference series in Probability and Statistics.

López-de-Ipiña, K., Alonso, J.-B., Travieso, C. M., Solé-Casals, J., Egiraun, H., Faundez-Zanuy, M., Ezeiza, A., Barroso, N., Ecay-Torres, M., Martinez-Lage, P. and Lizardui, U. M. d. (2013) On the Selection of Non-Invasive Methods Based on Speech Analysis Oriented to Automatic Alzheimer Disease Diagnosis. *Sensors*, **13**, 6730.

Masters, C. L., Simms, G., Weinman, N. A., Multhaup, G., McDonald, B. L. and Beyreuther, K. (1985) Amyloid Plaque Core Protein in Alzheimer Disease and Down Syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, **82**, 4245–4249.

McCullagh, P. and Nelder, J. (1989) *Generalized Linear Models*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D. and Stadlan, E. M. (1984) Clinical Diagnosis of Alzheimer's Disease: Report of the NINCDS-ADRDA Work Group Under the Auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, **34**, 939.

McLachlan, G. and Peel, D. (2000) *Finite Mixture Models, Willey Series in Probability and Statistics*. John Wiley & Sons, New York.

Mclachlan, G. J. (2004) *Discriminant Analysis and Statistical Pattern Recognition*. Wiley Series in Probability and Statistics. Wiley-Interscience.

Milanzi, E., Alonso, A. and Molenberghs, G. (2012) Ignoring Overdispersion in Hierarchical Loglinear Models: Possible Problems and Solutions. *Statistics in Medicine*, **31**, 1475–1482.

Molenberghs, G., Verbeke, G. and Deme'trio, C. (2007) An Extended Random-effects Approach to Modeling Repeated, Overdispersed Count Data. *Lifetime Data Analysis*, **13**, 513–531.

Molenberghs, G., Verbeke, G., Demètrio, C. G. B. and Vieira, A. M. C. (2010) A Family of Generalized Linear Models for Repeated Measures with Normal and Conjugate Random Effects. *Statistical science*, **25**, 325–347.

Muchene, L., Shkedy, Z., Arregay, Fantahun, M., Molenberghs, G., Jacobs, T. and Otava, M. (2016a) Hierarchical binomial-poisson overdispersed models for correlated dose-response data: A joint modelling approach. *Journal of Biopharmaceutical Statistics*, **Submitted**.

Muchene, L., Shkedy, Z., Bijnens, L., Manyakov, N., Van De Casteele, T., Jacobs, T., Verhoye, M., Praet, J., Van der Linden, A., Bottelbergs, A., Schmidt, M. and Pemberton, D. (2016b) *Applied Surrogate Endpoint Evaluation Methods with SAS and R*, chap. Evaluation of Magnetic Resonance Imaging as a Biomarker in Alzheimer's Disease.

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

O'Bryant, S. E., Gupta, V., Henriksen, K., Edwards, M., Jeromin, A., Lista, S., Bazenet, C., Soares, H., Lovestone, S., Hampel, H., Montine, T., Blennow, K., Foroud, T., Carrillo, M., Graff-Radford, N., Laske, C., Breteler, M., Shaw, L., Trojanowski, J. Q., Schupf, N., Rissman, R. A., Fagan, A. M., Oberoi, P., Umek, R., Weiner, M. W., Grammas, P., Posner, H. and Martins, R. (2015) Guidelines for the Standardization of Preanalytic Variables for Blood-based Biomarker Studies in Alzheimer's Disease Research . *Alzheimer's & Dementia*, **11**, 549 – 560.

O'Hara, R. and Sillanpaa, M. (2009) A Review of Bayesian Variable Selection Methods: What, How, and Which. *Bayesian Analysis*, **4**, 85–118.

Otava, M., Shkedy, Z., Lin, D., Gohlmann, H. W., Bijnens, L., Talloen, W. and Kasim, A. (2013) Dose-Response Modeling Under Simple Order Restrictions Using Bayesian Variable Selection Methods. *Statistics in Biopharmaceutical Research*, **0**, null.

Perl, D. P. (2010) Neuropathology of Alzheimer's Disease. *The Mount Sinai Journal of Medicine, New York*, **77**, 32–42.

Pinheiro, J., Bornkamp, B., Glimm, E. and Bretz, F. (2014) Model-based Dose Finding Under Model Uncertainty Using General Parametric Models. *Statistics in Medicine*, **33**, 1646–1661.

Plummer, M. (2008) Penalized Loss Functions for Bayesian Model Comparison. *Biostatistics*, **9**, 523–539.

Plummer, M. (2015) *Rjags: Bayesian Graphical Models Using MCMC*. R package version 3-15.

Prentice, R. L. (1989) Surrogate Endpoints in Clinical Trials: Definition and Operational Criteria. *Statist. Med.*, **8**, 431–440.

Radde, R., Bolmont, T., Kaeser, S. A., Coomaraswamy, J., Lindau, D., Stoltze, L., Calhoun, M. E., Jäggi, F., Wolburg, H., Gengler, S., Haass, C., Ghetti, B., Czech, C., Hölscher, C., Mathews, P. M. and Jucker, M. (2006) Aβ42-driven Cerebral Amyloidosis in Transgenic Mice Reveals Early and Robust Pathology. *EMBO reports*, **7**, 940–946.

Raftery, A. E. (1995) Bayesian Model Selection in Social Research. *Sociological Methodology*, **25**, 111–163.

Ramsay, J., Hooker, G. and Graves, S. (2009) *Functional Data Analysis with R and MATLAB*. Use R! Springer New York.

Richards, J. B., Sabol, K. E. and Seiden, L. S. (1993) DRL Interresponse-time Distributions: Quantification by Peak Deviation Analysis. *Journal of the Experimental Analysis of Behavior*, **60**, 361–385.

Sabbagh, M. N., Malek-Ahmadi, M., Kataria, R., Belden, C. M., Connor, D. J., Pearson, C., Jacobson, S., Davis, K., Yaari, R. and Singh, U. (2010) The Alzheimer's Questionnaire: A Proof of Concept Study for a New Informant-Based Dementia Assessment. *Journal of Alzheimer's disease : JAD*, **22**, 1015–1021.

Schwarz, G. (1978) Estimating the Dimension of a Model. *Ann. Statist.*, **6**, 461–464.

Serrano-Pozo, A., Frosch, M. P., Masliah, E. and Hyman, B. T. (2011) Neuropathological Alterations in Alzheimer Disease. *Cold Spring Harbor Perspectives in Medicine*, **1**, a006189.

Shkedy, Z. and Barbosa, F. T. (2005) Bayesian Evaluation of Surrogate Endpoints. In: *The Evaluation of Surrogate Endpoints* (Eds. T. Burzykowski, G. Molenberghs and M. Buyse), Statistics for Biology and Health. Springer New York.

Shkedy, Z., Molenberghs, G., Craenendonck, H. V., Steckler, T. and Bijnens, L. (2005) A Hierarchical Binomial-Poisson Model for the Analysis of a Crossover Design for Correlated Binary Data When the Number of Trials is Dose-dependent. *Journal of Biopharmaceutical Statistics*, **15**, 225–239.

Skellam, J. G. (1948) 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*, **10**, 257–261.

Snyder, H. M., Carrillo, M. C., Grodstein, F., Henriksen, K., Jeromin, A., Lovestone, S., Mielke, M. M., O'Bryant, S., Sarasa, M., Sjø gren, M., Soares, H., Teeling, J., Trushina, E., Ward, M., West, T., Bain, L. J., Shineman, D. W., Weiner, M. and Fillit, H. M. (2015) Developing Novel Blood-based Biomarkers for Alzheimer's Disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, **10**, 109–114.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002) Bayesian Measures of Model Complexity and Fit. *Journal of the royal statistical society: Series B (Statistical methodology)*, **64**, 583–639.

Stoline, M. R. (1993) Comparison of Two Medians Using a Two-sample Lognormal Model in Environmental Contexts. *Environmetrics*, **4**, 323–339.

Veraart, J., Poot, D. H. J., Van Hecke, W., Blockx, I., Van der Linden, A., Verhoye, M. and Sijbers, J. (2011) More Accurate Estimation of Diffusion Tensor Parameters Using Diffusion Kurtosis Imaging. *Magnetic Resonance in Medicine*, **65**, 138–145.

Veraart, J., Rajan, J., Peeters, R. R., Leemans, A., Sunaert, S. and Sijbers, J. (2013) Comprehensive Framework for Accurate Diffusion MRI Parameter Estimation. *Magnetic Resonance in Medicine*, **70**, 972–984.

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

Verbyla, A. P. and Cullis, B. R. (1990) Modelling in Repeated Measures Experiments. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **39**, 341–356.

Verdonck, A., De Ridder, L., Verbeke, G., Bourguignon, J., Carels, C., Kühn, E., Darras, V. and de Zegher, F. (1998) Comparative Effects of Neonatal and Prepubertal Castration on Craniofacial Growth in Rats . *Archives of Oral Biology*, **43**, 861 – 871.

Whitney, M. and Ryan, L. (2009) Quantifying Dose-response Uncertainty Using Bayesian Model Averaging. *Uncertainty Modeling in Dose Response: Bench Testing Environmental Toxicity*, 165–79.

Winblad, B., Amouyel, P., Andrieu, S., Ballard, C., Brayne, C., Brodaty, H., Cedazo-Minguez, A., Dubois, B., Edvardsson, D., Feldman, H., Fratiglioni, L., Frisoni, G. B., Gauthier, S., Georges, J., Graff, C., Iqbal, K., Jessen, F., Johansson, G., Jönsson, L., Kivipelto, M., Knapp, M., Mangialasche, F., Melis, R., Nordberg, A., Rikkert, M. O., Qiu, C., Sakmar, T. P., Scheltens, P., Schneider, L. S., Sperling, R., Tjernberg, L. O., Waldemar, G., Wimo, A. and Zetterberg, H. (2016) Defeating Alzheimer's Disease and Other Dementias: A Priority for European Science and Society. *The Lancet Neurology*, **15**, 455–532.

y Palacios, R. D., Campo, A., Henningsen, K., Verhoye, M., Poot, D., Dijkstra, J., Van Audekerke, J., Benveniste, H., Sijbers, J., Wiborg, O. *et al.* (2011) Magnetic resonance imaging and spectroscopy reveal differential hippocampal changes in anhedonic and resilient subtypes of the chronic mild stress rat model. *Biological psychiatry*, **70**, 449–457.

Zucchini, W. (2000) An Introduction to Model Selection. *J. Math. Psychol.*, **44**, 41–61.

Summary

Alzheimer's Disease (AD) is an age-dependent disease that has recently generated a lot of interest in the scientific community. The pathological changes induced by AD are irreversible and impacts on the quality of life of the patient especially because, AD affects the patient's motor coordination and cognitive ability, thereby rendering them dependent on care givers. The impact of AD on the well-being of patients is further complicated by the fact that, timely and accurate diagnosis of AD is currently lacking. Therefore, the main goal of most of the ongoing AD research is to identify biomarkers that will allow for accurate diagnosis of AD and monitoring of the disease progression in the development of treatments aimed at slowing or inhibiting the spread of AD pathology. For optimum clinical benefit, such biomarkers ought to be easy to acquire and preferably obtained non-invasively, or with minimum discomfort to the patient. Further, it is desirable that such a biomarker can accurately predict the AD pathology progression while the disease is still at an early stage. Currently, the true status of the disease pathology can only be known through a post-mortem examination, which is useless for the patient's management.

In the first part of this dissertation, we propose to use Magnetic Resonance Imaging (MRI) diffusion metrics as biomarkers for AD. MRI is an imaging technique that's already in clinical use for diagnosis and management of patients with brain trauma amongst other applications. MRI quantifies the flow of water molecules across the brain tissue. The diffusion of water molecules in a healthy brain is assumed to follow a Gaussian distribution, while any tissue malformations may induce deviations from the Gaussian flow of water molecules. The parameters derived from MRI include Diffusion Tensor Imaging (DTI) parameters and Diffusion Kurtosis Imaging (DKI) parameters. We used a mouse model for AD, with Amyloid Precursor Protein gene and Presenilin 1 genetic mutation (APP/PS1) in order to evaluate if MRI parameters can capture the AD pathology in mice brains. Note that this mouse model only mimics some aspects of AD in human patients, thereby allowing us to study specific AD pathology. AD however, is a manifestation of several pathological changes that sometimes occur simultaneously. The methodology presented in the first part of this dissertation addresses three broad goals with respect to AD.

First, can the evolution of AD in different regions in the brain be quantified using

MRI parameters? In Chapter 3, a linear model for the longitudinal MRI data is presented. The analysis enables us to identify the brain regions and age at which significant differences between healthy and APP/PS1 mice are observed. The cortex regions and in particular, the motor cortex was the region in which significant differences are observed especially at 6 and 8 months of age. In addition, DKI parameters (especially axial kurtosis and mean kurtosis) are more promising at discriminating between the two mice genotypes.

Secondly, which set of MRI parameters between DTI and DKI parameters can discriminate between diseased and healthy mice? More so, does DKI parameters improve the classification of disease status compared to classification based on DTI alone? Chapter 4 addresses this classification problem using linear discriminant analysis which is performed on three classifiers namely; DTI parameters, DKI parameters and a combination of both DTI and DKI parameters for each of the brain regions separately. As expected, as the disease pathology develops with age, the misclassification error reduces for the three classifiers. The classifier based on either DTI or DKI has comparable misclassification error for all ages while a classifier based on both DTI and DKI substantially reduces the misclassification error. Note that all these results vary with the brain region although the motor cortex shows the best discrimination ability between the two mice genotypes.

Finally, Chapters 5-7 are dedicated to the evaluation of MRI parameters as biomarkers for the AD pathology. We adopt the methodology for surrogate markers evaluation commonly used in clinical trials. While the setting posed by the MRI experiment presented in this dissertation is different, the methodology is still useful and is applied with minimal modifications. In the context of AD, rather than evaluating whether the treatment effect on the biomarker is predictive for the *treatment effect* on the true endpoint, we are interested in evaluating whether the disease effects on the biomarker are predictive for the *disease effects* on the true endpoint. Moreover, rather than multiple centres or trials as is the case in clinical trials, in the MRI experiment, these are substituted with mice cohorts of ages 2, 4, 6, 8 and 10 months. We present a frequentist and Bayesian approach to biomarker evaluation which are applied to the different combinations of MRI parameters, pathological histology stains and brain regions. Results indicate that disease-level surrogacy of different MRI parameters varies with the specific histology stain used and the brain region. Note that for some regions such as the amygdala and olfactory bulb, none of the MRI parameters shows evidence of being a good surrogate for the four histology stains.

The results presented in the first part of this dissertation indicate that although MRI parameters have potential to be surrogates for disease pathology in AD, there are still issues that ought to be addressed before they can be adopted for clinical use. The accuracy of disease-level surrogacy measures presented here is relatively low, possibly due to the small sample sizes. Simulation studies showed that the sample size used for the estimation of disease-level surrogacy has an impact on the accuracy of the estimated correlations. We presented a region-level analysis but potentially, a model based on data at the voxel-level resolution might be more interesting. Moreover, the brain is heterogeneous and the spatial arrangement of the different regions may have an impact on the spread of the disease pathology that should be accounted for in the analysis.

The focus of the second part of this dissertation was on dose-response models. In dose-response modelling, it is often the case that a choice has to be made of the "best" model on which to base subsequent estimation and inference. The set of models on which to choose the best models can result from different formulations of the model, for instance, linear and non-linear dose-response profiles or in some cases, by the exclusion of a subset of parameters from the "full" model. Statisticians then use information criteria such as the Akaike Information Criterion (AIC) , Bayesian Information Criterion (BIC) or Deviance Information Criterion (DIC) amongst others in order to decide on the best fitting model for the data. Once this choice of the "best" model is made, subsequent analysis and inference proceed completely ignoring the uncertainty introduced by model selection step.

Chapter 8-12 of this dissertation is dedicated to addressing the uncertainty in model selection in several dose-response experiments. The overarching theme is Bayesian model averaging whereby, rather than select a single "best" model, we propose to fit the set of all models simultaneously and compute a model-averaged estimate of the parameters of interest over all the models. This is implemented within the Bayesian framework whereby, we propose to use Bayesian Variable Selection (BVS) for the dose-response case studies presented in this dissertation. Note that the BVS approach presented in this dissertation is appropriate for cases in which the models are nested within one "full" model and the parameters for which BVS is to be performed are categorical. By specifying appropriate indicators, BVS allows for selection of a subset of the parameters to be included in the model during each iteration of the model fit. For each model, the posterior model probability is computed which is used as the model weight in computing the model-averaged estimates. Further, the posterior model probability is a measure of relative importance of each of the models in the set.

The advantage of BVS is that, the methodology can be applied only to a subset of the parameters (such as the dose parameters) while some other parameters are always included in the model (such as the random effects). Moreover, by using BVS other design considerations, for instance, order restriction of the dose parameters can be incorporated in the model. We applied BVS to several case studies of varying complexity. Both univariate models and joint models were considered. The Bayesian framework allows us to also investigate additional design issues such as overdispersion in Poisson and binomial distributions as well as order-restriction. Although we only performed estimation for the dose parameters of interest, inference in the context of BVS can be investigated. Further, the posterior model probability may be useful in performing model selection post-estimation, thereby taking into account the uncertainly unlike in the case of model selection pre-estimation where selection precedes estimation of the parameters of interest.

A few issues still need to be addressed with regards to BVS. First, since the model fitted changes in each iteration, classical Bayesian measures of goodness of fit such as DIC are no longer valid hence, an appropriately defined information criterion is needed. This is especially so when models incorporating BVS have to be compared with models without BVS. Secondly, while the posterior model probability is an indicator of the relative importance of the models, it is not clear what the impact of using this measure for model

selection may have. Finally, there is need for a simulation study to evaluate the impact of the model priors on the posterior model probability.

Samenvatting

De Ziekte van Alzheimer (ZA) is een leeftijd gerelateerde aandoening die recent aan enorme belangstelling heeft gewonnen binnen de wetenschappelijke gemeenschap. De pathologische veranderingen, veroorzaakt door ZA, zijn onomkeerbaar en hebben een invloed op de kwaliteit-van-leven van de patiënt. Dit komt omdat ZA een impact heeft op zowel de motorische coördinatie als de cognitieve mogelijkheden van patiënten, waardoor deze afhankelijk worden van zorgverleners. De invloed van ZA op het welbevinden van patiënten wordt daarenboven vergroot omdat tijdige en accurate diagnose van de ziekte voorlopig ontbreekt. Daarom is huidig ZA onderzoek vooral gericht op de identificatie van biomerkers die zowel een accurate diagnose van ZA als de opvolging van de ziekte in termen van ziekteprogressie toelaten. Dit heeft als doel om de ontwikkeling van behandelingen, die de symptomen van ZA kunnen vertragen of zelfs voorkomen, te faciliteren. Om klinisch optimaal te zijn, dienen zulke biomerkers gemakkelijk en bij voorkeur non-invasief te bekomen zijn, om zodoende het ongemak van de patiënt tot een minimum te beperken. Daarnaast is het wenselijk dat zulke biomerkers pathologische ZA progressie accuraat kunnen voorstellen wanneer de ziekte zich nog in een vroeg stadium bevindt. Op dit moment kan de ware status van de ziekte pathologie enkel post-mortem worden vastgesteld. Dit is vruchteloos vanuit een patiëntbeheer standpunt.

In het eerste deel van deze thesis onderzoeken we Magnetic Resonance Imaging (MRI) diffusie indicatoren als biomerkers voor ZA. MRI is een beeldvormingstechniek die onder andere al wordt gebruikt voor de diagnose en opvolging van patiënten met hersentrauma. MRI kwantificeert de stroom van watermoleculen doorheen hersenweefsel. Van de diffusie van watermoleculen in gezonde hersenen wordt aangenomen dat deze een Gaussiaanse distributie volgt, terwijl weefsel-misvorming kan zorgen voor deviaties van de Gaussiaanse stroom van watermoleculen. De parameters bekomen na MRI omvatten Diffusion Tensor Imaging (DTI) en Diffusion Kurtosis Imaging (DKI). We hebben gebruik gemaakt van een muismodel voor ZA, met Amyloid Precursor Protein gen en Presenilin 1 genetische mutatie (APP/PS1), om te onderzoeken of MRI parameters ZA pathologie kunnen oppikken in muishersenen. Merk op dat dit muismodel enkel bepaalde aspecten van ZA in mensen nabootst en dus toelaat om enkel een specifieke ZA pathologie te bestuderen. ZA komt tot uiting in verschillende pathologische veranderingen die mogelijks terzelfdertijd optreden. De besproken methodologie in het eerste deel van de

thesis richt zich op drie algemene doelen met betrekking tot ZA.

Ten eerste; kan de evolutie van ZA in verschillende delen in de hersenen gekwantificeerd worden, gebruik makend van MRI parameters? In hoofdstuk 3 wordt een lineair model voor longitudinale MRI data voorgesteld. Deze analyse laat ons toe om de hersengebieden te identificeren en de leeftijd te voorspellen, waar en wanneer significante verschillen tussen gezonde en APP/PS1 muizen geobserveerd worden. De cortex regio's in het algemeen, en de motor cortex in het bijzonder, waren de regio's waar significante verschillen worden geobserveerd vooral op 6 en 8 maanden. Bijkomend wordt vastgesteld dat DKI parameters (vooral axiale kurtosis en gemiddelde kurtosis) veelbelovend zijn om te discrimineren tussen de twee muis genotypes.

Ten tweede; welke verzameling van MRI parameters tussen DTI en DKI parameters kunnen discrimineren tussen zieke en gezonde muizen? In het bijzonder, leidt de toevoeging van DKI parameters tot een betere classificatie van de ziekte status vergeleken met de classificatie op basis van enkel DTI. Hoofdstuk 4 behandelt dit classificatie probleem gebruik makend van lineaire discriminant analyse. Deze analyse wordt toegepast op drie classificeerders, namelijk; DTI parameters, DKI parameters en een combinatie van zowel DTI als DKI parameters, apart voor ieder hersengebied. Zoals verwacht verkleint de misclassificatie fout van alle classificeerders naarmate de ziekte pathologie verergert met de leeftijd. De classificeerder gebaseerd op enkel DTI of DKI heeft een vergelijkbare misclassificatie fout voor alle leeftijden. Aan de andere kant verkleint een classificeerder gebaseerd op zowel DTI als DKI, de misclassificatie fout substantieel. Merk op dat alle resultaten variëren op basis van hersengebied, hoewel de motor cortex tot de beste discriminantie tussen de twee muis genotypes leidt.

Tenslotte zijn hoofdstukken 5 tot 7 gewijd aan de evaluatie van MRI parameters als biomerkers voor ZA pathologie. We nemen de methodologie voor de evaluatie van surrogaat merkers over. Deze wordt algemeen gebruikt in klinische studies. Hoewel de setting van het MRI experiment, zoals besproken in deze thesis, hiervan verschilt, is de methodologie nog steeds nuttig en kan ze worden toegepast na minimale aanpassing. In de context van ZA gaan we na of de ziekte-effecten op de biomerkers de ziekte-effecten op het ware eindpunt kunnen voorspellen, in plaats van, of het behandelingseffect op de biomerker de behandelingseffecten kunnen voorspellen op het ware eindpunt. Meer nog, in plaats van meerdere centra of studies, zoals gebruikelijk in klinische studies, worden deze in het MRI experiment vervangen door muiscohorten van 2, 4, 6, 8, en 10 maanden. We presenteren zowel een Frequentiste als Bayesiaanse benardering tot biomerker evaluatie, toegepast op de verschillende combinaties van MRI parameters, pathologische histologie kleuring en hersengebieden. Resultaten tonen aan dat ziekte-niveau surrogaatheid van verschillende MRI parameters varieert met de specifiek gebruikte histologiekleuring en het hersengebied. Merk op dat voor sommige regio's, zoals de amygdala en olfactorische bulb, geen van de MRI parameters blijk geven goede surrogaten te zijn voor de vier histologiekleuringen.

De gepresenteerde resultaten in het eerste deel van deze thesis tonen aan dat hoewel de MRI parameters potentieel hebben om surrogaten voor ziektepathologie in ZA te

zijn, er nog steeds kwesties zijn die dienen bekeken te worden vooraleer ze kunnen aangepast worden voor klinisch gebruik. De accuraatheid van de ziekte-niveau surrogaat maten die hier gepresenteerd worden, zijn relatief laag, mogelijks kan dit toegewezen worden aan de kleine steekproefgrootte. Simulatiestudies toonden aan dat de gebruikte steekproefgrootte voor de schatting van ziekte-niveau surrogaatheid een impact had op de accuraatheid van de geschatte correlaties. We toonden een regio-niveau analyse, maar mogelijks is een model gebaseerd op data op voxel-niveau resolutie interessanter. Meer nog, de hersenen zijn heterogeen, en de spatiale schikking van de verschillende regio's kan mogelijk een impact hebben op de spreiding van de ziekte pathologie waarmee rekening gehouden dient te worden in de analyse.

In het tweede deel van deze thesis werd dieper ingegaan op dosis-respons modellen. In dosis-respons modellen is het vaak het geval dat een 'best' model gekozen dient te worden. Dit model zal dan vervolgens dienen als basis van schatting en inferentie. De verzameling van modellen waaruit de beste modellen gekozen kunnen worden, kan gebaseerd zijn op verschillende formuleringen van het model. Dit kan, bijvoorbeeld, door de aanname van lineaire of non-lineaire dosis-respons profielen of in sommige gevallen, door de uitsluiting van een subset van de parameters uit het 'full' model. Statistici gebruiken in dit geval informatie criteria, zoals onder andere, het Akaike Informatie Criterium (AIC), Bayesiaans Informatie Criterium (BIC) of Deviantie Informatie Criterium (DIC) om tot een best fittend model te komen. Eenmaal de keuze voor het 'beste' model gevallen is, gaat men verder met analyse en inferentie, waarbij men de onzekerheid geïntroduceerd door de model-selectie-stap volledig negeert.

Hoofdstukken 8-12 van deze thesis zijn gewijd aan het in rekening brengen van de onzekerheid in model selectie in verschillende dosis-respons experimenten. Het overkoepelende thema is Bayesian model averaging, waarbij in plaats van de selectie van één 'beste' model, we voorstellen om de verzameling van alle modellen simultaan te fitten, waarna men een model-gemiddelde schatting over alle modellen bekomt voor de parameters van interesse. Deze methode wordt geïmplementeerd in een Bayesiaans kader, waarbij we voorstellen om gebruik te maken van Bayesian Variable Selection (BVS) in de dosis-respons studies gepresenteerd in deze thesis. Merk op dat de voorgestelde BVS benadering geschikt is voor gevallen waarin de modellen genest zijn binnen één 'full' model en de parameters, waarvoor BVS is verreist, categorisch zijn. Door de specificatie van geschikte indicatoren laat BVS toe om tijdens iedere model fit iteratie een subset te selecteren van parameters die in het model dienen te worden opgenomen. Voor ieder model wordt de posterior model kans berekend. Deze kansen worden vervolgens gebruikt als model gewichten om tot de model-gemiddelde schattingen te komen. Meer nog, de posterior model kansen zijn een maat voor de relatieve belangrijkheid van elk van de modellen in de set.

Het voordeel van BVS is dat de methode enkel kan toegepast worden op een subset van de parameters (zoals bv. dosis parameters) terwijl andere parameters altijd in het model worden opgenomen (zoals bv random effecten). Daarnaast is het zo dat door het gebruik van BVS ook andere ontwikkelingsconsideraties, zoals bijvoorbeeld, volgorde restricties van de dosis parameters kunnen worden geïncorporeerd

in het model. We hebben het BVS toegepast op verschillende studies van variërende complexiteit. Zowel univariate als joint modellen werden beschouwd. Het Bayesiaans kader laat ons toe om ook bijkomende design kwesties, zoals over-dispersie in Poisson en binomiale distributies, alsook volgorde-restrictie, te onderzoeken. Hoewel we enkel schattingen bepaalden voor de dosis parameters van interesse, kan inferentie in de context van BVS ook onderzocht worden. Verder zou ook de posterior model kans nuttig kunnen zijn in de uitvoering van post-schatting-model-selectie. Hierbij zouden we al de onzekerheid in rekening kunnen brengen, in vergelijking met het pre-schatting-model-selectie, in welke selectie voorafgaat aan schatting van de parameters van interesse.

Enkele kwesties dienen nog steeds te worden onderzocht betreffende BVS. Ten eerste, aangezien het model dat gefit wordt verschillend is in iedere iteratie, zijn de klassieke Bayesiaanse maten voor goodness-of-fit, zoals DIC, niet meer geschikt. Hierdoor hebben we nood aan een overeenkomstig gespecifieerd informatie criterium. Dit is vooral belangrijk wanneer BVS modellen vergeleken worden met modellen niet bekomen met BVS. Ten tweede, terwijl posterior model kans een indicator is van het relatieve belang van modellen, is het niet duidelijk wat de impact van het gebruik van deze maat op model selectie is. Tenslotte is er nood aan een simulatie studie die de impact van model priors op de posterior model kans onderzoekt.