Statistical Models for Count Data from Multiple Sclerosis Clinical Trials and their Applications

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the

Graduate School of The Ohio State University

By

Mallikarjuna Rao Rettiganti, M.Sc.

Graduate Program in Statistics

The Ohio State University

2010

Dissertation Committee:

Haikady N. Nagaraja, Advisor

Jason C. Hsu

Eloise Kaizar

Thomas J. Santner

ABSTRACT

Multiple sclerosis (MS) is an autoimmune disease in which the body's own immune system attacks the central nervous system. Relapsing remitting MS (RRMS) is an initial stage of the disease where the patient experiences distinct phases of relapse and remittance. Magnetic resonance imaging (MRI) is commonly used to monitor the RRMS disease progression. MRI scans of the brain are taken each month and the total number of new MRI lesions seen during the follow-up period is used as the response variable of interest. The Negative Binomial (NB) and the Poisson-Inverse Gaussian (P-IG) distributions have been shown to fit this over-dispersed data well. Currently, only nonparametric tests are being used to test for the treatment effect in RRMS trials, but the NB and P-IG distributions have been used for simulating the MRI data for the power analyses of these tests and determination of the associated sample sizes.

We consider three different trial designs in our study, namely parallel group (PG), baseline vs. treatment (BVT), and parallel group with a baseline correction (PGB). We identify the treatment effect by the parameter γ, with $1-\gamma$ representing the proportion reduction in the mean count of new lesions. For these designs we investigate the finite-sample properties of likelihood based parametric tests such as the likelihood ratio test (LRT) and Rao's score test (RST) for γ, and Wald tests (WT) for $g(\gamma)$ with $g(\gamma) = \gamma, \gamma^2, \sqrt{\gamma}$, and $\log(\gamma)$.

We use the NB and the P-IG models for PG trials and propose optimal likelihood based tests. Recently, tests based on the NB model have been proposed for PG trials; they rely on the χ^2 approximation and do not maintain Type I error rates for small samples. We propose simulation based tests that maintain Type I error rates, and for the NB model we also consider the case of unequal dispersion parameters for the two groups. For BVT and PGB trials, assuming a bivariate NB (BNB) model, we investigate various parametric tests and compare them. We perform power analyses and sample size estimation using the simulated percentiles of the exact distribution of the preferred test statistics for all the above scenarios.

We compare the sample sizes of our recommended parametric tests with those of the nonparametric tests published in the literature. For the NB models the exact LRT, RST, and WT for $\log(\gamma)$ remained unbiased and generally did equally well for all the three designs. When compared to the corresponding nonparametric test, the LRT gave 30-45% reduction in sample sizes for the PG trials, 25-60% for the BVT trials, and 70-80% for the PGB trials. The WT for γ^2, though not unbiased, had the highest power for $\gamma < 1$ and provided a further reduction of around 10-20% over the LRT in terms of sample sizes. Hence, it is best suited for RRMS clinical trials. For the P-IG model for PG trials, the LRT provided a sample size reduction of 30-50% compared to the Wilcoxon Rank Sum test and the exact WT for γ provided a reduction of 40-50%.

To my wife, Priya

To my parents, Bala and Gopal

ACKNOWLEDGMENTS

The years I spent at Ohio State have been by far the hardest, but also the most enjoyable years I have faced in my career so far. There are several people that made this journey possible and I would like to thank them all.

My advisor Prof. Nagaraja has been a source of guidance and inspiration throughout my years as a graduate student here at Ohio State. That my research and thesis is largely due to him goes without saying. My interactions with him each week have taught me to be not only a better statistician and a researcher, but also a better person. None of this would have been possible without his encouragement, support and advice every single step of the way in both academics and otherwise. For this and everything else I am sincerely grateful to him. I would also like to thank my committee members Prof. Jason Hsu, Prof. Eloise Kaizar, and Prof. Thomas Santner whose valuable insights have only enhanced my thesis and expanded my research. I thank Prof. Jackie Miller for guiding me through my very tough initial years as a graduate student and as a teaching assistant. I would like to thank the Department of Statistics for providing the facilities to work and for its continuous financial support all along. I would also like to thank Brian Smith and Eric Drake for providing computer support all these years.

I would like to thank all my professors at Loyola College, Chennai, India for encouraging me to pursue my higher studies. My conversation with Prof. Balakrishnan,

McMaster University, Canada, one summer morning, has been largely responsible for me coming to Ohio State and I thank him for that.

I thank my parents Bala and Gopal, my sister Aparna and my brother Ram for having stood by me all these years. Their belief in my abilities kept driving me to work harder towards my goal. Last but not least, I thank my wife Priya for her constant support and encouragement. Her love, affection, and sacrifice have helped me overcome some of the toughest days of my life and I will be forever indebted to her for that.

VITA

1999	B.Sc. Statistics, University of Madras, India
2001	M.Sc. Statistics, University of Madras, India
2005	M.Sc. Statistics, McMaster University, Canada
2005-2009	Graduate Teaching Associate, Department of Statistics, The Ohio State University
June-September 2009	Graduate Summer Intern, ACHRI, University of Arkansas for Medical Sciences

FIELDS OF STUDY

Major Field: Statistics

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

Introduction

1.1 Multiple Sclerosis

Multiple sceloris (MS) is an autoimmune demyelinating disease in which the immune system attacks the central nervous system and damages the myelin sheath covering the nerves resulting in a break down of the blood-brain barrier (BBB). This affects the ability of the brain cells and the spinal cord to communicate with each other. The term "multiple sclerosis" refers to scars or lesions that result from the destruction of the tissues that wrap around the nerves (myelin sheath). This destruction is called demyelination.[1] In the United States alone, there are approximately 400,000 individuals affected by MS. It is more common among women and starts usually among people aged between 20 and 40.

There are many possible symptoms of MS since the central nervous system controls much of the body's functioning. Since these are very general symptoms, it often takes many years for patients to be diagnosed with MS. Vision loss, muscle weakness, pain in the arms and legs, fatigue, loss of balance, problems with speech, paralysis etc., are some of, but not limited to, symptoms of MS. These symptoms vary with each patient

[1] *http://www.merck.com/mmhe/sec06/ch092/ch092b.html*

and it is also very difficult to predict the progression of the disease for any individual patient. During the initial stages of the disease the symptoms are usually experienced few and far between. The disease generally progresses from a relapsing-remitting phase to a more severe stage where symptoms are experienced fairly regularly. The different stages of MS are described in the next section below.

1.1.1 Different Stages of MS

There are four basic stages of MS. These are:

1. **Relapsing-Remitting Multiple Sclerosis (RRMS):** At this stage, the patient experiences periods called a *relapse* which are characterized by the acute attacks and worsening of symptoms. This is then followed by a *remission* where full or partial recovery of the body function is observed. The relapse can last from a few days to several weeks followed by prolonged periods of remission. Usually, at this stage, the symptoms do not seem to worsen between attacks.

2. **Secondary Progressive Multiple Sclerosis (SPMS):** This initially begins with a relapsing-remitting stage and later develops into a progressive stage. This may start immediately after the onset of the MS or may appear several years later.

3. **Primary Progressive Multiple Sclerosis (PPMS):** This stage is characterized by gradual progression of the disease without any obvious periods of remission although the disease tends to level out in between. This is the most common type of MS in people who develop the disease after the age of 40.

4. **Progressive Relapsing Multiple Sclerosis (PRMS):** This is the least common type of MS where patients experience very steady disease progression with severe attacks with little to no relief from the symptoms.

Our research work is primarily focussed on the study of clinical trials involing RRMS patients.

1.1.2 Monitoring Disease Progression in MS

Several measures of clinical evaluations are used to monitor disease progression in various types of MS. Degree of disability, relapse rate and time to clinical deterioration are used as the primary outcome measures in RRMS clinical trials.

Magnetic Resonance Imaging (MRI)

MRI is an imaging technique that has increasingly been used to diagnose and monitor the pathological progression of the disease MS. MRI is able to identify lesions in about 95% of the patients with clinically established MS. This tool can produce high quality images of the brain and makes it possible to visualize the lesions inside the brain. Serial MRI scans, possibly taken every month, can be used to study the evolution of the lesions and to determine the disease progress.

Gadolinium-enhancing lesions are accepted as markers of subclinical disease activity. However, the correlation between enhancing lesions and relapse rate is poor even when observed for long periods of time. When looking at individual cases, there may not be any relationship between the two for a long period of time. This happens when the disease activity occurs in clinically silent regions of the central nervous system (CNS). On the contrary, a single lesion appearing in the functionally critical

region of the CNS may show severe clinical consequences. Also, even though the short term relationship between enhancing lesions and clinical disability is poor or non-existent, a weak but significant relationship between the enhancing lesions and long-term disability is observed.

Other Measures

Disease activity in MS has been traditionally measured on clinical grounds such as the Expanded Disability Status Scale (EDSS) (Kurtzke [1983]). It is a method of quantifying disability in MS and is the most standard measure for monitoring changes in the level of disability over time. The EDSS scale ranges from 0 to 10 in increments of 0.5. Higher EDSS score corresponds to a greater level of disability of the patient. Although EDSS scale is used as a primary end point in Phase III clinical trials, it is not without its limitations. Firstly, it is not a continous scale and takes on only ordinal values (0 to 10, in steps of 0.5). Secondly, since it places greater importance on the ambulatory functions it is somewhat insensitive to neurological and cognitive dysfunctions. EDSS is a very subjective measure and calculating EDSS scores can be extremely complicated. Finally, due to the limitations in its design, it is very susceptible to "jumps" from one end of the scale to the other rather than seeing a smooth decline or improvement.

Another measure used in MS trials is the *relapse rate*, the rate at which relapses occur. This is also often used as a primary endpoint in Phase III MS trials. The quantification of relapses suffers due to the possible large variability among MS symptoms. Furthermore, a relapse can be a very rare occurrence since the manifestation of clinical evidence can be very modest in a trial lasting 2 to 3 years.

MRI detects 5-10 times more disease activity in RRMS and SPMS patients than is clinically apparent. MRI has two further advantages over the EDSS. Firstly, it is quantitative in the sense that the information gathered is the number of active lesions in contrast to the ordinal nature of the scales such as EDSS (measured from 0-10). These are objective while the clinical scales such as EDSS and relapse rate are subjective. Cumulative MRI lesion counts are used as a primary end point in Phase II clinical trials and as a secondary end point in Phase III clinical trials. In this thesis, we will be interested only in modelling the number of *new* enhancing lesions (lesions not seen during the previous month) seen each month in clinical trials involving RRMS patients.

1.2 Clinical Trial Designs in MS

Parallel Group (PG) trials, Baseline vs. Treatment (BVT) trials and Parallel Group trials with a baseline correction scan (PGB) are some of the most important and widely used experimental designs in clinical trials. In this section, we describe these designs in the context of RRMS clinical trials.

1.2.1 Parallel Group (PG) Trials

In PG trials, also called the randomized placebo-controlled trials, the patients are randomly split into two groups where one group receives the treatment while the other receives the placebo. This design has become the gold standard of clinical research. To establish the causality between the treatment and the outcome, steps are taken to ensure that the treatment and the control arms are similar in every way except the actual treatment itself. This ensures that no other variables are *confounded*

with the treatment. Double *blinding*, if feasible, ensures that neither the patients nor the investigators are aware of the patient assignment. The purpose of blinding is to minimize the patients receiving different care, or their data be interpreted differently, based on their assignment. Blinding is especially important when the end point of the study is subjective, but less important when it is objective as it is in the case of MS trials when the end point of the study is the number of enhanced MRI lesion counts.

1.2.2 Baseline vs. Treatment Trials (BVT)

In this design, all the patients would be subjected to an initial set of MRI scans, typically for 3 to 6 months, before the treatment is initiated. After that monthly MRI scans are taken over a pre-determined interval usually for up to 6 months. The first period is often called the baseline period and the second is called as the treatment period. The cumulative number of new enhancing lesions seen in both the baseline period and the treatment period are used as the primary outcome measures. This type of design is also known as the open-label cross-over design.

In this design, each patient serves as his or her own control. The within patient variation is reduced and this leads to smaller sample sizes to detect a significant treatment effect. A limitation of this design however is that even when the decrease in the number of lesions is not due to the treatment it may be perceived to be of treatment related.

1.2.3 Parallel Group Trials with a Baseline Correction Scan (PGB)

In this design, the baseline correction scan is obtained before the treatment is initiated. Once it is obtained, the rest of the trial design and protocol is exactly the same as for the PG trials. The two control groups have to be as similar as possible before the treatment is initiated. In some cases, even when randomization is done, the two groups are different with respect to the mean of the end point. The baseline correction scan is obtained to counter any significant differences among the two arms of the trial even before the treatment is initiated.

1.3 Statistical Models and Methods

1.3.1 Distributions

We give a brief introduction to the three distributions that will be used in later chapters.

Negative Binomial Distribution

The probability mass function of a random variable (rv) Y which is distributed according to a negative binomial (NB) distribution with mean μ and dispersion α is

$$P_Y(y) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\mu^y}{y!} \frac{\Gamma(y+\alpha)}{(\mu+\alpha)^{y+\alpha}}, \quad y = 0, 1, 2, \ldots; \quad \mu, \alpha > 0. \tag{1.3.1}$$

We say Y has NB(μ, α). Here $E(Y) = \mu$ and $Var(Y) = \mu + \mu^2/\alpha$. Note that the variance of Y depends on the mean μ and the dispersion parameter α. Also, the variance of Y exceeds its mean by a quantity μ^2/α. Further, the Poisson distribution is a special case of the NB distribution. When α goes to ∞ the $Var(Y)$ converges

to μ and Y converges in distribution to a Poisson rv with rate parameter μ. These properties make the NB distribution ideally suited to model over dispersed count data. The applications of this distribution to PG trials involving RRMS patients will be seen in Chapter 2.

Bivariate Negative Binomial Distribution

The joint pmf of two random variables X and Y that are distributed according to a bivariate negative binomial (BNB) distribution parameters μ_1, μ_2 and α is

$$P_{X,Y}(x,y) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\mu_1^x}{x!} \frac{\mu_2^y}{y!} \frac{\Gamma(x+y+\alpha)}{(\mu_1+\mu_2+\alpha)^{x+y+\alpha}}, \quad x,y = 0,1,2,\ldots; \quad \mu_1,\mu_2,\alpha > 0. \tag{1.3.2}$$

We then say $(X,Y) \sim \text{BNB}(\mu_1, \mu_2, \alpha)$. Marginally X and Y are each distributed according to $\text{NB}(\mu_1, \alpha)$ and $\text{NB}(\mu_2, \alpha)$ respectively. The applications of this distribution to RRMS BVT and PGB trials will be seen in Chapters 3 and 4 respectively.

The Poisson-Inverse Gaussian (P-IG) Distribution

The pmf of a rv Y distributed according to P-IG distribution with mean μ and shape parameter λ is

$$P(Y=y) = \frac{\tau^y}{y!} \left(\frac{2\theta}{\pi}\right)^{\frac{1}{2}} \exp\left(\frac{\lambda}{\mu}\right) K_{y-\frac{1}{2}}(\theta), \qquad y = 0,1,\ldots, \tag{1.3.3}$$

Here $\tau = \left[\frac{1}{\mu^2} + \frac{2}{\lambda}\right]^{-\frac{1}{2}}$, $\theta = \frac{\lambda}{\tau}$ and $K(\cdot)$ is the modified Bessel function of the third kind. It is defined as

$$K_\nu(z) = \frac{\pi}{2} \cdot \frac{I_{-\nu}(z) - I_\nu(z)}{\sin \nu\pi},$$

where $I(\cdot)$ is the modified Bessel function of the first kind defined by

$$I_\nu(z) = \sum_{m=0}^{\infty} \frac{(\frac{z}{2})^{\nu+2m}}{m!\Gamma(m+\nu+1)}.$$

Here, $E(Y) = \mu$ and $Var(Y) = \mu + \mu^3/\lambda$. This distribution is sometimes used as an alternative to the NB distribution to model over dispersed count data. As λ goes to ∞, Y converges in distribution to a Poisson rv with mean μ, as in the case of the NB distribution.

1.3.2 Parametric Tests

Let $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ be an independently and identically distributed (iid) sample from a distribution with probability function $P_Y(\boldsymbol{y}|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a $p \times 1$ parametric vector. Suppose $\boldsymbol{\theta}$, is partitioned into θ_1, a scalar parameter of interest and θ_2, a $(p-1) \times 1$ vector nuisance parameters. Also let the score vector, the vector of first order derivatives, $s(\boldsymbol{\theta}) = [s_1(\boldsymbol{\theta}), s_2(\boldsymbol{\theta})]'$ and the Fisher information matrix (FIM)

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}$$

be partitioned accordingly. Let the log-likelihood be $\ell(\boldsymbol{\theta}|\boldsymbol{y})$. Suppose $H_0 : \theta_1 = \theta_1^0$ be the hypothesis of interest. Let $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2)$ and $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1, \tilde{\theta}_2)$ be the MLE of $\boldsymbol{\theta}$ obtained under the unrestricted and restricted hypothesis respectively. We present below three popular parametric tests.

The Likelihood Ratio Test (LRT)

The LRT statistic (Rao [2005]) for testing $H_0 : \theta_1 = \theta_1^0$ vs. $H_1 : \theta_1 \neq \theta_1^0$ is

$$LR = -2(\ell(\tilde{\boldsymbol{\theta}}|\boldsymbol{y}) - \ell(\hat{\boldsymbol{\theta}}|\boldsymbol{y})) \tag{1.3.4}$$

where $\ell(\tilde{\boldsymbol{\theta}}|\boldsymbol{y})$ and $\ell(\hat{\boldsymbol{\theta}}|\boldsymbol{y})$ are the log-likelihood functions evaluated at the restricted and the unrestricted MLEs respectively.

The Rao Score Test (RST)

The RST statistic (Rao [1948, 2005]) for testing $H_0 : \theta_1 = \theta_1^0$ vs. $H_1 : \theta_1 \neq \theta_1^0$ is

$$RS = \left[s_1(\tilde{\boldsymbol{\theta}})\right]^2 [I_{1.2}(\tilde{\boldsymbol{\theta}})]^{-1}, \tag{1.3.5}$$

where $I_{1.2} = I_{11} - I_{12}I_{22}^{-1}I_{21}$ is a scalar and equals $\sigma^2_{\hat{\theta}_1}(\tilde{\boldsymbol{\theta}})$, the asymptotic variance of $\hat{\theta}_1$ evaluated at the MLEs obtained under H_0.

The Wald Test (WT)

The WT statistic (Wald [1943], Rao [2005]) for testing $H_0 : \theta_1 = \theta_1^0$ vs. $H_1 : \theta_1 \neq \theta_1^0$ is

$$WT = \left[\frac{\hat{\theta}_1 - \theta_1^0}{\sigma_{\hat{\theta}_1}(\hat{\boldsymbol{\theta}})}\right]^2 \tag{1.3.6}$$

where $\sigma^2_{\hat{\theta}_1}(\hat{\boldsymbol{\theta}})$ is the asymptotic variance of $\hat{\theta}_1$ evaluated at $\hat{\boldsymbol{\theta}}$.

Each of the above three tests are asymptotically equivalent and have a χ^2_1 distribution under H_0. The LRT requires computation of the MLEs of $\boldsymbol{\theta}$ under both the restricted and the unrestricted cases, while the RST requires the MLEs under only the restricted case and the WT requires the MLEs under only the unrestricted case.

1.4 The Research Problem

This work is motivated by the problem of sample size and power calculations for clinical trials in MS. These trials routinely use MRI lesion counts as their outcome measure of interest and are very expensive since they involve taking repeated MRI scans of the brain. Typically the patients are followed up for several months and the MRI scans of the brain are taken each month. Currently only nonparametric tests such as the Wilcoxon Rank Sum (WRS) and the Wilcoxon Signed Rank (WSR) test

are being used in MS trials even though it has been shown that the NB and P-IG distributions provide a good fit to such count data. Using parametric tests for the assumed parametric models generally increase the power to detect a treatment effect and thus reduce the sample sizes required. The goals of this thesis are (i) to develop desirable parametric tests for the treatment effect based on the NB and the P-IG model for different trial designs in RRMS clinical trials, and (ii) to examine the magnitude of the increase in power and reduction in sample size over the corresponding nonparametric tests. In Section 1.1 we have given a brief introduction to the disease of multiple sclerosis and discuss its different stages and progression. Three of the most common clinical trial designs used in multiple sclerosis clinical trials, the PG, BVT, and PGB designs were described in Section 1.2. Section 1.3 has introduced the NB, BNB and P-IG models to model MRI lesion count data and the LRT, RST and WTs that will be used in future chapters to test for the treatment efficacy.

In Chapter 2 we apply the NB model to MRI lesion count data in RRMS PG trials. In Chapter 3 we propose a bivariate NB (BNB) distribution to model lesion count data arising out of RRMS BVT trials. The BNB distribution is used to model data from RRMS PGB trials in Chapter 4. In Chapter 5 the P-IG distribution is used as an alternative to model lesion count data from PG trials. For all these models we obtain maximum likelihood estimates of the parameters and propose LRT, RST and WTs to test for the treatment effect. Through a detailed simulation study we evaluate and compare the performance of these tests and suggest appropriate tests for each trial design and parametric model considered. For each of the above scenarios we obtain sample size estimates based on selected tests and compare them to those obtained using nonparametric tests. Finally, in Chapter 6, we conclude by summarizing our

results and discuss avenues for further work. All the computational work required for this thesis was done using the software R Development Core Team [2010].

The Appendix contains basic notations and abbreviations used in this thesis. There we present data sets from MRI studies and an epilepsy study where the NB models provide a good fit. It also contains a brief description of the simulation method for different trial designs and sample R codes.

CHAPTER 2

Univariate Negative Binomial Models for Parallel Group Trials

2.1 Introduction

Use of the Poisson distribution to model the number of new enhancing lesions seen in RRMS patients requires the highly unrealistic assumption that the mean number of new enhancing lesions in any particular time period is the same for all the patients within a group. When subjects are randomly chosen and the subject effect is taken to be random where the mean parameter is assumed to be distributed as a Gamma variable, the distribution of the observed count data turns out to be negative binomial. When $X|\theta$ is assumed to be distributed according to Poisson(θ) and θ is further assumed to be distributed according to Gamma($\alpha, \mu/\alpha$), the marginal distribution of X is

$$\begin{aligned} P_X(x) &= \int_0^\infty P_{X|\theta}(x) f(\theta) d\theta \\ &= \int_0^\infty \frac{e^{-\theta}\theta^x}{x!} \frac{1}{\Gamma(\alpha)(\mu/\alpha)^\alpha} \theta^{\alpha-1} e^{-\theta\alpha/\mu} \\ &= \frac{\Gamma(x+\alpha)}{\Gamma(\alpha)x!} \left(\frac{\mu}{\mu+\alpha}\right)^x \left(\frac{\alpha}{\mu+\alpha}\right)^\alpha, \quad x = 0, 1, 2, \ldots; \alpha, \beta > 0. \end{aligned} \tag{2.1.1}$$

The expected value of X is μ and its variance is $\mu + \mu^2/\alpha$ where α is called the dispersion parameter. As α goes to infinity, X converges in distribution to a Poisson random variable with rate parameter μ but it has larger variance than the Poisson for all real α. Thus the NB distribution is used as an alternative to Poisson distribution when modeling over dispersed count data. This relationship between the NB and Poisson distributions was first derived by Greenwood and Yule [1920] who used it to represent "accident proneness".

There are several other chance mechanisms that give rise to the NB distribution. In a sequence of independent Bernoulli trials with probability of success p, if X denotes the number of failures that precede the r^{th} success, then X has a NB distribution with the following pmf:

$$P_X(x) = \binom{x+r-1}{r-1} p^r (1-p)^x, \quad x = 0, 1, \ldots; 0 < p < 1. \tag{2.1.2}$$

If in the above experiment, Y denotes the number of "trials" required to attain r successes, then $Y = X + r$ is said to have a NB distribution with pmf

$$P_Y(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}, \quad y = r, r+1, \ldots; 0 < p < 1. \tag{2.1.3}$$

The NB distribution in (2.1.2) and (2.1.3) can also be derived as a sum of r i.i.d. geometric random variables which are appropriately defined. Johnson et al. [2005] provide a detailed description of this distribution and its properties and discuss several other chance mechanisms which lead to it.

In this chapter we focus on the NB distribution that arises as a gamma mixture of Poisson random variables as given in (2.1.1). In Section 2.2 this distribution is used to model MRI lesion counts data arising out of PG trials involving RRMS patients. The maximum likelihood estimation of the model parameters is done in Section 2.3. Exact

parametric tests for the treatment effect are developed in Section 2.4. Section 2.5 compares the performance of these tests and Section 2.6 provides sample size estimates applicable to RRMS PG trials. The findings are summarized in Section 2.8.

2.2 The NB Model

Sormani et al. [1999] show that the NB distribution fits better than the Poisson distribution to the data on the number of new enhancing lesions counted in 56 MS patients followed for 9 months. Sormani et al. [2001a] show that the NB distribution fits well to other larger data sets (see Appendix Section C.1.3 for the data) of MRI counts from RRMS patients. We use the empirical evidence provided by these data sets to build univariate NB models for PG trials on RRMS patients. In PG trials (also called randomized placebo-controlled trials), the patients are randomly assigned to one of two groups where one group receives the treatment while the other receives the placebo. Let Y_1 (Y_2) be the total number of *new* enhancing lesions seen during the follow-up period of an RRMS patient in the placebo (treatment) group. Let Z_1, Z_2 denote the random subject effects, assumed to be independent Gamma(α, α^{-1}) rvs. We assume that

$$Y_1|Z_1 = z \sim \text{Poisson}(\mu z) \text{ and } Y_2|Z_2 = z \sim \text{Poisson}(\gamma\mu z), \quad \mu, \gamma > 0.$$

Then Y_1 is NB(μ, α), and Y_2 is NB$(\gamma\mu, \alpha)$, and since the patients are nested within groups, Y_1 and Y_2 are assumed to be independent.

If there is evidence that the dispersion parameter α is also affected by the treatment, we assume

$$Y_1|Z_1 = z_1 \sim \text{Poisson}(\mu z_1) \text{ and } Y_2|Z_2 = z_2 \sim \text{Poisson}(\gamma\mu z_2),$$

where Z_i is distributed according to Gamma$(\alpha_i, {\alpha_i}^{-1})$, $i = 1, 2$. Then Y_1 and Y_2 are independent, Y_1 is NB(μ, α_1), and Y_2 is NB$(\gamma\mu, \alpha_2)$.

We assume there are n_1 subjects in the placebo group and n_2 in the treatment group.

2.3 Estimation

2.3.1 Equal Dispersion Parameters

We now find the MLEs of the parameters of the NB model.[2] Since the subjects in each group and between groups are independent of each other, the joint likelihood for this model with data $\mathbf{y}_1 = (y_{11}, y_{12}, \ldots, y_{1n_1})$ and $\mathbf{y}_2 = (y_{21}, y_{22}, \ldots, y_{2n_2})$ is given by

$$\begin{aligned} L(\gamma, \mu, \alpha | \mathbf{y_1}, \mathbf{y_2}) = & \prod_{i=1}^{n_1} P_{Y_1}(y_{1i}) \prod_{j=1}^{n_2} P_{Y_2}(y_{2j}) \\ = & \left[\frac{\alpha^\alpha}{\Gamma(\alpha)}\right]^{n_1+n_2} \frac{\mu^{\sum y_{1i}}}{(\mu+\alpha)^{\sum y_{1i}+n_1\alpha}} \frac{(\gamma\mu)^{\sum y_{2j}}}{(\gamma\mu+\alpha)^{\sum y_{2j}+n_2\alpha}} \\ & \times \prod_{i=1}^{n_1} \frac{\Gamma(y_{1i}+\alpha)}{y_{1i}!} \prod_{j=1}^{n_2} \frac{\Gamma(y_{2j}+\alpha)}{y_{2j}!}, \end{aligned} \tag{2.3.1}$$

and the parameter space is $\Theta = \{(\gamma, \mu, \alpha) : \gamma, \mu, \alpha > 0\}$. The log-likelihood is

$$\begin{aligned} \ell(\gamma, \mu, \alpha) = & (n_1 + n_2) \{\alpha \log \alpha - \log \Gamma(\alpha)\} + (n_1 \bar{y}_1 + n_2 \bar{y}_2) \log(\mu) \\ & - n_1(\bar{y}_1 + \alpha) \log(\mu + \alpha) + n_2 \bar{y}_2 \log(\gamma) - n_2(\bar{y}_2 + \alpha) \log(\gamma\mu + \alpha) \\ & + \sum_{i=1}^{n_1} \{\log \Gamma(y_{1i} + \alpha) - \log(y_{1i}!)\} + \sum_{j=1}^{n_2} \{\log \Gamma(y_{2j} + \alpha) - \log(y_{2j}!)\}. \end{aligned} \tag{2.3.2}$$

[2]These are available in Aban et al. [2009]. We present the likelihood function and the parameter estimates here for completeness.

The first order derivatives of the log-likelihood with respect to the three parameters are:

$$\frac{\partial \ell(\gamma,\mu,\alpha)}{\partial \gamma} = \frac{n_2\alpha(\bar{y}_2-\gamma\mu)}{\gamma(\gamma\mu+\alpha)}, \tag{2.3.3}$$

$$\frac{\partial \ell(\gamma,\mu,\alpha)}{\partial \mu} = \frac{n_1\bar{y}_1+n_2\bar{y}_2}{\mu} - \frac{n_1(\bar{y}_1+\alpha)}{\mu+\alpha} - \frac{n_2(\bar{y}_2+\alpha)}{\gamma\mu+\alpha}\gamma, \tag{2.3.4}$$

$$\begin{aligned}\frac{\partial \ell(\gamma,\mu,\alpha)}{\partial \alpha} =& -(n_1+n_2)[\psi(\alpha)-1] + n_1\log\left(\frac{\alpha}{\mu+\alpha}\right) - \frac{n_1(\bar{y}_1+\alpha)}{\mu+\alpha} + \sum_{i=1}^{n_1}\psi(y_{1i}+\alpha) \\ &+ n_2\log\left(\frac{\alpha}{\gamma\mu+\alpha}\right) - \frac{n_2(\bar{y}_2+\alpha)}{\gamma\mu+\alpha} + \sum_{j=1}^{n_2}\psi(y_{2j}+\alpha).\end{aligned} \tag{2.3.5}$$

The MLEs of the parameters are obtained by setting the above first order score vector equations to 0 and solving for the three parameters. This gives us the following result:

Lemma 2.3.1. *The MLEs of γ and μ are*

$$\hat{\gamma} = \bar{y}_2/\bar{y}_1 \text{ and } \hat{\mu} = \bar{y}_1. \tag{2.3.6}$$

The MLE of α, $\hat{\alpha}$, is obtained by numerically maximizing the profile log-likelihood $\ell(\hat{\gamma},\hat{\mu},\alpha)$ with respect to α.

When $\gamma = \gamma_0$ (known), the MLEs of μ and α are obtained by setting equations (2.3.4) and (2.3.5), with $\gamma = \gamma_0$, to 0 and solving for the two parameters. This leads us to the following result:

Lemma 2.3.2. *(Aban et al. [2009]) When $\gamma = \gamma_0$ is known, the MLEs of μ and α satisfy the following constraints:*

$$\begin{aligned}\tilde{\mu} =& \frac{n_1(\gamma_0\bar{y}_1-\tilde{\alpha}) + n_2(\bar{y}_2-\gamma_0\tilde{\alpha})}{2\gamma_0(n_1+n_2)} \\ &+ \frac{\sqrt{[n_1(\gamma_0\bar{y}_1-\tilde{\alpha}) + n_2(\bar{y}_2-\gamma_0\tilde{\alpha})]^2 + 4\gamma_0\tilde{\alpha}(n_1+n_2)(n_1\bar{y}_1+n_2\bar{y}_2)}}{2\gamma_0(n_1+n_2)};\end{aligned}$$

$$\begin{aligned}0 =& -(n_1+n_2)[\psi(\tilde{\alpha})-1] + n_1\log\left(\frac{\tilde{\alpha}}{\tilde{\mu}+\tilde{\alpha}}\right) + n_2\log\left(\frac{\tilde{\alpha}}{\gamma_0\tilde{\mu}+\tilde{\alpha}}\right) \\ &- \frac{n_1(\bar{y}_1+\tilde{\alpha})}{\tilde{\mu}+\tilde{\alpha}} - \frac{n_2(\bar{y}_2+\tilde{\alpha})}{\gamma_0\tilde{\mu}+\tilde{\alpha}} + \sum\psi(y_{1i}+\tilde{\alpha}) + \sum\psi(y_{2j}+\tilde{\alpha}).\end{aligned}$$

When $\gamma_0 = 1$, $\tilde{\mu}$ above simplifies to

$$\tilde{\mu} = \frac{n_1\bar{y}_1 + n_2\bar{y}_2}{n_1 + n_2},$$

and is free of $\tilde{\alpha}$ and in that case $\tilde{\alpha}$ is obtained by numerically maximizing the profile log-likelihood $\ell(\gamma_0, \tilde{\mu}, \alpha)$ with respect to α.

The second order derivatives of (2.3.2) are:

$$\begin{aligned}
\frac{\partial^2\ell(\gamma,\mu,\alpha)}{\partial\gamma^2} =& \frac{-n_2\alpha\left[\gamma\mu(\gamma\mu+\alpha)+(\bar{y}_2-\gamma\mu)(2\gamma\mu+\alpha)\right]}{\left[\gamma(\gamma\mu+\alpha)\right]^2},\\
\frac{\partial^2\ell(\gamma,\mu,\alpha)}{\partial\gamma\partial\mu} =& -\frac{n_2\alpha(\bar{y}_2+\alpha)}{(\gamma\mu+\alpha)^2} = \frac{\partial^2\ell(\gamma,\mu,\alpha)}{\partial\mu\partial\gamma},\\
\frac{\partial^2\ell(\gamma,\mu,\alpha)}{\partial\gamma\partial\alpha} =& -\frac{n_2\mu(\gamma\mu-\bar{y}_2)}{(\gamma\mu+\alpha)^2} = \frac{\partial^2\ell(\gamma,\mu,\alpha)}{\partial\alpha\partial\gamma},\\
\frac{\partial^2\ell(\gamma,\mu,\alpha)}{\partial\mu^2} =& -\frac{n_1\bar{y}_1+n_2\bar{y}_2}{\mu^2} + \frac{n_1(\bar{y}_1+\alpha)}{(\mu+\alpha)^2} + \frac{n_2\gamma^2(\bar{y}_2+\alpha)}{(\gamma\mu+\alpha)^2},\\
\frac{\partial^2\ell(\gamma,\mu,\alpha)}{\partial\mu\partial\alpha} =& -\frac{n_1(\mu-\bar{y}_1)}{(\mu+\alpha)^2} - \frac{n_2\gamma(\gamma\mu-\bar{y}_2)}{(\gamma\mu+\alpha)^2} = \frac{\partial^2\ell(\gamma,\mu,\alpha)}{\partial\alpha\partial\mu},\\
\frac{\partial^2\ell(\gamma,\mu,\alpha)}{\partial\alpha^2} =& -(n_1+n_2)\psi'(\alpha) + \frac{n_1\mu}{\alpha(\mu+\alpha)} + \frac{n_2\gamma\mu}{\alpha(\gamma\mu+\alpha)} - \frac{n_1(\mu-\bar{y}_1)}{(\mu+\alpha)^2}\\
& - \frac{n_2(\gamma\mu-\bar{y}_2)}{(\gamma\mu+\alpha)^2} + \sum_{i=1}^{n_1}\psi'(y_{1i}+\alpha) + \sum_{j=1}^{n_2}\psi'(y_{2j}+\alpha).
\end{aligned} \tag{2.3.7}$$

The Fisher Information Matrix (FIM) is $\mathbf{I}(\boldsymbol{\theta}) = \big(I(\boldsymbol{\theta})\big)_{i,j} = -E\left\{\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right\}$ where $\theta = (\theta_1,\theta_2,\theta_3) = (\gamma,\mu,\alpha)$ is the parameter vector.

Lemma 2.3.3. *Using the fact that $E(\bar{Y}_1) = \mu$ and $E(\bar{Y}_2) = \gamma\mu$, and from the fact that $Y_{1i}, i = 1,\ldots,n_1$ and $Y_{2j}, j = 1,\ldots,n_2$, are respectively identically distributed,*

we get the elements of the FIM to be

$$
\begin{aligned}
I_{11}(\boldsymbol{\theta}) &= -E\left\{\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial\gamma^2}\right\} = \frac{n_2\mu\alpha}{\gamma(\gamma\mu+\alpha)},\\
I_{22}(\boldsymbol{\theta}) &= -E\left\{\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial\mu^2}\right\} = \frac{n_1\alpha}{\mu(\mu+\alpha)} + \frac{n_2\gamma\alpha}{\mu(\gamma\mu+\alpha)},\\
I_{33}(\boldsymbol{\theta}) &= -E\left\{\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial\alpha^2}\right\} = n_1\left\{\psi'(\alpha) - \frac{\mu}{\alpha(\mu+\alpha)}\right\} - n_1E(\psi'(Y_1+\alpha))\\
&\quad + n_2\left\{\psi'(\alpha) - \frac{\gamma\mu}{\alpha(\gamma\mu+\alpha)}\right\} - n_2E(\psi'(Y_2+\alpha)),\\
I_{12}(\boldsymbol{\theta}) &= -E\left\{\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial\gamma\partial\mu}\right\} = \frac{n_2\alpha}{\gamma\mu+\alpha} = I_{21}(\boldsymbol{\theta}), \quad \textit{and}\\
I_{ij}(\boldsymbol{\theta}) &= 0, \quad \textit{otherwise.}
\end{aligned}
$$

2.3.2 Unequal Dispersion Parameters

When the dispersion parameters differ,

$$
\begin{aligned}
L(\gamma,\mu,\alpha_1,\alpha_2|\mathbf{y_1},\mathbf{y_2}) = &\left[\frac{\alpha_1^{\alpha_1}}{\Gamma(\alpha_1)}\right]^{n_1}\frac{\mu^{\sum y_{1i}}}{(\mu+\alpha_1)^{\sum y_{1i}+n_1\alpha_1}}\left[\frac{\alpha_2^{\alpha_2}}{\Gamma(\alpha_2)}\right]^{n_2}\frac{(\gamma\mu)^{\sum y_{2j}}}{(\gamma\mu+\alpha_2)^{\sum y_{2j}+n_2\alpha_2}}\\
&\times\prod_{i=1}^{n_1}\frac{\Gamma(y_{1i}+\alpha_1)}{y_{1i}!}\prod_{j=1}^{n_2}\frac{\Gamma(y_{2j}+\alpha_2)}{y_{2j}!},
\end{aligned}
$$

and the parameter space is $\Theta = \{(\gamma,\mu,\alpha_1,\alpha_2) : \gamma,\mu,\alpha_1,\alpha_2 > 0\}$. The log-likelihood is

$$
\begin{aligned}
\ell(\gamma,\mu,\alpha_1,\alpha_2) = &\, n_1\left\{\alpha_1\log\alpha_1 - \log\Gamma(\alpha_1)\right\} + n_2\left\{\alpha_2\log\alpha_2 - \log\Gamma(\alpha_2)\right\}\\
&+ (n_1\bar{y}_1 + n_2\bar{y}_2)\log(\mu) - n_1(\bar{y}_1+\alpha_1)\log(\mu+\alpha_1) + n_2\bar{y}_2\log(\gamma)\\
&- n_2(\bar{y}_2+\alpha_2)\log(\gamma\mu+\alpha_2) + \sum_{i=1}^{n_1}\left\{\log\Gamma(y_{1i}+\alpha_1) - \log(y_{1i}!)\right\}\\
&+ \sum_{j=1}^{n_2}\left\{\log\Gamma(y_{2j}+\alpha_2) - \log(y_{2j}!)\right\}. \qquad (2.3.8)
\end{aligned}
$$

The score vector components are

$$\frac{\partial \ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial \gamma} = \frac{n_2\alpha_2(\bar{y}_2-\gamma\mu)}{\gamma(\gamma\mu+\alpha_2)}, \tag{2.3.9}$$

$$\frac{\partial \ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial \mu} = \frac{n_1\bar{y}_1}{\mu}-\frac{n_1(\bar{y}_1+\alpha_1)}{\mu+\alpha_1}+\frac{n_2\bar{y}_2}{\mu}-\frac{n_2(\bar{y}_2+\alpha_2)}{\gamma\mu+\alpha_2}\gamma, \tag{2.3.10}$$

$$\begin{aligned}\frac{\partial \ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial \alpha_1} = & -n_1[\psi(\alpha_1)-1]+n_1\log\left(\frac{\alpha_1}{\mu+\alpha_1}\right)-\frac{n_1(\bar{y}_1+\alpha_1)}{\mu+\tilde{\alpha}_1} \\ & +\sum_{i=1}^{n_1}\psi(y_{1i}+\alpha_1),\end{aligned} \tag{2.3.11}$$

$$\begin{aligned}\frac{\partial \ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial \alpha_2} = & -n_2[\psi(\alpha_2)-1]+n_2\log\left(\frac{\alpha_2}{\gamma\mu+\alpha_2}\right)-\frac{n_2(\bar{y}_2+\alpha_2)}{\gamma\mu+\alpha_2} \\ & +\sum_{j=1}^{n_2}\psi(y_{2j}+\alpha_2).\end{aligned} \tag{2.3.12}$$

where $\psi(\cdot)$ is the digamma function.

Lemma 2.3.4. *The MLEs of γ and μ are given by (2.3.6) as in the equal dispersion parameter case, and the MLEs of α_1 and α_2 are obtained by numerically maximizing the profile loglikelihoood $\ell(\hat{\gamma},\hat{\mu},\alpha_1,\alpha_2)$ with respect to (α_1,α_2).*

Lemma 2.3.5. *When $\gamma=\gamma_0$ and known, the MLEs of μ, α_1 and α_2 satisfy the following:*

$$\begin{aligned}\tilde{\mu} =& \frac{n_1\tilde{\alpha}_1(\gamma_0\bar{y}_1-\tilde{\alpha}_2)+n_2\tilde{\alpha}_2(\bar{y}_2-\gamma_0\alpha_1)}{2\gamma_0(n_1\tilde{\alpha}_1+n_2\tilde{\alpha}_2)} \\ & +\frac{\sqrt{[n_1\tilde{\alpha}_1(\gamma_0\bar{y}_1-\tilde{\alpha}_2)+n_2\tilde{\alpha}_2(\bar{y}_2-\gamma_0\tilde{\alpha}_1)]^2+4\gamma_0\tilde{\alpha}_1\tilde{\alpha}_2(n_1\tilde{\alpha}_1+n_2\tilde{\alpha}_2)(n_1\bar{y}_1+n_2\bar{y}_2)}}{2\gamma_0(n_1\tilde{\alpha}_1+n_2\tilde{\alpha}_2)},\end{aligned}$$

$$0 = -n_1[\psi(\tilde{\alpha}_1)-1]+n_1\log\left(\frac{\tilde{\alpha}_1}{\tilde{\mu}+\tilde{\alpha}_1}\right)-\frac{n_1(\bar{y}_1+\tilde{\alpha}_1)}{\tilde{\mu}+\tilde{\alpha}_1}+\sum_{i=1}^{n_1}\psi(y_{1i}+\tilde{\alpha}_1),$$

$$0 = -n_2[\psi(\tilde{\alpha}_2)-1]+n_2\log\left(\frac{\tilde{\alpha}_2}{\gamma_0\tilde{\mu}+\tilde{\alpha}_2}\right)-\frac{n_2(\bar{y}_2+\tilde{\alpha}_2)}{\gamma_0\tilde{\mu}+\tilde{\alpha}_2}+\sum_{j=1}^{n_2}\psi(y_{2j}+\tilde{\alpha}_2).$$

The second order derivatives of (2.3.8) are

$$\begin{aligned}
\frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\gamma^2} =& \frac{-n_2\alpha_2\left[\gamma\mu(\gamma\mu+\alpha_2)+(\bar{y}_2-\gamma\mu)(2\gamma\mu+\alpha_2)\right]}{\left[\gamma(\gamma\mu+\alpha_2)\right]^2},\\
\frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\gamma\partial\mu} =& -\frac{n_2\alpha_2(\bar{y}_2+\alpha_2)}{(\gamma\mu+\alpha_2)^2} = \frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\mu\partial\gamma},\\
\frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\gamma\partial\alpha_1} =& 0 = \frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\alpha_1\partial\gamma},\\
\frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\gamma\partial\alpha_2} =& -\frac{n_2\mu(\gamma\mu-\bar{y}_2)}{(\gamma\mu+\alpha_2)^2} = \frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\alpha_2\partial\gamma},\\
\frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\mu^2} =& -\frac{n_1\bar{y}_1+n_2\bar{y}_2}{\mu^2}+\frac{n_1(\bar{y}_1+\alpha_1)}{(\mu+\alpha_1)^2}+\frac{n_2\gamma^2(\bar{y}_2+\alpha_2)}{(\gamma\mu+\alpha_2)^2},\\
\frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\mu\partial\alpha_1} =& -\frac{n_1(\mu-\bar{y}_1)}{(\mu+\alpha)^2} = \frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\alpha_1\partial\mu},\\
\frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\mu\partial\alpha_2} =& -\frac{n_2\gamma(\gamma\mu-\bar{y}_2)}{(\gamma\mu+\alpha)^2} = \frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\alpha_2\partial\mu},\\
\frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\alpha_1^2} =& -n_1\psi'(\alpha_1)+\frac{n_1\mu}{\alpha_1(\mu+\alpha_1)}-\frac{n_1(\mu-\bar{y}_1)}{(\mu+\alpha_1)^2}+\sum_{i=1}^{n_1}\psi'(y_{1i}+\alpha_1),\\
\frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\alpha_2^2} =& -n_2\psi'(\alpha_2)+\frac{n_2\gamma\mu}{\alpha_2(\gamma\mu+\alpha_2)}-\frac{n_2(\gamma\mu-\bar{y}_2)}{(\gamma\mu+\alpha_2)^2}+\sum_{j=1}^{n_2}\psi'(y_{2j}+\alpha_2),\\
\frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\alpha_1\partial\alpha_2} =& 0 = \frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\alpha_2\partial\alpha_1}.
\end{aligned} \tag{2.3.13}$$

Lemma 2.3.6. *Using the fact that $E(\bar{Y}_1)=\mu$ and $E(\bar{Y}_2)=\gamma\mu$, and from the fact that $Y_{1i}, i=1,\ldots,n_1$ and $Y_{2j}, j=1,\ldots,n_2$ are respectively identically distributed, we get the elements of the FIM $\mathbf{I}(\boldsymbol{\theta})$ to be*

$$\begin{aligned}
I_{11}(\boldsymbol{\theta}) =& -E\left\{\frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\gamma^2}\right\} = \frac{n_2\mu\alpha_2}{\gamma(\gamma\mu+\alpha_2)},\\
I_{22}(\boldsymbol{\theta}) =& -E\left\{\frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\mu^2}\right\} = \frac{n_1\alpha_1}{\mu(\mu+\alpha_1)}+\frac{n_2\gamma\alpha_2}{\mu(\gamma\mu+\alpha_2)},\\
I_{33}(\boldsymbol{\theta}) =& -E\left\{\frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\alpha_1^2}\right\} = n_1\left\{\psi'(\alpha_1)-\frac{\mu}{\alpha_1(\mu+\alpha_1)}\right\}-n_1E(\psi'(Y_1+\alpha_1)),\\
I_{44}(\boldsymbol{\theta}) =& -E\left\{\frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\alpha_2^2}\right\} = n_2\left\{\psi'(\alpha_2)-\frac{\gamma\mu}{\alpha_2(\gamma\mu+\alpha_2)}\right\}-n_2E(\psi'(Y_2+\alpha_2)),\\
I_{12}(\boldsymbol{\theta}) =& -E\left\{\frac{\partial^2\ell(\gamma,\mu,\alpha_1,\alpha_2)}{\partial\gamma\partial\mu}\right\} = \frac{n_2\alpha_2}{\gamma\mu+\alpha_2} = I_{21}(\boldsymbol{\theta}), \quad \textit{and}\\
I_{ij}(\boldsymbol{\theta}) =& \ 0, \quad \textit{otherwise},
\end{aligned}$$

where $\psi'(\cdot)$ is the trigamma function and $\boldsymbol{\theta} = (\gamma, \mu, \alpha_1, \alpha_2)$ is the parameter vector.

Lemma 2.3.7. *The asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ is $\Sigma(\boldsymbol{\theta}) = [\mathbf{I}(\boldsymbol{\theta})]^{-1}$. Further, $\sigma^2_{\hat{\gamma}} = \Sigma_{11}(\boldsymbol{\theta})$, the asymptotic variance of $\hat{\gamma}$, is given by*

$$\sigma^2_{\hat{\gamma}}(\boldsymbol{\theta}) = \frac{\gamma\left[n_1\alpha_1(\gamma\mu + \alpha_2) + n_2\alpha_2\gamma(\mu + \alpha_1)\right]}{n_1 n_2 \alpha_1 \alpha_2 \mu}. \tag{2.3.14}$$

Remarks: The MLEs are determined using a combination of solution of likelihood equations for some of the parameters and direct numerical maximization of the likelihood with respect to the others. For PG trials with common α, the solutions of μ and γ (or just of μ when $\gamma_0 = 1$) are obtained from likelihood equations while keeping the α fixed. It can be shown that these solutions correspond to the maximum value of the likelihood for a fixed α. When the model is parametrized in terms of (μ_1, μ_2, α), the solutions for μ_1 and μ_2 obtained from the score equations are $\bar{y}_1$ and $\bar{y}_2$ respectively. For a fixed α, the determinant of the second order derivative matrix of the log-likelihood evaluated at $\hat{\mu}_1 = \bar{y}_1$ and $\hat{\mu}_2 = \bar{y}_2$ is

$$\begin{vmatrix} \frac{\partial^2 \ell(\mu_1,\mu_2,\alpha)}{\partial \mu_1^2} & \frac{\partial^2 \ell(\mu_1,\mu_2,\alpha)}{\partial \mu_1 \partial \mu_2} \\ \frac{\partial^2 \ell(\mu_1,\mu_2,\alpha)}{\partial \mu_1 \partial \mu_2} & \frac{\partial^2 \ell(\mu_1,\mu_2,\alpha)}{\partial \mu_2^2} \end{vmatrix} = \begin{vmatrix} \frac{-n_1\alpha}{\bar{y}_2(\bar{y}_2+\alpha)} & 0 \\ 0 & \frac{-n_2\alpha}{\bar{y}_2(\bar{y}_2+\alpha)} \end{vmatrix} = \frac{n_1 n_2 \alpha^2}{\bar{y}_1\bar{y}_2(\bar{y}_1+\alpha)(\bar{y}_2+\alpha)} > 0 \quad \forall\ \alpha > 0.$$

Since $\frac{\partial^2 \ell(\mu_1,\mu_2,\alpha)}{\partial \mu_1^2} < 0$, it can be said that $\bar{y}_1$ and $\bar{y}_2$ are the MLEs of μ_1 and μ_2 respectively. Since these are independent of α, we can say that $L(\hat{\mu}_1, \hat{\mu}_2, \alpha) \geq L(\mu_1, \mu_2, \alpha)$ for all μ_1 and $\mu_2 > 0$. Thus, to get the MLE of α, the profile likelihood (or log-likelihood) at these MLEs can be looked at and maximized with respect to α.

For PG trials assuming equal α, Figure 2.1 shows the plots of $g(\alpha) = \ell(\bar{y}_1, \bar{y}_2, \alpha)$ (Plot (a)) and $g'(\alpha) = \partial\ell(\mu_1, \mu_2, \alpha)/\partial\alpha$ at $\mu_1 = \bar{y}_1$ and $\mu_2 = \bar{y}_2$ (Plot (b)) as a function of α for one simulated dataset. It can be seen that the profile log-likelihood $g(\alpha)$ increases until a certain point and then decreases as α becomes larger. Plot (b) of Figure 2.1 shows that $g'(\alpha)$ decreases with α until a certain point (crossing zero

only once) and then increases as α becomes larger without ever reaching zero again. This shows that maximizing $g(\alpha)$ numerically with respect to α yields its MLE $\hat{\alpha}$ as does solving the equation $g'(\alpha) = 0$. Of course, in the above discussion, the log-likelihood function is in terms of (μ_1, μ_2, α) instead of the original parametrization (γ, μ, α) but since there is a one to one relation between the two parametrizations and because of the invariance property of the MLEs the above method can be used to obtain the MLEs under the (γ, μ, α) parametrization as well.

For PG trials with $\alpha_1 \neq \alpha_2$, when γ is unknown, we use numerical maximization of ℓ with respect to α_1 and α_2, and when γ is known, we use numerical maximization with respect to all unknown parameters. A similar argument to the one given above can be used to show that this method gives the MLEs of the required parameters. In all these cases one could also obtain the MLEs identified through the solutions of relevant likelihood equations.

2.4 Testing for the Treatment Effect

The *treatment effect*, $100(1-\gamma)\%$, represents the proportion of reduction in the mean new lesion counts. If the treatment had no effect, γ would be equal to 1 and in RRMS trials the treatment is considered ineffective if $\gamma \geq 1$.

2.4.1 Equal Dispersion Parameters

We now present parametric tests for the general case $H_0 : \gamma = \gamma_0$ vs. $H_1 : \gamma \neq \gamma_0$. To test if the treatment is effective in the RRMS clinical trials, γ_0 is taken to be 1. Wang et al. [2001] have considered this comparison assuming the common dispersion parameter α is *known*. Then, the likelihood function given in (2.3.1) belongs to a

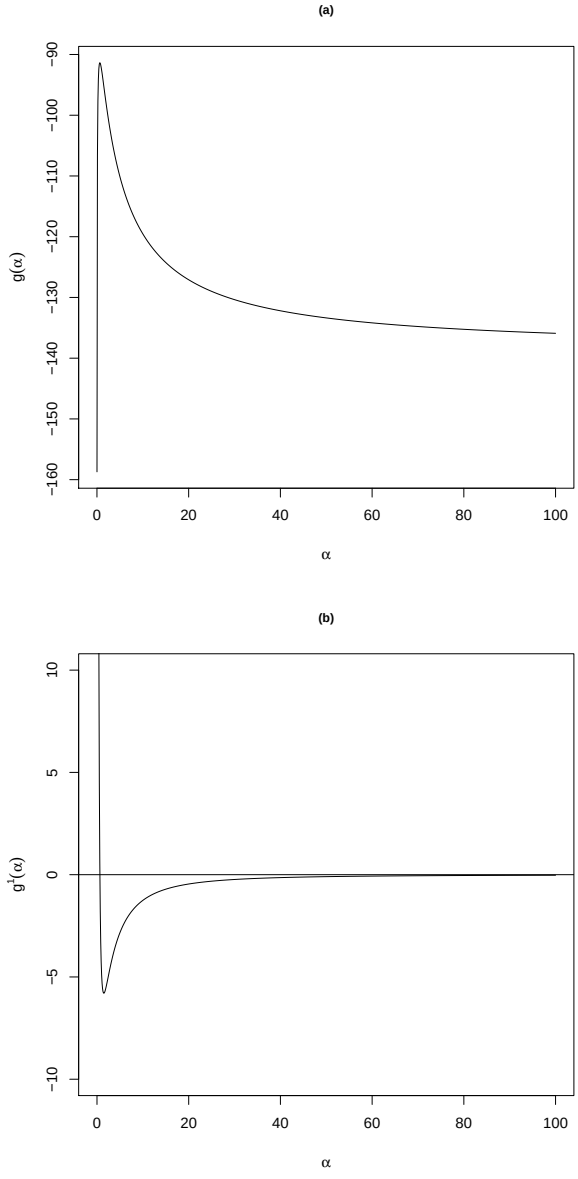

Figure 2.1: Plot of the log-likelihood $g(\alpha)$ and first derivative $g'(\alpha)$ for PG trials assuming equal dispersion parameters with μ_1 and μ_2 fixed at their MLEs $\bar{y}_1$ and $\bar{y}_2$ respectively.

two-parameter exponential family and consequently, they use classical results (see, e.g., Section 4.4 of Lehmann [1986]) to derive a uniformly most powerful unbiased (UMPU) test. But when α is unknown, it is not known whether a UMPU test exists.

Aban et al. [2009] derive the LRT for testing $H_0 : \gamma = \gamma_0$ vs. $H_1 : \gamma \neq \gamma_0$ for PG trials using the NB model with common unknown dispersion parameter. To construct the LRT, the MLEs of the parameters are needed under both Θ and Θ_0 specified by H_0. These are described in Lemmas 2.3.1 and 2.3.2, respectively.

Theorem 2.4.1. *The LRT statistic for testing $H_0 : \gamma = \gamma_0$ vs. $H_1 : \gamma \neq \gamma_0$ is given by*

$$L_{PG} = -2 \log \left(\frac{\sup_{\Theta_0} L(\gamma, \mu, \alpha)}{\sup_{\Theta} L(\gamma, \mu, \alpha)} \right) = -2(\ell(\gamma_0, \tilde{\mu}, \tilde{\alpha}) - \ell(\hat{\gamma}, \hat{\mu}, \hat{\alpha})).$$

When H_0 is true, this statistic is asymptotically distributed as a χ^2_1 rv as n_1, n_2 increase. An approximate level ν test rejects H_0 if $L_{PG} > \chi^2_1(1 - \nu)$.

For testing the treatment effect Aban et al. [2009] also propose and compare other *asymptotic* parametric tests such as the Rao's Score Test (RST) and Wald tests (using several transformations $g(\gamma)$ of γ such as γ, γ^2, and $\log(\gamma)$) for testing the treatment effect. The LRT needs the MLEs of parameters for both parameter spaces Θ and Θ_0. The Wald tests need them only for Θ whereas the RST needs MLEs only under Θ_0. While these three tests are asymptotically equivalent, they behave very differently for small samples and the χ^2 critical value may fail to maintain the assumed significance level. Using simulation methods these authors show that the LRT and Wald tests are not appropriate for sample sizes under 50 as they have a highly inflated Type I error rate. For larger sample sizes these tests have more power than the RST and their significance levels are close to the nominal level. Although RST is less powerful than either the LRT or the Wald test, its Type I error rate is well below the nominal

5% level even for small sample sizes. Their paper recommends the use of RST when individual sample sizes are under 50, and either the LRT or the Wald test that tests $\log \gamma = 0$ for larger sample sizes.

The PROC GENMOD procedure in SAS can be used to compare the means of two independent NB populations assuming equal dispersion parameters. It provides LRT, and WT(log(γ)) when we use log link function and takes $c_0 = \chi^2_1(1-\nu)$ as the critical value. With $\mu_1 = \mu$ and $\mu_2 = \gamma\mu$ as the parameters of interest, PROC GENMOD can be used with appropriate link functions to generate LRT and Wald tests for $H_0 : g(\mu_1) = g(\mu_2)$ where $g(\mu) = \mu, \mu^2$ or $\sqrt{\mu}$. However those Wald tests are different from the Wald tests for $g(\gamma)$ while the LRT is invariant of these transformations. See also Remark 7 of Aban et al. [2009].

2.4.2 Unequal Dispersion Parameters

When α_1 and α_2 are not equal and assumed to be known, the likelihood function can still be shown to be a two-parameter exponential family and a two-sided UMPU test for $H_0 : \gamma = 1$ can be derived using the results given in Section 4.4 of Lehmann [1986]. The above likelihood based tests can be derived when the dispersion parameters are unknown and assumed unequal. They are described next.

Theorem 2.4.2. *(LRT). When $\alpha_1 \neq \alpha_2$, the LRT statistic to test $H_0 : \gamma = \gamma_0$ vs. $H_1 : \gamma \neq \gamma_0$ is given by $-2(\ell(\gamma_0, \tilde{\mu}, \tilde{\alpha}_1, \tilde{\alpha}_2) - \ell(\hat{\gamma}, \hat{\mu}, \hat{\alpha}_1, \hat{\alpha}_2))$, where the MLEs are described in Lemmas 2.3.4 and 2.3.5. The LRT statistic is asymptotically distributed as a χ^2_1 rv under H_0.*

Theorem 2.4.3. *(RST). When $\alpha_1 \neq \alpha_2$, the RST statistic for testing $H_0 : \gamma = \gamma_0$ vs. $H_1 : \gamma \neq \gamma_0$ is given by*

$$R_{PG} = \left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma}\right]^2_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} \cdot \sigma^2_{\hat{\gamma}}(\tilde{\boldsymbol{\theta}}) = \frac{n_2\tilde{\alpha}_2}{n_1\tilde{\alpha}_1\gamma_0\tilde{\mu}} \frac{(\bar{y}_2 - \gamma_0\tilde{\mu})^2}{(\gamma_0\tilde{\mu} + \tilde{\alpha}_2)^2} \left[n_1\tilde{\alpha}_1(\gamma_0\tilde{\mu} + \tilde{\alpha}_2) + n_2\tilde{\alpha}_2\gamma_0(\tilde{\mu} + \tilde{\alpha}_1)\right]$$

where $\tilde{\mu}$, $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ are the MLEs under H_0 of μ, α_1 and α_2, respectively. Then an approximate level ν test rejects H_0 if $R_{PG} > \chi^2_1(1-\nu)$.

Theorem 2.4.4. *(Wald Test ($WT(g(\gamma))$, See Rao [2005], Sec 1.2). When $\alpha_1 \neq \alpha_2$ and $g(\cdot)$ is a $1-1$ function with nonzero derivative, a Wald Test for testing $H_0 : g(\gamma) = g(\gamma_0)$ vs. $H_1 : g(\gamma) \neq g(\gamma_0)$ is based on*

$$W_{PG}(g(\gamma)) = \left[\frac{g\left(\bar{y}_2/\bar{y}_1\right) - g(\gamma_0)}{g\prime(\hat{\gamma})\sigma_{\hat{\gamma}}(\hat{\theta})}\right]^2$$

where

$$\sigma^2_{\hat{\gamma}}(\hat{\theta}) = \frac{\hat{\gamma}\left[n_1\hat{\alpha}_1(\hat{\gamma}\hat{\mu} + \hat{\alpha}_2) + n_2\hat{\alpha}_2\hat{\gamma}(\hat{\mu} + \hat{\alpha}_1)\right]}{n_1 n_2 \hat{\alpha}_1\hat{\alpha}_2\hat{\mu}}$$

is the asymptotic variance of $\hat{\gamma}$ (see Lemma 2.3.7) evaluated at the MLEs. An approximate level ν test rejects the null hypothesis H_0 if $W_{PG}(g(\gamma)) > \chi^2_1(1-\nu)$.

2.4.3 Exact Parametric Tests and Critical Values

The choice of c_0 (= 3.8415) as the critical value for a 5% significance level may be inappropriate for small studies. For small PG trials, as observed by Aban et al. [2009], it may lead to very liberal (LRT, WT) or too conservative (RST) tests. A more precise critical value (referred to as the *exact* critical value from now on) of the test statistic can be obtained through simulation. For example, for a given sample size, the LRT statistic is simulated from the null distribution a large number of times (say 200,000). The $100(1-\nu)^{\text{th}}$ sample percentile provides an exact critical value for a $100\nu\%$ level test. The null hypothesis is then rejected if the LRT statistic is greater

than the simulated critical value. The properties of such exact and approximate LRTs, RSTs and Wald tests for PG trials are compared in Section 2.5.

For the PG trials, the simulated 95th percentile values for the different tests are displayed as a function of the sample sizes in Figure 2.2. There we take $\mu = 5.9$, $\alpha = 0.49$; these values are suggested by data sets from RRMS studies Sormani et al. [2001b]. For the LRT and Wald tests they are well above c_0 for small samples in all cases. As expected, the exact percentile values converge to c_0 as sample sizes increase although the convergence is quicker for the LRT than for any of the Wald tests. The convergence appears to be the slowest for the Wald test for γ^2. The exact percentile values for the RST are below c_0 for small samples sizes. This means RST tends to be conservative while the other tests fail to maintain the nominal level for small samples.

We conducted simulation studies to check on the effect of μ and α on the simulated percentiles of LRT and $WT(\gamma^2)$, and actual Type I error rates when one uses c_0 as the critical value. We considered several values of μ ranging from 1 to 10 and α ranging from 0.2 to 5. Sample sizes of 10, 20, 50 and 100 subjects per group were considered. For the LRT, μ has little to no effect on the simulated percentiles or the Type I error rates (Figures 2.3 and 2.4). Increasing α results in smaller simulated percentiles and consequently reduced levels for PG trials. For smaller sample sizes ($n = 10, 20$) it can be seen that the *exact* percentiles are well above c_0. For larger sample sizes they seem to be reasonably close to c_0. The 'effect' of these *exact* percentiles can be seen on the asymptotic Type I error rates. For small sample sizes ($n = 10, 20$) the Type I error rates using the asymptotic LRT seem to be very high even for larger values of α. For $n = 50$, the Type I error rates seem to higher than the nominal level when α is very small. Thus, for sample sizes of 50, it is still advisable to use *exact* percentile based

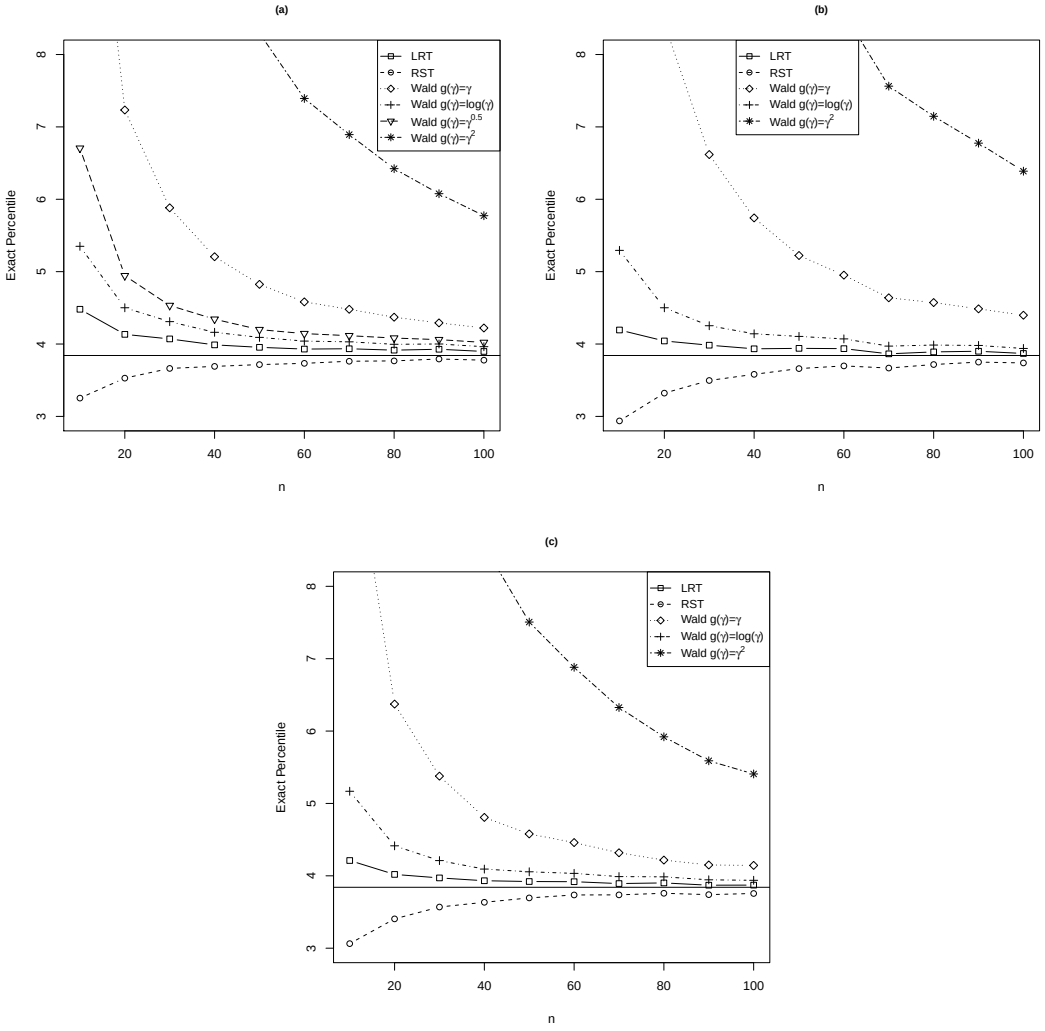

Figure 2.2: PG Trial: Simulation based 95th percentile value for the null distribution of LRT, RST and Wald test statistics as a function of common sample size n; $\mu = 5.9$, $\alpha_1 = 0.49$ and (a) PG Trial with $\alpha_2 = \alpha_1$, (b) PG Trial with $\alpha_2 = 0.75\alpha_1$, (c) PG trial with $\alpha_2 = 1.25\alpha_1$. The solid horizontal line refers to $c_0 = 3.8415 (= \chi^2_1(0.95))$.

tests if α is very small. For $n = 100$ per group the Type I error rates seem to be very close to 0.05 for all values of μ and α and hence the asymptotic approximation can safely be used.

For $WT(\gamma^2)$ (Figures 2.5 and 2.6)), the effects are more pronounced. Here increasing μ does not affect the simulated percentiles or the levels. Increasing α brings exact percentile values closer to c_0. In all the cases considered, the effect of α decreases with increasing sample size. The asymptotic Type I error rates using the WT(γ^2) seem to be much higher than the nominal level of 5% even for very high sample sizes. Unless the sample size is very large (100 or greater) and α is very large (5 or greater) it is not advisable to use asymptotic approximation for the WT(γ^2).

The Type I error rates and *exact* percentile values for the WT(log(γ)) (Figures not shown) were very similar to those for the LRT. For the RST (Figures not shown), for sample sizes less than 50, the exact percentile values were less than c_0 leading to conservative tests for all values of α. For sample sizes >= 50, the exact percentiles are very close to c_0 leading to Type I error rates close to nominal level of 0.05 except for $\alpha = 0.2$. We did not consider the other tests as they were not the "best" tests as will be seen in the next section (see Figure 2.7).

2.5 Power Analysis

Sormani et al. [2001b] used the NB model and simulation methods to enumerate PG trial sample sizes for the nonparametric WRS test. For this purpose they used parameter estimates from large data sets from control groups from RRMS trials. One can use the associated parametric assumptions to produce more powerful parametric tests. Aban et al. [2009] show the advantages of using likelihood based asymptotic tests assuming the NB model with equal dispersion parameter. We now compare the power and sample size estimates of our exact tests and of the asymptotic tests for PG trials with both equal and unequal dispersion parameters. For the equal dispersion

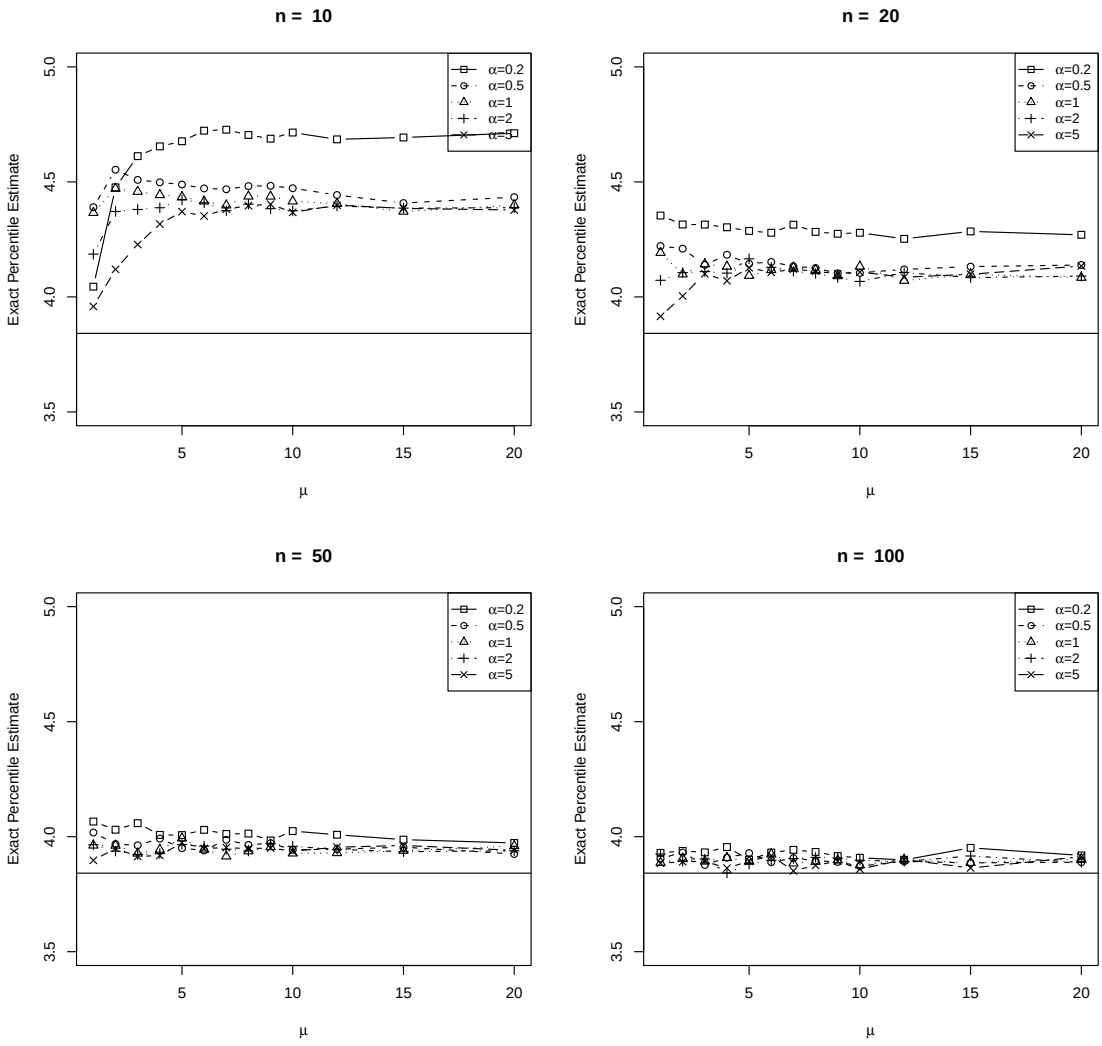

Figure 2.3: PG Trial: Simulation based 95th percentile values for the null distribution of LRT for different values of μ and α and $n = 10, 20, 50, 100$ subjects per group. The solid horizontal line refers to the $c_0 = 3.8415 (= \chi^2_1(0.95))$.

case, the estimated sample sizes are compared to the sample sizes given by Sormani et al. [2001b]. We use their parameter estimates for the control group, and as they have done, we assume monthly scans and use 3 and 6 month observation periods.

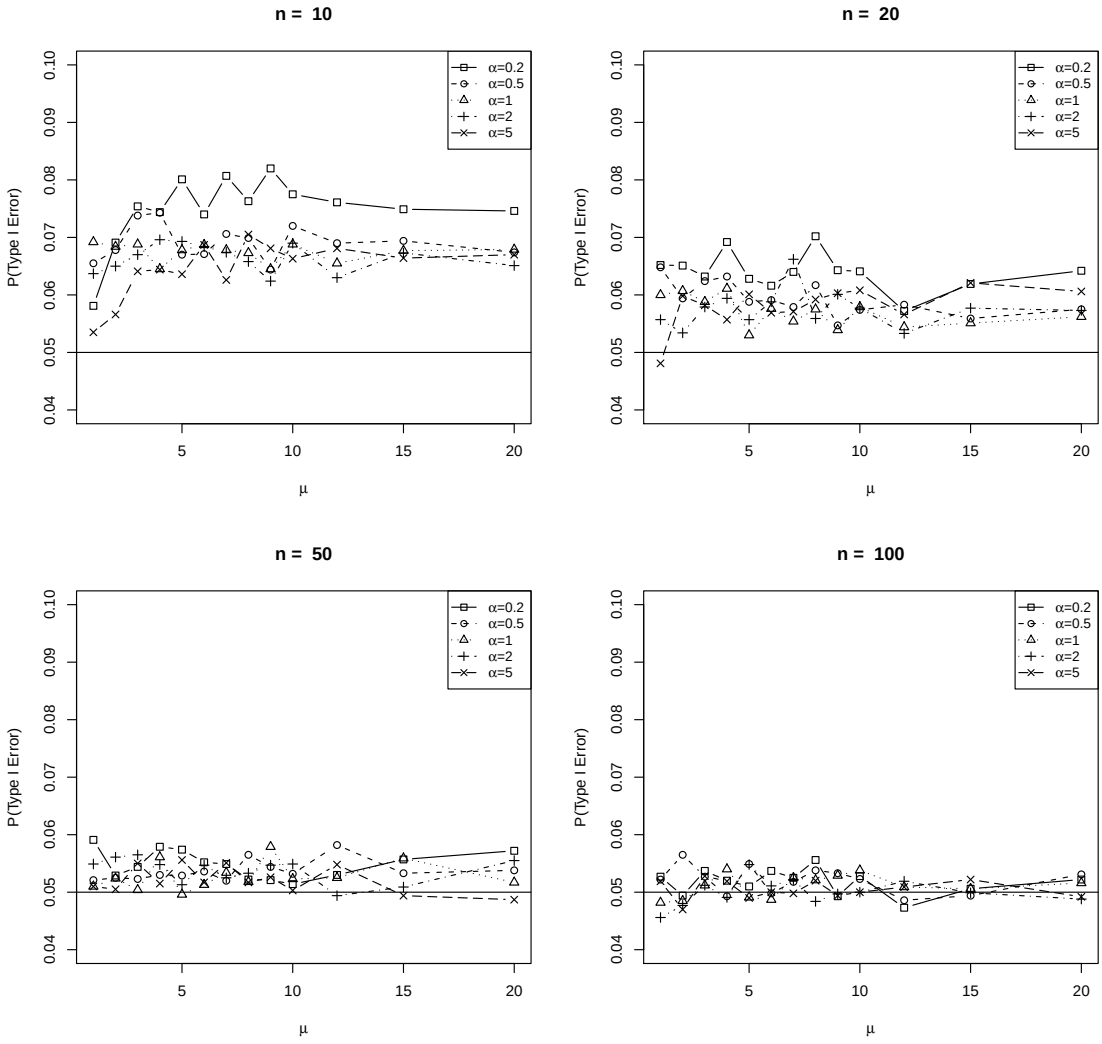

Figure 2.4: PG Trial: Type I error rates for LRT using asymptotic approximation for different values of μ and α and $n = 10, 20, 50, 100$ subjects per group. The solid horizontal line refers to nominal level $\nu = 0.05$.

Figure 2.7 presents the power curves as a function of γ for the various tests using exact critical values for PG trials assuming equal dispersion parameter across the two groups. The null hypothesis takes $\gamma_0 = 1$ and we use the initial parameter estimates

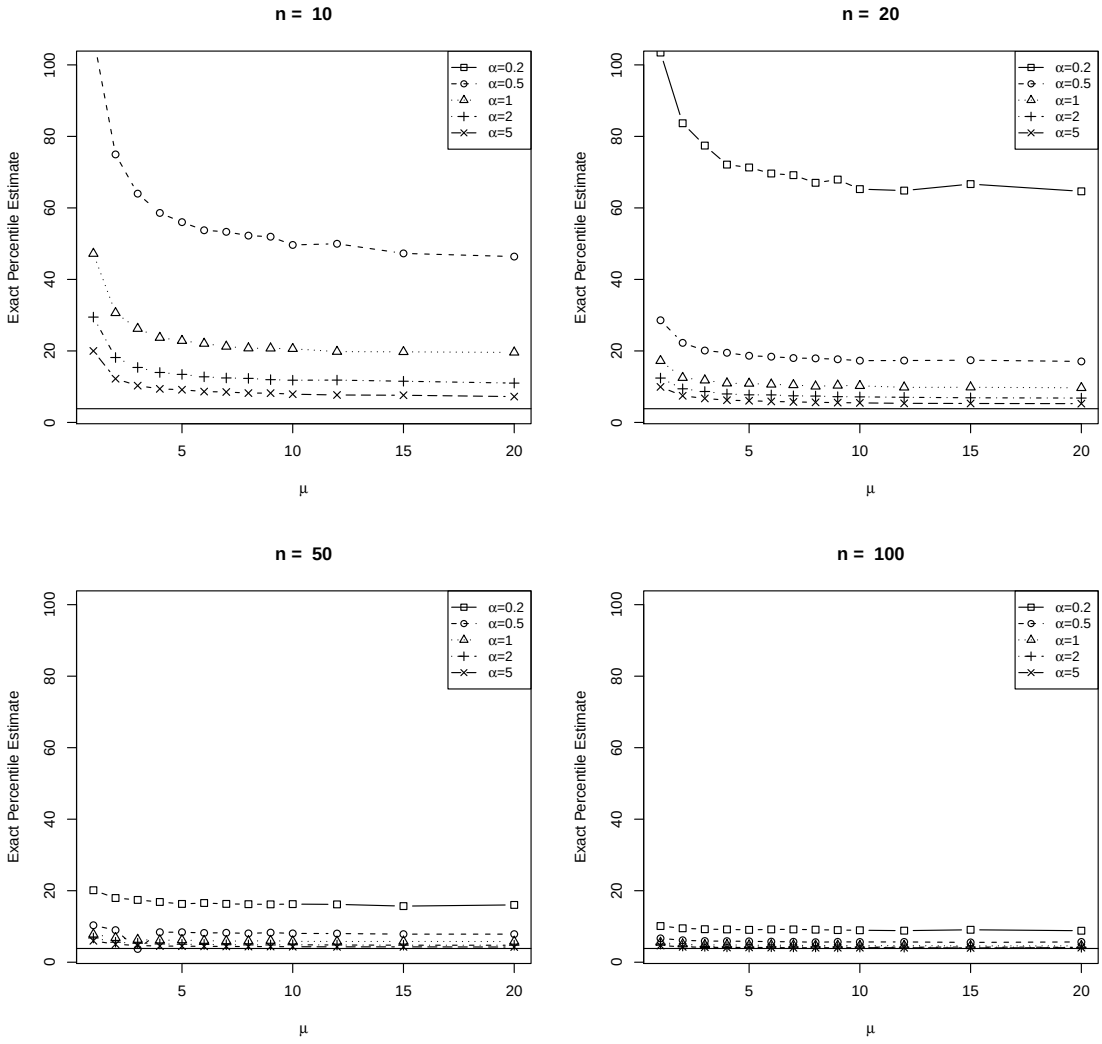

Figure 2.5: PG Trial: Simulation based 95th percentile value for the null distribution of WT(γ^2) for different values of μ and α and $n = 10, 20, 50, 100$ subjects per group. The solid horizontal line refers to the $c_0 = 3.8415 (= \chi^2_1(0.95))$.

of $\mu = 5.9, \alpha = 0.49$ for the control group as suggested by Sormani et al. [2001b] (for 3 month trials). The power curves correspond to $n_1 = n_2 = 50$, and a nominal level $\nu = 0.05$ based on simulated critical values reported in Figure 2.2 (a). The power

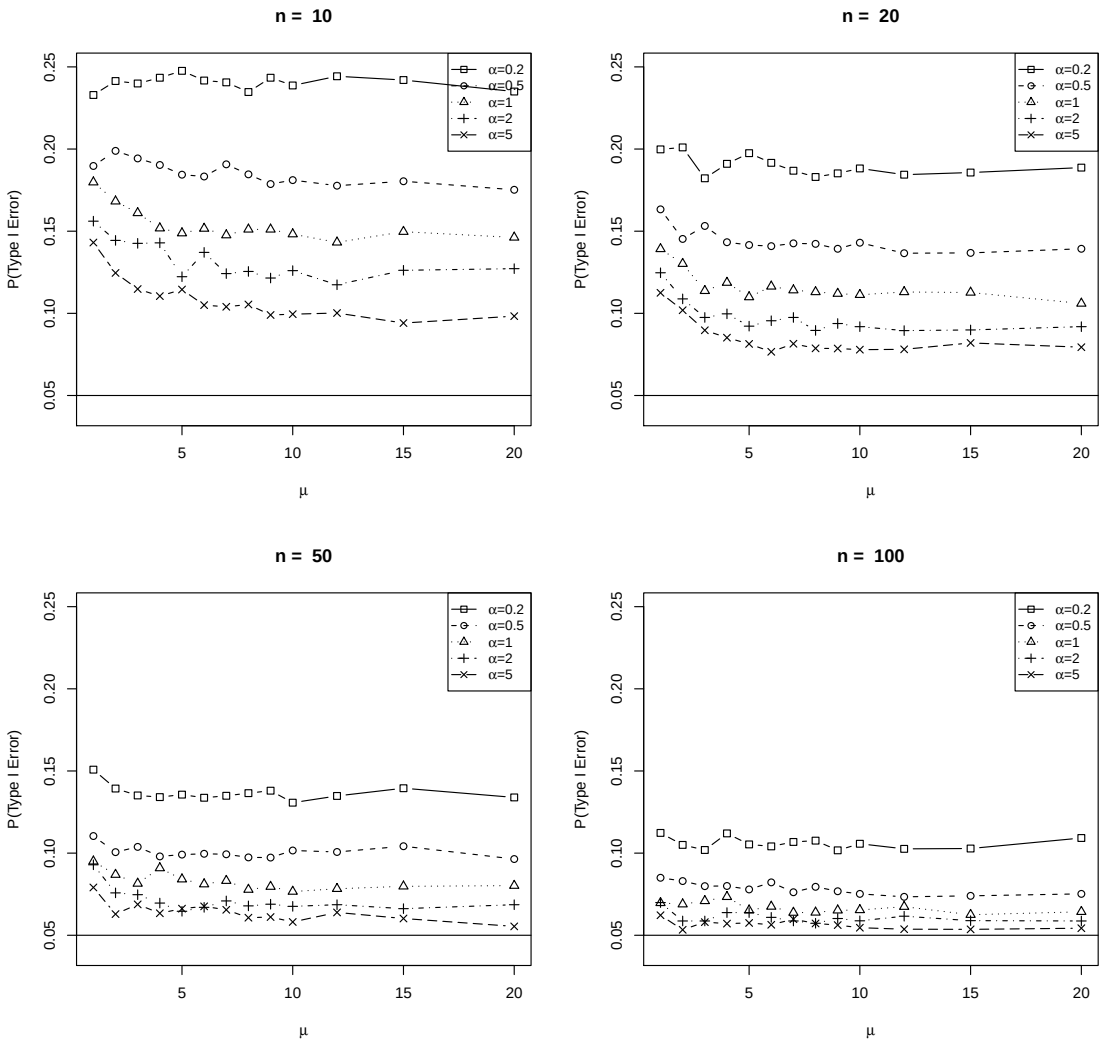

Figure 2.6: PG Trial: Type I error rates for WT(γ^2) using asymptotic approximation for different values of μ and α and $n = 10, 20, 50, 100$ subjects per group. The solid horizontal line refers to nominal level $\nu = 0.05$.

estimates for LRT, RST and the WT(log(γ)) are very close and show that they are unbiased (empirically). For $\gamma < 1$, the WT(γ^2) has a slightly higher power than the WT(γ) and both have better power than the three unbiased tests. For $\gamma > 1$, the

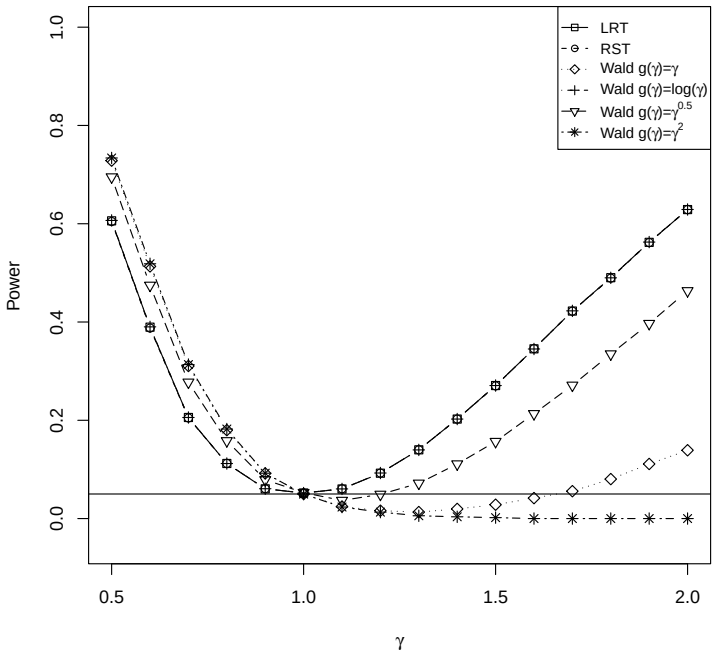

Figure 2.7: PG Trial, Common Dispersion: Power of Exact 5% level LRT, RST, and Wald tests for treatment effect, assuming (i) initial parameter estimates of $\mu = 5.9, \alpha = 0.49$, and (iii) $n_1 = n_2 = 50$. The solid horizontal line refers to nominal level $\nu = 0.05$.

power for WT(γ^2) goes to zero very quickly. For the test for γ it slowly rises above the 5% level and remains well below the power estimates for the unbiased tests. For the RRMS clinical trials where the treatment is expected to reduce the number of new enhancing lesions, the emphasis is on the region $\gamma < 1$, and hence the Wald test for γ^2 is the most suitable test for detecting such a change. It also has a desirable property of small power for $\gamma > 1$ since if the treatment results in increased counts, the Wald

test fails to reject H_0 with probability close to 1. The WT($\sqrt{\gamma}$) is not unbiased and does not excel in any region under H_1, and hence is not considered further.

2.6 Sample Size Estimation

In Tables 2.1-2.3, sample sizes required to achieve 80% and 90% power for LRT and the WT(γ^2) are obtained under three different scenarios. For the LRT both the chi-squared percentile (c_0) and the exact percentile obtained through simulation are used. For the Wald test sample sizes are presented using only the exact percentile as c_0 is a very poor underapproximation even for large sample sizes. The sample sizes obtained by Sormani et al. [2001b] using the nonparametric Wilcoxon Rank Sum (WRS) test are given for comparison. The exact percentiles for the LRT and the Wald test for the corresponding sample sizes are also given. They are helpful in carrying out the tests when the associated data sets are available at the end of the study.

The power was computed as the proportion of trials out of $B = 10,000$ that yielded a significant result. Due to the inherent natural variation of this type of simulation, one standard error above the estimate obtained ($p + \sqrt{p(1-p)/B}$) is used in determining whether the required power is attained. That is, for $B = 10{,}000$ trials, when the required power is 80% (say), a simulated power of 79.6% or higher is considered enough and when the required power is 90%, a simulated power of 89.7% is considered sufficient. Two-sided significance level was set at 5% and only equal sample sizes ($n_1 = n_2$) were considered in our power analyses.

The MLEs of μ and α of the NB model were obtained for a follow-up period (t) of 3 and 6 months (Sormani et al. [2001b]). They were, respectively, 5.9 and 0.49

for $t = 3$, and 13 and 0.52 for $t = 6$. Sormani et al. [2001b] use these estimates and nominal 5% significance level to obtain sample sizes using the WRS test. We use the same parameter estimates in our sample size calculations.

Sample sizes for the equal dispersion parameter case are given in Table 2.1 as $(1 - \gamma)$ ranges from 0.50 to 0.80. There is a 30-45% reduction in sample sizes by the use of LRT when compared to those using the nonparametric test for the treatment effects and follow-up time periods considered. The exact and asymptotic 5% level LRTs produce almost identical sample sizes when they are large. The latter test produces slightly lower n when the sample sizes are small; it also inflates the level beyond the nominal 5%. The 5% level WT(γ^2) reduces the required sample sizes by further 18-28% resulting in about half the sample sizes demanded by the WRS test. As expected, the sample sizes for the 6 month follow-up period are smaller than those needed for the 3 month follow-up period. But the doubling of the follow-up period from 3 to 6 months reduces the sample sizes by only about 10-15%. This is understandable since the number of new enhancing lesions seen during the first three months tend to be highly correlated with the number of new enhancing lesions seen during the next 3 months. Increased power from 80% to 90% will require an increase in sample sizes of 22-36% for the LRT and 30-40% for the WT(γ^2).

Similar sample size estimates are also obtained for the unequal dispersion parameter case. For Case (a): $\alpha_2 = 0.75\alpha_1$, sample sizes for the LRT (exact and asymptotic) and WT(γ^2) are given in Table 2.2. When compared to the equal dispersion case, the sample sizes have increased by about 12-27%. Table 2.3 provides a similar comparison for Case (b): $\alpha_2 = 1.25\alpha_1$. Here the required sample sizes have decreased by 10% or less. As seen in Figure 2.9, the power curves for the LRT, RST and WT(log(γ)) are

Table 2.1: PG Trial: Sample sizes per group to achieve 80% and 90% power for $100(1-\gamma)\%$ treatment effect and follow-up period (t) of 3 and 6 months assuming equal dispersion parameter, and initial estimates $(\mu, \alpha) = (5.9, 0.49)$ for $t = 3$, and $(13, 0.52)$ for $t = 6$; level $\nu = 0.05$.

$1-\gamma$	t	80% power; 5% level					
		Test				Critical Value	
		WRS	LRT Asymptotic	LRT Exact	WT(γ^2) Exact	LRT Exact	WT(γ^2) Exact
0.50	3	125	76	76	61	3.892	7.354
	6	118	68	69	54	3.928	7.420
0.60	3	75	45	45	35	3.988	10.821
	6	65	39	40	32	3.724	10.790
0.70	3	48	27	28	21	4.083	17.612
	6	40	24	25	19	4.073	17.731
0.80	3	28	16	18	13	4.162	34.260
	6	24	14	15	11	4.242	38.103
$1-\gamma$	t	90% power; 5% level					
		Test				Critical Value	
		WRS	LRT Asymptotic	LRT Exact	WT(γ^2) Exact	LRT Exact	WT(γ^2) Exact
0.50	3	150	102	103	83	3.908	6.298
	6	140	90	92	73	3.914	6.410
0.60	3	98	60	60	49	3.939	8.499
	6	85	52	52	43	3.935	8.573
0.70	3	64	36	36	28	4.021	13.143
	6	58	32	33	25	4.050	13.366
0.80	3	40	22	22	17	4.146	23.177
	6	35	19	19	15	4.133	24.106

Table 2.2: PG Trial: Sample sizes per group to achieve 80% and 90% power for $100(1-\gamma)\%$ treatment effect and follow-up period (t) of 3 and 6 months with unequal dispersion parameters and initial estimates of $(\mu, \alpha_1, \alpha_2)$ at $(5.9, 0.49, 0.3675)$ for $t = 3$, and $(13, 0.52, 0.39)$ for $t = 6$; level $\nu = 0.05$.

$1-\gamma$	t	80% power; 5% level				
		Test			Critical Value	
		LRT Asymptotic	LRT Exact	WT(γ^2) Exact	LRT Exact	WT(γ^2) Exact
0.50	3	87	89	68	3.902	7.765
	6	79	80	62	3.883	7.967
0.60	3	52	53	41	3.915	11.267
	6	47	47	37	3.936	11.702
0.70	3	33	33	25	3.932	18.797
	6	28	30	22	3.996	19.539
0.80	3	21	22	15	4.017	39.749
	6	18	19	13	4.083	43.209
$1-\gamma$	t	90% power; 5% level				
		Test			Critical Value	
		LRT Asymptotic	LRT Exact	WT(γ^2) Exact	LRT Exact	WT(γ^2) Exact
0.50	3	115	115	95	3.862	6.534
	6	104	104	83	3.905	6.694
0.60	3	69	70	55	3.892	8.977
	6	62	63	49	3.911	9.158
0.70	3	43	43	34	3.951	13.139
	6	36	38	30	3.976	13.719
0.80	3	27	27	20	3.992	25.066
	6	22	23	17	4.010	27.252

Table 2.3: PG Trial: Sample sizes per group to achieve 80% and 90% power for $100(1-\gamma)\%$ treatment effect and a follow-up period (t) of 3 and 6 months with unequal dispersion parameters and initial estimates of $(\mu, \alpha_1, \alpha_2)$ at $(5.9, 0.49, 0.6125)$ for $t = 3$ and $(13, 0.52, 0.65)$ for $t = 6$; level $\nu = 0.05$.

$1-\gamma$	t	80% power; 5% level				
		Test			Critical Value	
		LRT Asymptotic	LRT Exact	WT(γ^2) Exact	LRT Exact	WT(γ^2) Exact
0.50	3	68	70	53	3.920	7.209
	6	62	62	48	3.911	7.252
0.60	3	41	42	31	3.945	10.212
	6	36	37	28	3.950	10.294
0.70	3	26	26	19	3.991	15.904
	6	22	23	17	4.037	16.462
0.80	3	17	18	12	4.053	28.975
	6	14	15	10	4.128	32.698
$1-\gamma$	t	90% power; 5% level				
		Test			Critical Value	
		LRT Asymptotic	LRT Exact	WT(γ^2) Exact	LRT Exact	WT(γ^2) Exact
0.50	3	93	94	76	3.881	6.074
	6	82	83	66	3.900	6.161
0.60	3	55	55	44	3.904	8.052
	6	47	48	40	3.939	8.058
0.70	3	33	35	26	3.963	11.756
	6	29	30	24	3.975	11.816
0.80	3	21	22	16	4.068	19.176
	6	18	18	14	4.043	20.107

very close and consequently the last two tests produce very similar sample sizes and they differ from the LRT sample sizes by at most 2. The sample sizes for the Wald test for γ^2 are around 17-32% smaller than the sample sizes needed for LRT.

2.7 A Robustness Study

A robustness study was done to evaluate the performance of these exact parametric tests which assume equal dispersion parameter $\alpha = 0.49$ across the two groups. Two cases with $\alpha_1 = 0.49$ were considered: (a) $\alpha_2 = 0.5\alpha_1$ and (b) $\alpha_2 = 2\alpha_1$. In case (a) [(b)] variability in the treatment group response is increased [decreased] when compared to the equal α case. Figure 2.8 (a) and (b) show the power curves for these parametric tests for cases (a) and (b) respectively for $n_1 = n_2 = 50$. When $\alpha_2 = 0.5\alpha_1$, the power estimates are slightly lower than the corresponding estimates for the equal dispersion case (given in Figure 2.7) while the simulated Type I error rates are higher than the nominal 5% level. The departure is the most serious for the WT(γ^2) but it has the highest power for $\gamma < 1$. WT(γ) behaves very similar.

As Figure 2.8 (b) shows, the estimated power when $\alpha_2 = 2\alpha_1$ is higher than when $\alpha_2 = \alpha_1$. The WT(γ^2) has Type I error rate below the 5% level at 0.0367 but still has a higher power than other tests for $\gamma < 1$. The Type I error rates for LRT, RST and WT(log(γ)) are still slightly above 5%. Thus the LRT, RST and WT(log(γ)) seem to be more robust when the suspicion is that $\alpha_2 = 0.5\alpha_1$ and the Wald test for γ^2 is the best test when $\alpha_2 = 2\alpha_1$ and the region of interest is $\gamma < 1$. In the latter case, the treatment reduces the mean as well as the variance of the new lesion counts.

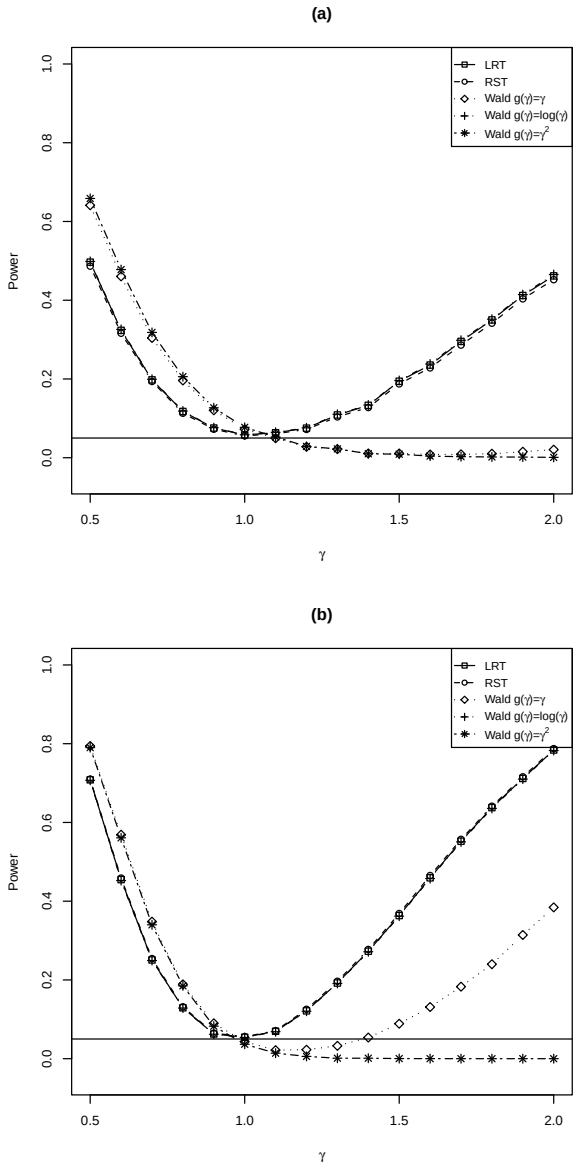

Figure 2.8: PG Trial, Robustness Study: Powers of exact parametric tests with 5% level that assume equal dispersion parameter across groups when the data come from groups with unequal dispersion parameters; $\mu = 5.9, \alpha_1 = 0.49, n_1 = n_2 = 50$; (a) $\alpha_2 = 0.5\alpha_1$ (b) $\alpha_2 = 2\alpha_1$. The solid horizontal line refers to the nominal level $\nu = 0.05$.

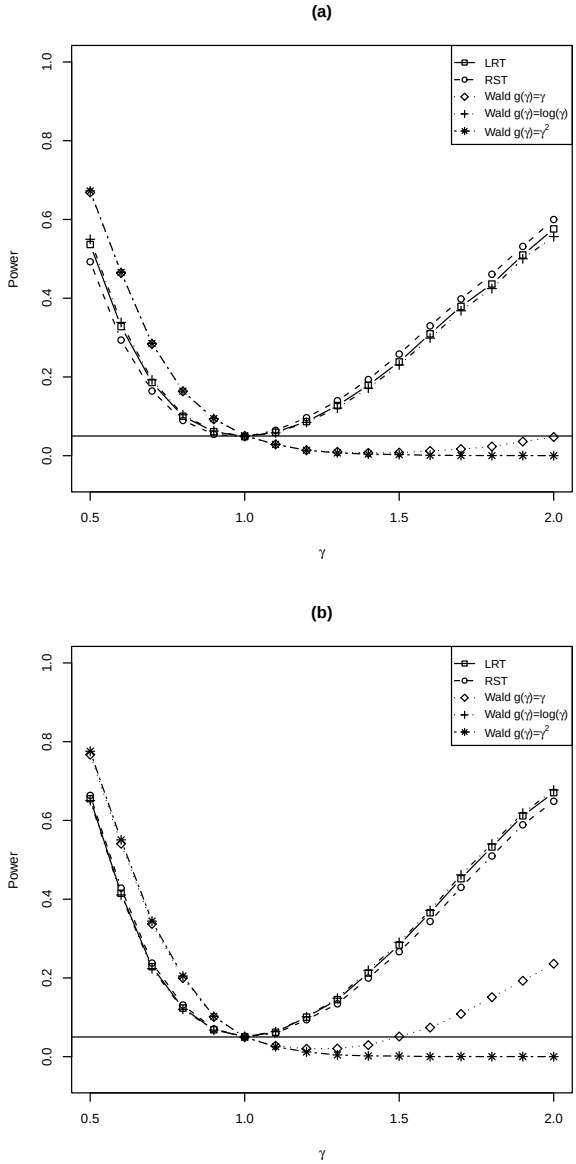

Figure 2.9: PG Trial, Unequal Dispersion: Power of Exact 5% level LRT, RST, and Wald tests for treatment effect, assuming initial parameter estimates $\mu = 5.9, \alpha_1 = 0.49$, sample sizes $n_1 = n_2 = 50$, and (a) $\alpha_2 = 0.75\alpha_1$ (b) $\alpha_2 = 1.25\alpha_1$. The solid horizontal line refers to the nominal level $\nu = 0.05$.

Unequal Dispersion Parameters

Figures 2.9 (a) and (b) show the power curves for five parametric tests for cases (a) $\alpha_2 = 0.75\alpha_1$ and (b) $\alpha_2 = 1.25\alpha_1$, respectively. In both these cases Wald test for γ^2 has the highest power for $\gamma < 1$ and for $\gamma > 1$ its power quickly goes to 0 as in the equal α case. The LRT, RST and WT($\log(\gamma)$) are unbiased. For case (a), the RST has the highest power for $\gamma > 1$ but has the least power for $\gamma < 1$. The WT($\log(\gamma)$) has a slightly better power than the LRT for $\gamma < 1$ but for $\gamma > 1$ the LRT does better. For case (b), the power of the three unbiased tests are very close to each other although the RST seems to have the highest power for $\gamma < 1$ and the LRT has the highest power for $\gamma > 1$. These differences in power among the unbiased tests observed in Figures 2.9 (a) and (b) are magnified respectively when $\alpha_2 = 0.5\alpha_1$ and $\alpha_2 = 2\alpha_1$ (figure not shown). Thus, for the unequal dispersion parameter case, the use of an appropriate test is suggested based on the region of interest and whether or not an unbiased test is desired.

2.8 Discussion

Here we have assumed NB models and used large data sets on new monthly MRI lesion counts to produce sample size calculations for future RRMS trials. Since the estimates of the parameters were obtained by fitting the NB model to RRMS patients not selected for MRI activity at baseline, it is highly suggested that the sample sizes reported here be applied only to clinical trials involving a similar group of RRMS patients. The parametric tests described in Section 2.4 are two-sided tests which test $H_0 : \gamma = 1$ vs. $H_1 : \gamma \neq 1$. Although the intrinsic research hypothesis is one-sided

($\gamma < 1$) as it is often believed the treatment could not harm the patient, the FDA prefers a two-sided test in order to be more conservative.

For PG trials with equal dispersion parameter α across groups, one of the three unbiased tests (LRT, RST and WT(log(γ))) can be used with exact percentiles. They are comparable in terms of computational efforts. For PG trials with unequal dispersion parameters across groups an unbiased test can be chosen based on the region of interest. For $\alpha_2 < \alpha_1$ case, when $\gamma < 1$, the WT(log(γ)) has the highest power among all unbiased tests considered, and when $\gamma > 1$ the RST has the highest power. The reverse is true when $\alpha_2 > \alpha_1$. The use of LRT has resulted in a 30-45% reduction in sample sizes for PG trials when compared to those based on nonparametric tests. When unbiasedness of a test is not important, as in the case of RRMS clinical trials where the researcher is interested in only one side of the null hypothesis, using the Wald test for γ^2 results in a further reduction in sample sizes by 10-30% when compared to the LRT.

Another important thing to note is that, for PG trials, increasing the mean parameter μ from 5.9 to 13 (or the follow-up period from 3 to 6 months) without much change in α, the sample sizes required to detect a significant effect seem to reduce by 10-15%. This is because, increasing the follow-up period adds more redundant data and thus does not result in a corresponding reduction in sample size.

So far only nonparametric sample size estimates are being used in RRMS clinical trials. The results presented in this chapter show the advantages of using parametric tests when the NB model assumption seems reasonable. Obtaining monthly MRI scans of the brain for each patient and counting the number of new enhancing lesions

is a very expensive and tedious process and these results will help reduce the cost and effort involved significantly.

CHAPTER 3

Bivariate Negative Binomial Models for Baseline vs. Treatment Trials

3.1 Introduction

Many forms of bivariate negative binomial (BNB) distributions exist in the literature. Subramaniam and Subramaniam [1992] give a detailed description of the several chance mechanisms that give rise to a BNB distribution. A few of them are discussed here.

Suppose a sequence of independent trials result in one of three distinct outcomes labeled A, B or C, with probabilities $P(A) = p_1, P(B) = p_2$ and $P(C) = 1 - p_1 - p_2$. Let the number of trials of C be fixed at r, and let X denote the number of occurrences of A and Y be the number of occurrences of B before the r^{th} occurrence of C. Then the joint distribution of X and Y is bivariate negative binomial with pmf

$$P(x, y) = \frac{(r + x + y - 1)!}{(r - 1)!x!y!} p_1^x p_2^y (1 - p_1 - p_2)^r, \quad x = 0, 1, \ldots, y = 0, 1, \ldots. \tag{3.1.1}$$

The form of the pmf given in (3.1.1) can be seen from the fact that the first $x+y+r-1$ trials constitute a trinomial distribution with index parameter $(x + y + r - 1)$ and probabilities p_1, p_2 and $1 - p_1 - p_2$ and the last trial has to be C. This distribution

was first developed by Guldberg [1934]. The probability generating function for this distribution is

$$G_{X,Y}(s_1, s_2) = (1 - p_1 - p_2)^r (1 - p_1 s_1 - p_2 s_2)^{-r}. \tag{3.1.2}$$

It can be seen that each of the marginal distributions of X and Y is univariate negative binomial. Also,

$$\text{Cov}(x, y) = \frac{r p_1 p_2}{(1 - p_1 - p_2)^2}$$

and

$$\rho_{X,Y}(x, y) = \left\{ \frac{p_1 p_2}{(1 - p_1)(1 - p_2)} \right\}^{1/2}$$

which is always positive. The pgf of the conditional distribution of $X|Y = y$ is

$$\begin{aligned} G_X(s|y) =& \frac{G^{(0,y)}(s, 0)}{G^{(0,y)}(1, 0)} \\ & \text{where } \left[G^{(x,y)}(u, v) = \frac{\partial^{x+y}}{\partial s_1^x s_2^y} G(s_1, s_2)|_{s_1=u, s_2=v} \right] \\ =& (1 - p_1)^{r+y} (1 - p_1 s)^{-(r+y)} \end{aligned} \tag{3.1.3}$$

which is $\text{NB}\Big((r + y)\frac{p_1}{1-p_1}, r + y\Big)$. Thus the regression of X on Y is

$$E(X|Y = y) = r \frac{p_1}{1 - p_1} + y \frac{p_1}{1 - p_1}$$

which is linear in y with a positive slope.

Downton [1970] comes up with a bivariate geometric distribution arising out of the following mechanism. Suppose shocks arrive attacking a two component system. Let p_1 be the probability that a shock is received by the first component and p_2 be the probability that a shock is received by the second component and $1 - p_1 - p_2$ be the probability that a shock is received by both components. Let X be the number of shocks suffered by component 1 prior to the first failure and Y be the number of

shocks suffered by component 2 prior to the first failure. Then the joint distribution of (X, Y) is bivariate geometric with pmf

$$P(x,y) = \binom{x+y}{x} p_1^x p_2^y (1-p_1-p_2), \quad x = 0,1,\ldots, y = 0,1,\ldots. \tag{3.1.4}$$

This model can also be generalized to a BNB distribution when X and Y represent the number of shocks to the components before the r^{th} failure of the system. This will lead to the bivariate negative binomial distribution given in (3.1.1).

In this chapter, however, we are interested in the BNB distribution arising out of compounding two independent Poisson random variables with a gamma random variable. Section 3.2 introduces such a BNB model and discusses its applications in modelling MRI lesion counts arising out of RRMS BVT trials. The MLEs of the model parameters are obtained in Section 3.3. Several parameteric tests for the treatment effect are proposed and compared in Section 3.4 and Section 3.5 respectively. Sample size estimates using the selected parametric tests are obtained in Section 3.6 and these are compared to sample sizes based on nonparametric tests. Finally, the results are summarized in Section 3.7.

3.2 The BNB Model

The compounding approach that generated the univariate NB model in Section 2.2 can be extended to the bivariate case where we compound two conditionally independent Poisson rvs with a Gamma rv. Assume that

$$X|Z = z \sim \text{Poisson}(\mu z) \text{ independent of } Y|Z = z \sim \text{Poisson}(\gamma\mu z), \quad \mu, \gamma > 0, \tag{3.2.1}$$

where Z is the random subject effect that is assumed to be a Gamma$(\alpha, 1/\alpha)$ rv. Then the joint distribution of (X, Y) is BNB with pmf

$$P_{X,Y}(x,y) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\mu^x}{x!} \frac{(\gamma\mu)^y}{y!} \frac{\Gamma(x+y+\alpha)}{(\mu+\gamma\mu+\alpha)^{x+y+\alpha}}, \quad x, y = 0, 1, 2 \ldots; \ \mu, \gamma, \alpha > 0. \tag{3.2.2}$$

Then $(X, Y) \sim$ BNB$(\mu, \gamma\mu, \alpha)$. Note that one can get the above expression from (3.1.1) with $p_1 = \frac{\mu}{\mu+\gamma\mu+\alpha}$, $p_2 = \frac{\gamma\mu}{\mu+\gamma\mu+\alpha}$ and $r = \alpha$. One thing to note is that, in (3.1.1), r by definition is a positive integer but α in (3.2.2) need not be an integer. The moment generating function (mgf) of the BNB distribution given in (3.2.1) can be obtained from the pgf given in (3.1.2) and using the parametric relations given before as follows:

$$\begin{aligned} M_{X,Y}(t_1, t_2) =& G_{X,Y}(e^{t_1}, e^{t_2}) \\ =& \left[\frac{\alpha}{\mu(1-e^{t_1}) + \gamma\mu(1-e^{t_2}) + \alpha} \right]^\alpha . \end{aligned} \tag{3.2.3}$$

The marginal distribution of X can be obtained from the above mgf using the relation

$$\begin{aligned} M_X(t_1) =& M_{X,Y}(t_1, 0) \\ =& \left[\frac{\alpha}{\mu(1-e^{t_1}) + \alpha} \right]^\alpha . \end{aligned} \tag{3.2.4}$$

This is the mgf of NB(μ, α). Similarly, the marginal distribution of Y can also be seen to be NB$(\gamma\mu, \alpha)$.

Several authors have used the compounding approach to produce families of BNB distributions. For example, Arbous and Kerrich [1951] and Bates and Neyman [1952] apply this BNB distribution to the study of accident proneness. Edwards and Gurland [1961] and Subramaniam [1966] relax the assumption of conditional independence given in (3.2.1) and assume X and Y are positively correlated.

In RRMS BVT trials, the subjects are observed for a specified period of time before and after the treatment is initiated. Assuming X (Y) to be the total number of new enhancing lesions seen during the baseline (treatment) period we apply the model given in (3.2.2).

3.3 Estimation

Assuming there are n subjects with data $(\mathbf{x}, \mathbf{y}) = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ the likelihood function of our BVT model is

$$\begin{aligned} L(\gamma, \mu, \alpha|(\mathbf{x}, \mathbf{y})) &= \prod_{i=1}^{n} P_{X,Y}(x_i, y_i) \\ &= \left\{ \frac{\alpha^\alpha}{\Gamma(\alpha)} \right\}^n \frac{\mu^{\sum x_i + \sum y_i}}{\prod_{i=1}^n x_i!} \frac{\gamma^{\sum y_i}}{\prod_{i=1}^n y_i!} \frac{\prod_{i=1}^n \Gamma(x_i + y_i + \alpha)}{(\mu + \gamma\mu + \alpha)^{\sum x_i + y_i + \alpha}}. \end{aligned} \tag{3.3.1}$$

The log-likelihood function is

$$\begin{aligned} \ell(\gamma, \mu, \alpha) =& n\{\alpha \log \alpha - \log \Gamma(\alpha)\} + n(\bar{x} + \bar{y}) \log \mu + n\bar{y} \log \gamma + \sum_{i=1}^{n} \log \Gamma(x_i + y_i + \alpha) \\ & - n(\bar{x} + \bar{y} + \alpha) \log(\mu + \gamma\mu + \alpha) - \sum_{i=1}^{n} \log(x_i!) - \sum_{i=1}^{n} \log(y_i!). \end{aligned} \tag{3.3.2}$$

The components of the score vector for the log-likelihood given in (3.3.2) are

$$\frac{\partial \ell(\gamma, \mu, \alpha)}{\partial \mu} = \frac{n(\bar{x} + \bar{y})}{\mu} - \frac{n(\bar{x} + \bar{y} + \alpha)}{\mu + \gamma\mu + \alpha}(1 + \gamma), \tag{3.3.3}$$

$$\frac{\partial \ell(\gamma, \mu, \alpha)}{\partial \gamma} = \frac{n\bar{y}}{\gamma} - \frac{n(\bar{x} + \bar{y} + \alpha)}{\mu + \gamma\mu + \alpha}\mu, \tag{3.3.4}$$

$$\begin{aligned} \frac{\partial \ell(\gamma, \mu, \alpha)}{\partial \alpha} =& -n[\psi(\alpha) - 1] + n \log \left(\frac{\alpha}{\mu + \gamma\mu + \alpha} \right) - \frac{n(\bar{x} + \bar{y} + \alpha)}{\mu + \gamma\mu + \alpha} \\ & + \sum_{i=1}^{n} \psi(x_i + y_i + \alpha). \end{aligned} \tag{3.3.5}$$

The MLEs of γ and μ obtained by setting the score equations (3.3.3)-(3.3.5) to 0 and solving for the parameters giving us the following result:

Lemma 3.3.1. *The MLEs of γ and μ are*

$$\hat{\gamma} = \frac{\bar{y}}{\bar{x}} \text{ and } \hat{\mu} = \bar{x}.$$

The MLE $\hat{\alpha}$ can be obtained by numerically maximizing the profile loglikelihood function $\ell(\hat{\gamma}, \hat{\mu}, \alpha)$ with respect to α.

Bates and Neyman [1952] have given the MLEs of the parameters of the BNB pmf given in (3.2.2) under a different parameterization.

When $\gamma = \gamma_0$ is known, the MLEs of μ and α are obtained by setting (3.3.4) and (3.3.5) to 0 and solving for the two parameters. The MLEs in this case are given by the lemma below.

Lemma 3.3.2. *For a known $\gamma = \gamma_0$, the MLE of μ is*

$$\tilde{\mu} = \frac{\bar{x} + \bar{y}}{1 + \gamma_0},$$

and $\tilde{\alpha}$, the MLE of α is obtained by numerically maximizing the profile loglikelihood $\ell(\gamma_0, \tilde{\mu}, \alpha)$ with respect to α.

The second order derivatives of the log-likelihood function in (3.3.2) are

$$\begin{aligned}
\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \gamma^2} &= -\frac{n\bar{y}}{\gamma^2} + \frac{n\gamma\mu(\bar{x}+\bar{y}+\alpha)}{(\mu+\gamma\mu+\alpha)^2}, \\
\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \gamma \partial \mu} &= -\frac{n\alpha(\bar{x}+\bar{y}+\alpha)}{(\mu+\gamma\mu+\alpha)^2} = \frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \mu \partial \gamma}, \\
\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \gamma \partial \alpha} &= -\frac{n\mu(\mu+\gamma\mu-\bar{x}-\bar{y})}{(\mu+\gamma\mu+\alpha)^2} = \frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \alpha \partial \gamma}, \\
\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \mu^2} &= -\frac{n(\bar{x}+\bar{y})}{\mu^2} + \frac{n(\bar{x}+\bar{y}+\alpha)(1+\gamma)^2}{(\mu+\gamma\mu+\alpha)^2}, \\
\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \mu \partial \alpha} &= -\frac{n(1+\gamma)(\mu+\gamma\mu-\bar{x}-\bar{y})}{(\mu+\gamma\mu+\alpha)^2} = \frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \alpha \partial \mu}, \\
\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \alpha^2} &= -n\psi'(\alpha) + \frac{n\mu(1+\gamma)}{\alpha(\mu+\gamma\mu+\alpha)} - \frac{n(\mu+\gamma\mu-\bar{x}-\bar{y})}{(\mu+\gamma\mu+\alpha)^2} \\
&\quad + \sum_{i=1}^{n} \psi'(x_i + y_i + \alpha).
\end{aligned} \tag{3.3.6}$$

Lemma 3.3.3. *The FIM* $\mathbf{I}(\boldsymbol{\theta})$ *where* $\left(I(\boldsymbol{\theta})\right)_{i,j} = -E\left\{\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j}\right\}$ *and* $\boldsymbol{\theta} = (\gamma, \mu, \alpha)$ *is the parameter vector. We use the fact that* $E(\bar{X}) = \mu$, $E(\bar{Y}) = \gamma\mu$ *and* $X_i's$, $Y_i's$ *are identical to get*

$$
\begin{aligned}
I_{11}(\boldsymbol{\theta}) &= -E\left\{\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \gamma^2}\right\} = \frac{n\mu(\mu+\alpha)}{\gamma(\mu+\gamma\mu+\alpha)},\\
I_{22}(\boldsymbol{\theta}) &= -E\left\{\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \mu^2}\right\} = \frac{n\alpha(1+\gamma)}{\mu(\mu+\gamma\mu+\alpha)},\\
I_{33}(\boldsymbol{\theta}) &= -E\left\{\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \alpha^2}\right\} = n\left\{\psi'(\alpha) - \frac{\mu(1+\gamma)}{\alpha(\mu+\gamma\mu+\alpha)}\right\} - nE\{\psi'(X_i+Y_i+\alpha)\}\\
I_{12}(\boldsymbol{\theta}) &= -E\left\{\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \gamma \partial \mu}\right\} = \frac{n\alpha}{\mu+\gamma\mu+\alpha} = I_{21}(\boldsymbol{\theta}), \quad \textit{and}\\
I_{ij}(\boldsymbol{\theta}) &= 0, \quad \textit{otherwise.}
\end{aligned}
$$

Lemma 3.3.4. *Proceeding as in the case of PG trials, the asymptotic variance of* $\hat{\gamma}$ *for the BNB model simplifies to*

$$\sigma^2_{\hat{\gamma}}(\boldsymbol{\theta}) = \frac{\gamma(1+\gamma)(\mu+\gamma\mu+\alpha)}{n\left[\mu(\mu+\alpha)(1+\gamma) - \alpha\gamma\right]}. \tag{3.3.7}$$

Remark: A similar argument to the one given in Chapter 2 for PG trials can be used to show that the estimation methods used in this section give the MLEs of the parameters as desired. When the model is parametrized in terms of (μ_1, μ_2, α), the solutions for μ_1 and μ_2 obtained from the score equations are $\bar{x}$ and $\bar{y}$ respectively. For a fixed α, the determinant of the second order derivative matrix of the log-likelihood evaluated at $\hat{\mu}_1 = \bar{y}_1$ and $\hat{\mu}_2 = \bar{y}_2$ is

$$
\begin{vmatrix} \frac{\partial^2 \ell(\mu_1,\mu_2,\alpha)}{\partial \mu_1^2} & \frac{\partial^2 \ell(\mu_1,\mu_2,\alpha)}{\partial \mu_1 \partial \mu_2} \\ \frac{\partial^2 \ell(\mu_1,\mu_2,\alpha)}{\partial \mu_1 \partial \mu_2} & \frac{\partial^2 \ell(\mu_1,\mu_2,\alpha)}{\partial \mu_2^2} \end{vmatrix} = \begin{vmatrix} \frac{-n(\bar{y}+\alpha)}{\bar{x}(\bar{x}+\bar{y}+\alpha)} & \frac{n}{\bar{x}+\bar{y}+\alpha} \\ \frac{n}{\bar{x}+\bar{y}+\alpha} & \frac{-n(\bar{x}+\alpha)}{\bar{y}(\bar{x}+\bar{y}+\alpha)} \end{vmatrix} = \frac{n^2\alpha}{\bar{x}\bar{y}(\bar{x}+\bar{y}+\alpha)} > 0 \quad \forall\ \alpha > 0.
$$

Also, since $\partial^2 \ell(\mu_1, \mu_2, \alpha)/\partial \mu_1^2 < 0$ for all $\alpha > 0$, we can say that $\bar{x}$ and $\bar{y}$ are the MLEs of μ_1 and μ_2 respectively. Furthermore, these two estimates are independent of α and hence we can maximize the profile log-likelihood obtained at these estimates to get $\hat{\alpha}$, the MLE of α.

3.4 Testing for the Treatment Effect

Parametric tests based on the likelihood can be developed for the treatment effect as in the case of PG trials.

3.4.1 Known Dispersion Parameter α

When the dispersion parameter α is known, classical results can be used to derive a UMPU test for testing $H_0 : \gamma = 1$ vs. $H_1 : \gamma \neq 1$. When α is known, the likelihood function for the BNB model given in (3.3.1) can be expressed as

$$\begin{aligned}\ell(\gamma, \mu) =& \frac{\prod_{i=1}^{n} \Gamma(x_i + y_i + \alpha)}{\prod_{i=1}^{n} x_i! \prod_{i=1}^{n} y_i!} \left[\frac{1}{\Gamma(\alpha)}\right]^n \exp\left\{n\bar{x} \log\left(\frac{\mu}{\mu + \gamma\mu + \alpha}\right)\right. \\ & \left. + n\bar{y} \log\left(\frac{\gamma\mu}{\mu + \gamma\mu + \alpha}\right) + n\alpha \log\left(\frac{\alpha}{\mu + \gamma\mu + \alpha}\right)\right\} \\ =& h(\mathbf{x}, \mathbf{y}) \exp\left\{\sum_{j=1}^{2} \eta_j(\boldsymbol{\theta}) T_j(\mathbf{x}, \mathbf{y}) - A(\boldsymbol{\theta})\right\} \end{aligned} \tag{3.4.1}$$

where

$$T_1(\mathbf{x}, \mathbf{y}) = n\bar{x} = T_1 \text{ (say)}, \; T_2(\mathbf{x}, \mathbf{y}) = n\bar{y} = T_2, \; \eta_1(\boldsymbol{\theta}) = \log\left(\frac{\mu}{\mu + \gamma\mu + \alpha}\right),$$

$$\eta_2(\boldsymbol{\theta}) = \log\left(\frac{\gamma\mu}{\mu + \gamma\mu + \alpha}\right), \; A(\boldsymbol{\theta}) = -n\alpha \log\left(\frac{\alpha}{\mu + \gamma\mu + \alpha}\right) \text{ and}$$

$$h(\mathbf{x}, \mathbf{y}) = \frac{\prod_{i=1}^{n} \Gamma(x_i + y_i + \alpha)}{\prod_{i=1}^{n} x_i! \prod_{i=1}^{n} y_i!} \left[\frac{1}{\Gamma(\alpha)}\right]^n.$$

Let $T_{12} = T_1 + T_2$. Since (T_1, T_2) belongs to a two-parameter exponential family as seen above, we know that it is a complete sufficient statistic for (η_1, η_2) and so is (T_{12}, T_2). Clearly $T_1 \sim \text{NB}(n\mu, n\alpha)$ and $T_2 \sim \text{NB}(n\gamma\mu, n\alpha)$. Also the joint distribution of

(T_1, T_2) is BNB$(n\mu, n\gamma\mu, n\alpha)$. The joint pmf of (T_2, T_{12}) can then be derived as

$$\begin{aligned} P(T_2 = t_2, T_{12} = t_{12}) =& P(T_1 = t_{12} - t_2, T_2 = t_2) \\ =& \frac{(n\alpha)^{n\alpha}}{\Gamma(n\alpha)} \frac{(n\mu)^{t_{12}-t_2}}{(t_{12}-t_2)!} \frac{(n\gamma\mu)^{t_2}}{t_2!} \frac{\Gamma(t_{12}+n\alpha)}{(n\mu+n\gamma\mu+n\alpha)^{t_{12}+n\alpha}} \\ =& \binom{t_{12}}{t_2} \frac{\Gamma(t_{12}+n\alpha)}{\Gamma(t_{12}+1)\Gamma(n\alpha)} \frac{(n\alpha)^{n\alpha}(n\mu)^{t_{12}}\gamma^{t_2}}{(n\mu+n\gamma\mu+n\alpha)^{t_{12}+n\alpha}}, \end{aligned} \tag{3.4.2}$$

where $t_2 = 0, 1, 2, \ldots, t_{12};\ t_{12} = t_2, t_2 + 1, \ldots$.

The conditional distribution of T_2 given $T_{12} = t_{12}$ is given by

$$\begin{aligned} P(T_2 = t_2 | T_{12} = t_{12}) =& \frac{P(T_2 = t_2, T_{12} = t_{12})}{P(T_{12} = t_{12})} \\ =& \frac{P(T_2 = t_2, T_{12} = t_{12})}{\sum_{k=0}^{t_{12}} P(T_2 = k, T_{12} = t_{12})} \\ =& \frac{\binom{t_{12}}{t_2}\gamma^{t_2}}{\sum_{k=0}^{t_{12}} \binom{t_{12}}{k}\gamma^k}, \quad t_2 = 0, 1, \ldots, t_{12}. \end{aligned} \tag{3.4.3}$$

This is known as a generalized power series distribution (Patil [1962]). Then, using the results in Section 4.4 of Lehmann [1986], we can obtain a UMPU test for testing $H_0 : \gamma = 1$ vs. $H_1 : \gamma \neq 1$. The critical function for this test is

$$\phi(t_2, t_{12}) = \begin{cases} 1 & \text{if } t_2 < c_1(t_{12}) \text{ or } t_2 > c_2(t_{12}) \\ \delta_1(t_{12}) & \text{if } t_2 = c_1(t_{12}) \\ \delta_2(t_{12}) & \text{if } t_2 = c_2(t_{12}) \\ 0 & \text{if } c_1(t_{12}) < t_2 < c_2(t_{12}) \end{cases} \tag{3.4.4}$$

where δ_1, δ_2, c_1 and c_2 are functions of t_{12} such that

$$E(\phi(T_2, T_{12}) | T_{12} = t_{12}) = \nu \tag{3.4.5}$$

and

$$E(T_2 \cdot \phi(T_2, T_{12}) | T_{12} = t_{12}) = \nu E(T_2 | T_{12} = t_{12}) \tag{3.4.6}$$

when $\gamma = 1$. Clearly, the distribution of T_2 given $T_{12} = t_{12}$ is symmetric about $E(T_2 | T_{12} = t_{12}) = t_{12}/2$ when $\gamma = 1$. Hence the above test can be further simplified

since any symmetric test that satisfies condition (3.4.5) must also satisfy (3.4.6). It then becomes a "symmetric" test with $\delta_1 = \delta_2$ and $c_2 = t_{12} - c_1$. Condition (3.4.5) now becomes

$$\frac{\sum_{l=0}^{c_1-1}\binom{t_{12}}{l} + \delta_1\binom{t_{12}}{c_1}}{\sum_{k=0}^{t_{12}}\binom{t_{12}}{k}} = \frac{\nu}{2}. \tag{3.4.7}$$

The p-value of the above UMPU test can be estimated using the following two methods:

$$\text{Regular } p\text{-value } = 2\frac{\sum_{l=0}^{t}\binom{t_{12}}{l}}{\sum_{k=0}^{t_{12}}\binom{t_{12}}{k}} \tag{3.4.8}$$

or

$$\text{Mid } p\text{-value } = \frac{2\sum_{l=0}^{t-1}\binom{t_{12}}{l} + \binom{t_{12}}{t}}{\sum_{k=0}^{t_{12}}\binom{t_{12}}{k}} \tag{3.4.9}$$

where $t = \min(t_1, t_2)$.

A simulation study was conducted to compare the power of the UMPU test using the regular p-value and the mid p-value methods. The power for these methods were computed as the proportion of $B = 10,000$ trials for which H_0 is rejected (p-value < 0.05). In addition to these two methods, the "Exact" power was also estimated using the definition of power, i.e., P(Reject $H_0|H_1$ is True). For each simulated data set (trial), the power of the conditional test given $T_{12} = t_{12}$ given by

$$\begin{aligned}\beta(\gamma|T_{12} = t_{12}) =& E_{H_1}\left[\phi(T_2, T_{12})|T_{12} = t_{12}\right] \\ =& \frac{\sum_{l=0}^{c_1-1}\binom{t_{12}}{t_2}\gamma^l + \sum_{l=c_2+1}^{t_{12}}\binom{t_{12}}{l}\gamma^l + \delta_1\binom{t_{12}}{c_1}\gamma^{c_1} + \delta_2\binom{t_{12}}{c_2}\gamma^{c_2}}{\sum_{k=0}^{t_{12}}\binom{t_{12}}{k}\gamma^k}.\end{aligned} \tag{3.4.10}$$

This is free of the nuisance parameters (μ, α) and thus has a known value. The average of these power values over B trials is taken to compute an estimate of the unconditional power $\beta(\gamma, \mu, \alpha)$. We call this estimate, the "Exact" power. This estimate is unbiased for the actual power and has the smallest variance among all unbiased estimates of $\beta(\gamma, \mu, \alpha)$ (see Section 4.5 of Lehmann [1986]). However, a

major disadvantage of this method of computing power is that it can only be found after the data has been observed and hence will not be suitable for computing sample sizes required to design an experiment. In such a case, the power can be estimated using the simulation method described before using the "Mid p- value" method. This method maintains its Type I error rate very well and also has higher power compared with the "Regular p- value" method even for very small sample sizes (see Figure 3.1).

3.4.2 Unknown Dispersion Parameter α

When the dispersion parameter α is unknown, the likelihood of the model in (3.3.1) does not belong to an exponential family and it is not known whether an UMPU test exists. In that case, likelihood based tests such as LRT, RST and Wald tests can be derived as in the case of PG trials. A test for the treatment effect tests $H_0 : \gamma = \gamma_0$ vs. $H_1 : \gamma \neq \gamma_0$ with $\gamma_0 = 1$. The tests are presented below:

Theorem 3.4.1. *The LRT statistic for testing $H_0 : \gamma = \gamma_0$ vs. $H_1 : \gamma \neq \gamma_0$ is given by*

$$L_{BVT} = -2(\ell(\gamma_0, \tilde{\mu}, \tilde{\alpha}) - \ell(\hat{\gamma}, \hat{\mu}, \hat{\alpha}))$$

where $\ell(\gamma_0, \tilde{\mu}, \tilde{\alpha})$ is the log-likelihood evaluated at the MLEs under H_0 (Lemma 3.3.2) and $\ell(\hat{\gamma}, \hat{\mu}, \hat{\alpha})$ is the log-likelihood evaluated at the unrestricted MLEs (Lemma 3.3.1).

Theorem 3.4.2. *The RST statistic for testing $H_0 : \gamma = \gamma_0$ vs. $H_1 : \gamma \neq \gamma_0$ for BVT trials is given by*

$$R_{BVT} = \frac{n\left[\tilde{\mu}(\bar{y} - \gamma_0\bar{x}) + \tilde{\alpha}(\bar{y} - \gamma_0\tilde{\mu})\right]^2(1+\gamma_0)}{\gamma_0(\tilde{\mu} + \gamma_0\tilde{\mu} + \tilde{\alpha})\left[\tilde{\mu}(1+\gamma_0)(\tilde{\mu}+\tilde{\alpha}) - \tilde{\alpha}\gamma_0\right]}.$$

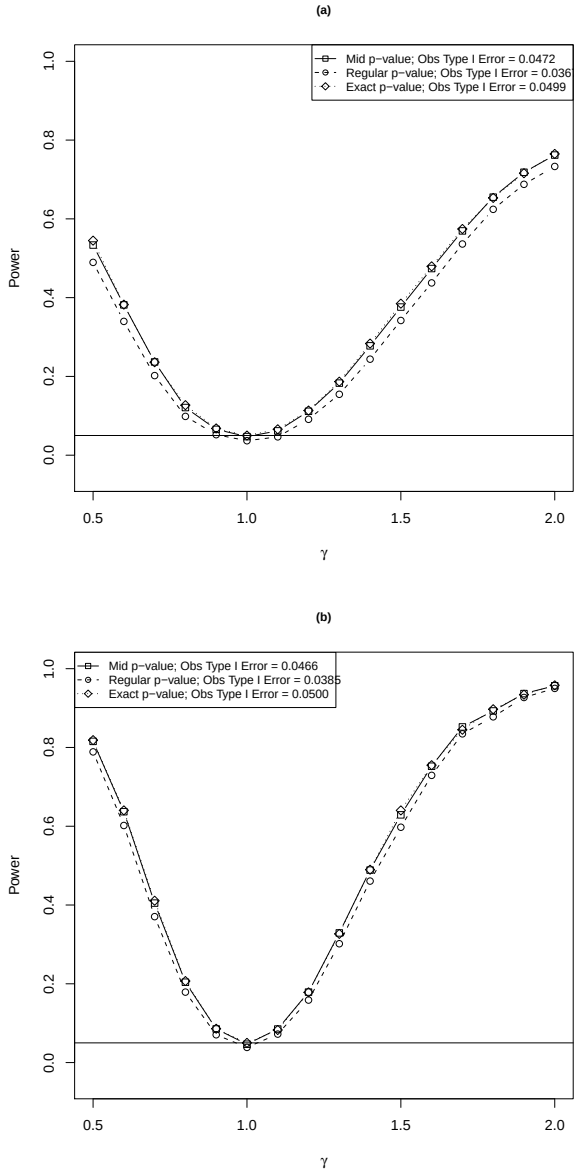

Figure 3.1: BVT Trial: Power as a function of γ for UMPU test with 5% levels for trials with 3-month baseline and 3-month treatment periods with sample sizes $(a)n = 5, (b)n = 10$ and $\mu = 5.9$, and $\alpha = 0.49$. The solid horizontal line refers to nominal level $\nu = 0.05$.

Theorem 3.4.3. *If $g(\cdot)$ is a $1-1$ function with nonzero derivative, the Wald test statistic for testing $H_0 : g(\gamma) = g(\gamma_0)$ vs. $H_1 : g(\gamma) \neq g(\gamma_0)$ for BVT trials is*

$$W_{BVT}(g(\hat{\gamma})) = \left[\frac{g\left(\bar{y}/\bar{x}\right) - g(\gamma_0)}{g'(\hat{\gamma})\sigma_{\hat{\gamma}}(\hat{\boldsymbol{\theta}})}\right]^2$$

where

$$\sigma^2_{\hat{\gamma}}(\hat{\boldsymbol{\theta}}) = \frac{\hat{\gamma}(1+\hat{\gamma})(\hat{\mu}+\hat{\gamma}\hat{\mu}+\hat{\alpha})}{n\left[\hat{\mu}(1+\hat{\gamma})(\hat{\mu}+\hat{\alpha}) - \hat{\alpha}\hat{\gamma}\right]}$$

is the asymptotic variance of $\hat{\gamma}$ evaluated at the MLEs (see Lemma 3.3.4)).

Under H_0, all the above test statistics converge in distribution to a χ^2_1 rv as $n \to \infty$ and consequently we choose c_0 as the critical value for an asymptotic 5% level test.

A simulation study was conducted to check the effect of changes in parameter values of the BNB model on the *exact* percentiles and estimated Type I error rates for LRT and WT(γ^2). The LRT can be used with asymptotic approximation safely even for very small sample sizes without compromising on Type I error rates (Figures 3.3 and 3.4). For RST and WT($\log(\gamma)$) the results were similar to the LRT (Figure not shown).

For the WT(γ^2), the *exact* percentiles can be much different than c_0 and the simulated levels higher than the nominal level for small sample sizes. The *exact* percentiles (Figure 3.5) tend to converge to c_0 for increasing values of μ (unlike for PG trials) and the convergence is quicker for larger sample sizes. Similarly, the Type I error rates seem to converge to the nominal level for increasing values of μ (Figure 3.6). Also, higher values of α tend to increase the significance level although the change is very small for α less than or equal to 2. The significance level for the WT(γ^2) can be much higher than the nominal level even for higher sample sizes if α is very high. Unless the sample size is 50 or greater and μ is 10 or greater, it may not reasonable to use the asymptotic approximation for the Wald test. We did not

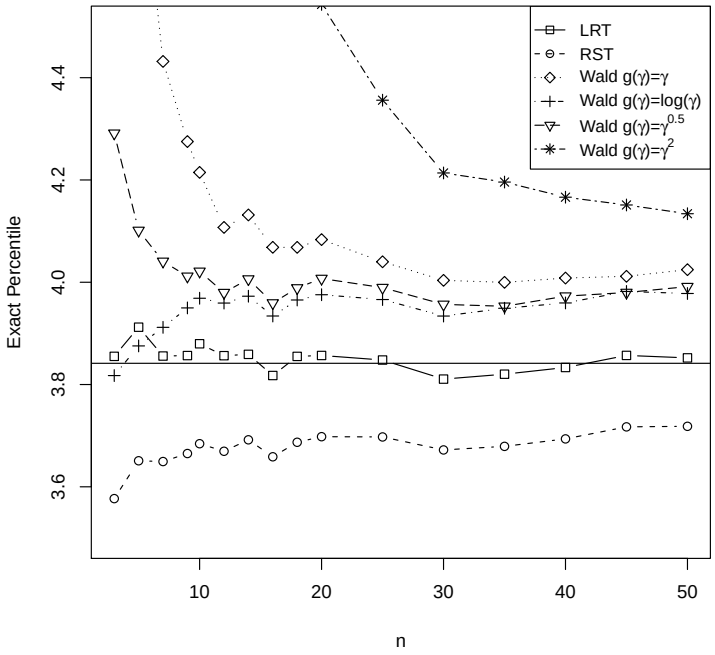

Figure 3.2: BVT Trial: Simulation based 95th percentile value for the null distribution of LRT, RST and Wald test statistics as a function of common sample size n; $\mu = 5.9$, $\alpha_1 = 0.49$. The solid horizontal line refers to $c_0 = 3.8415 (= \chi^2_1(0.95))$.

consider the WT($\sqrt{\gamma}$) and WT(γ) as they did not have the highest power on either side of H_0 as we will see in the next section.

3.5 Power Analysis

Figure 3.7 shows the power curves for the six parametric tests discussed in Section 3.4, using exact percentiles. A sample size of 10 is considered and the initial parameter estimates considered for simulation were taken to be the same as for PG trials. Here again, the power estimates for the LRT, RST and WT($\log(\gamma)$) are very

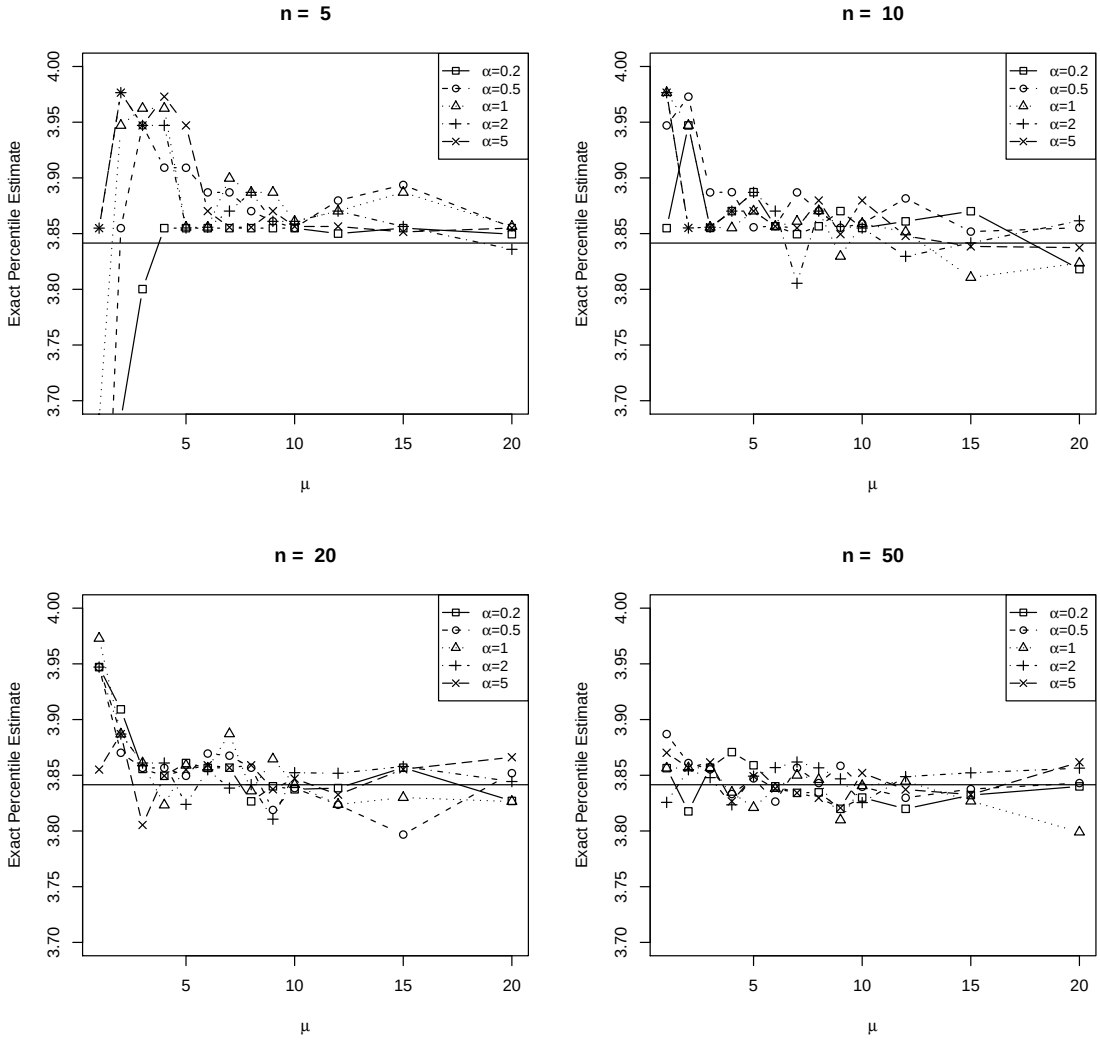

Figure 3.3: BVT Trial: Simulation based 95th percentile value for the null distribution of LRT for different values of μ and α and $n = 5, 10, 20, 50$ subjects. The solid horizontal line refers to the $c_0 = 3.8415 (= \chi^2_1(0.95))$.

similar and the tests are unbiased. The WT($\sqrt{\gamma}$) also appears to be unbiased and has a slightly higher power than the other three unbiased tests for $\gamma < 1$, but has lower power for $\gamma > 1$. A suitable unbiased test can be chosen from these four tests

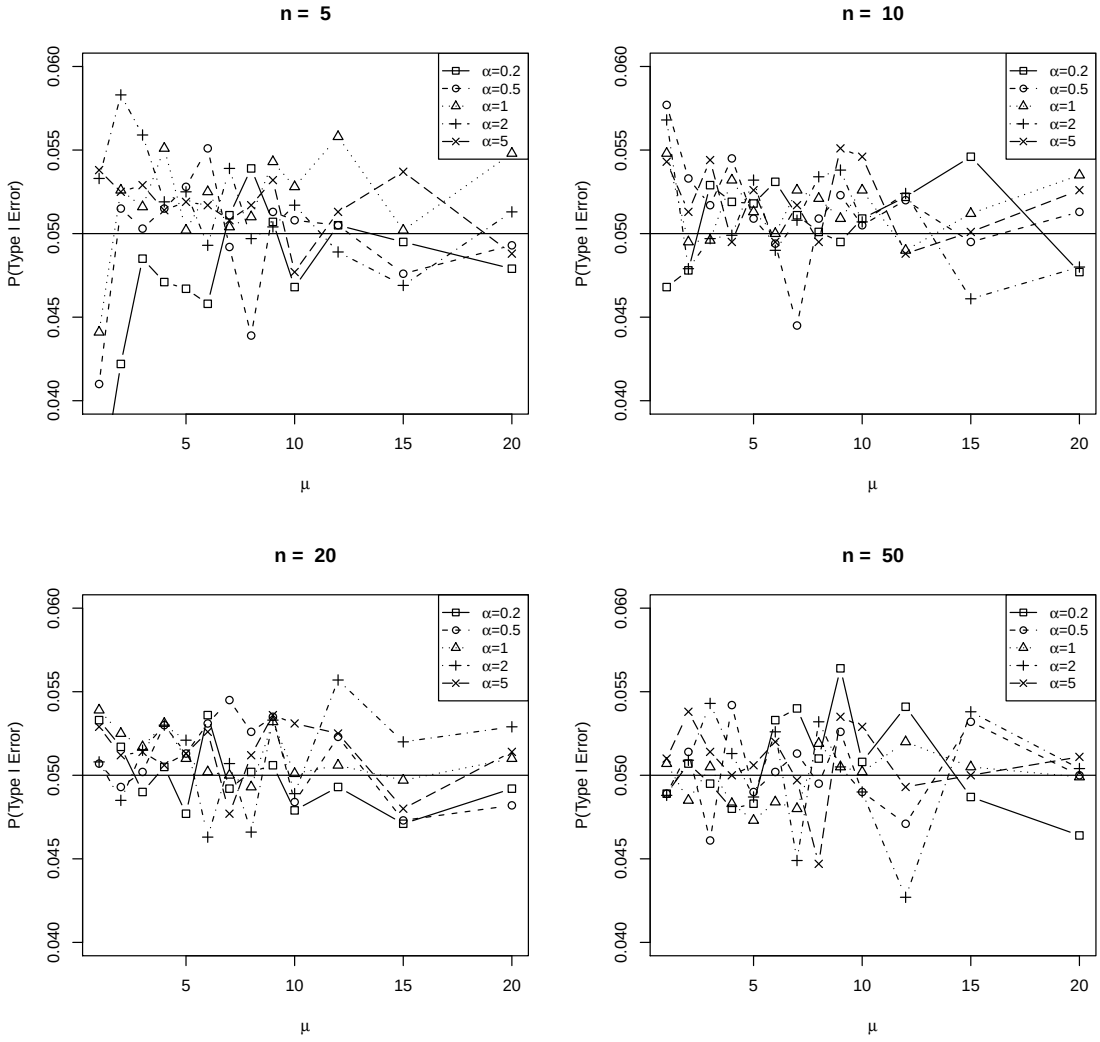

Figure 3.4: BVT Trial: Type I error rates for LRT using asymptotic approximation for different values of μ and α and $n = 5, 10, 20, 50$ subjects per group. The solid horizontal line refers to nominal level $\nu = 0.05$.

depending on the region of interest. The WT(γ^2) has the highest power for $\gamma < 1$ but is biased and has the least power for $\gamma > 1$. The WT(γ) is also biased, has a lower power than the WT (γ^2) for $\gamma < 1$ and for $\gamma > 1$ it has lower power than any of

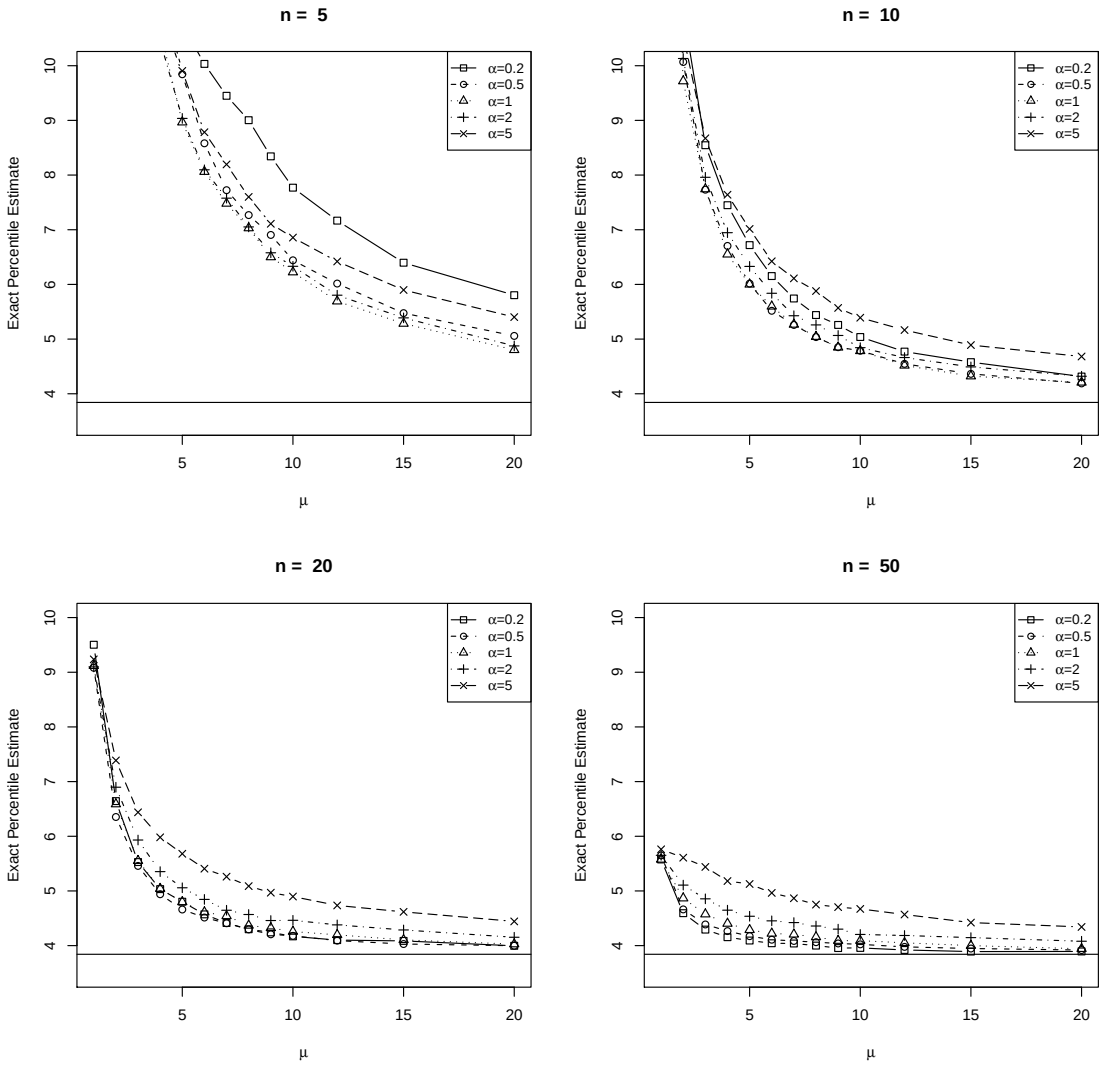

Figure 3.5: BVT Trial: Simulation based 95th percentile value for the null distribution of $\mathrm{WT}(\gamma^2)$ for different values of μ and α and $n = 5, 10, 20, 50$ subjects. The solid horizontal line refers to the $c_0 = 3.8415 (= \chi^2_1(0.95))$.

the unbiased tests. Thus, for RRMS clinical trials where the interest is on the region $\gamma < 1$, the $\mathrm{WT}(\gamma^2)$ is perhaps the best choice.

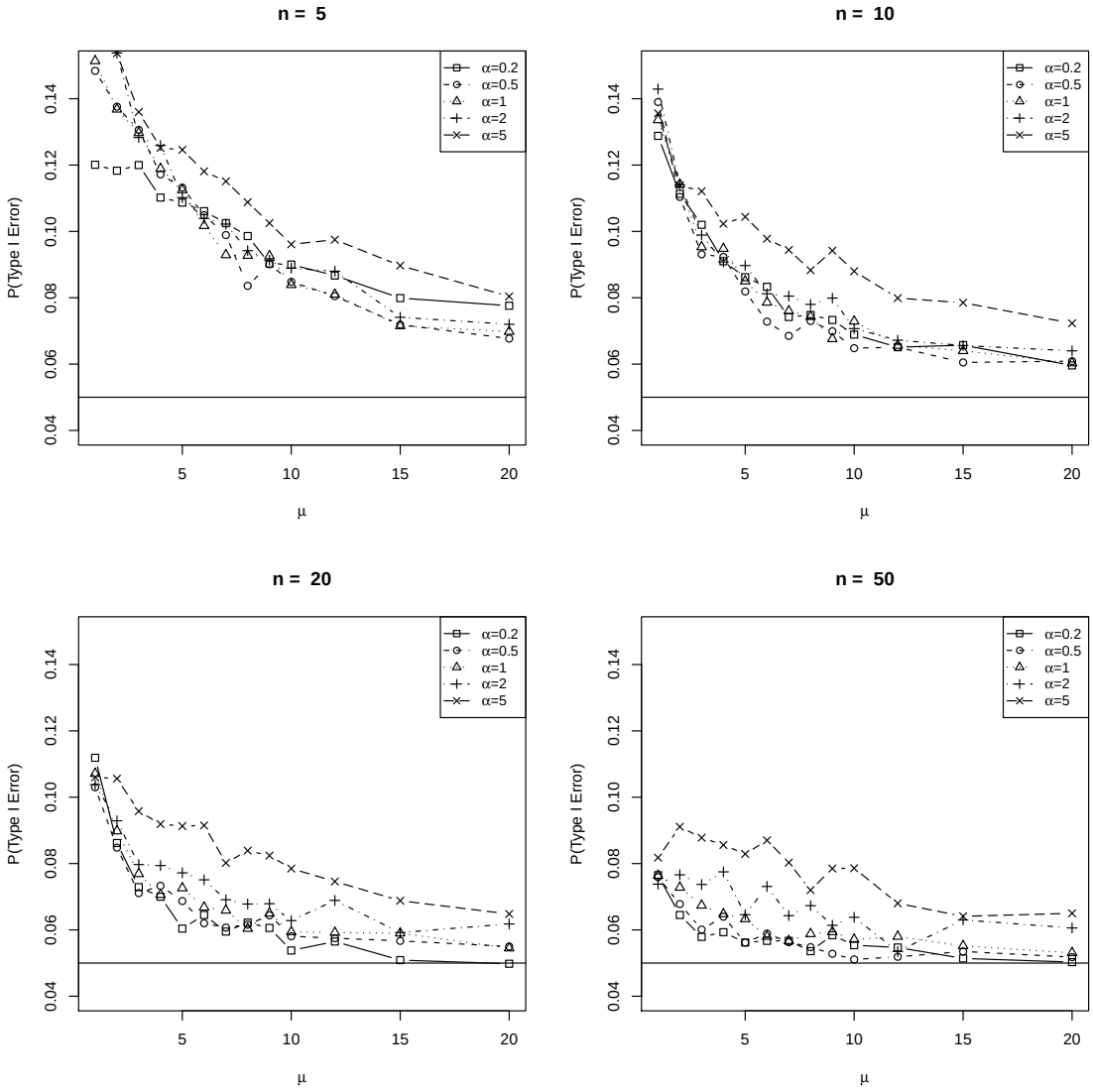

Figure 3.6: BVT Trial: Type I error rates for $\mathrm{WT}(\gamma^2)$ using asymptotic approximation for different values of μ and α and $n = 5, 10, 20, 50$ subjects per group. The solid horizontal line refers to nominal level $\nu = 0.05$.

A comparison of the power estimates when we use asymptotic and exact percentiles for LRT, RST and $\mathrm{WT}(\gamma^2)$ is presented in Figure 3.8. As apparent from the

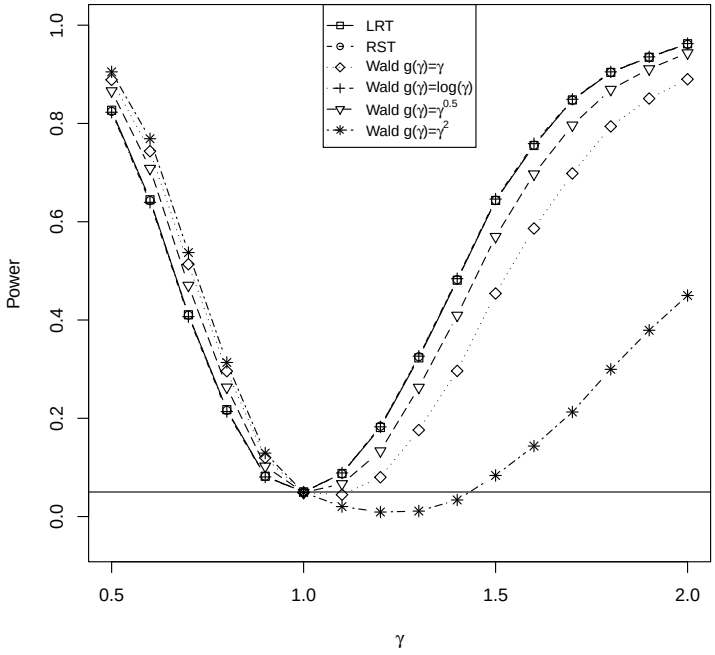

Figure 3.7: BVT Trial: Power curve as a function of γ for six tests with exact 5% levels for 3-month baseline vs 3-month treatment trials with sample size $n = 10$, $\mu = 5.9$, and $\alpha = 0.49$. The solid horizontal line refers to nominal level $\nu = 0.05$.

exact 5% critical values given in Figure 3.2, the power estimates for the exact and asymptotic tests are very similar for LRT, and the exact test produces slightly higher power for RST. In contrast, the power curves for the exact and asymptotic WT(γ^2) differ substantially and the asymptotic test has an inflated level of significance. In conclusion, we suggest the use of c_0 as the critical value for the LRT, and it works reasonably well for RST and Wald test for $\log(\gamma)$, for samples of size 10 or higher. We suggest using the exact critical value for Wald test for γ^2.

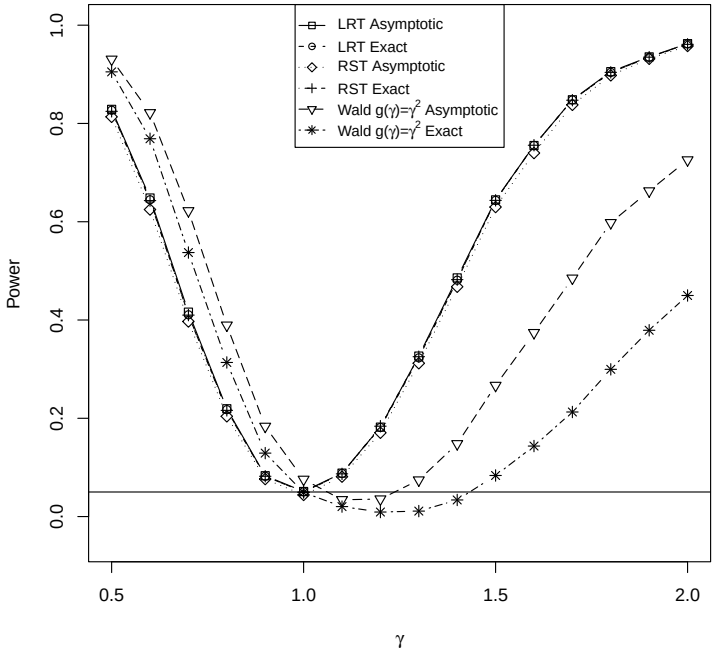

Figure 3.8: BVT Trial: Power comparisons for the exact and asymptotic 5% level LRT, RST and Wald test for γ^2 for 3-month baseline vs 3-month treatment trials with sample size $n = 10$, $\mu = 5.9$, and $\alpha = 0.49$. The solid horizontal line refers to nominal level $\nu = 0.05$.

3.6 Sample Size Estimation

Table 3.1 gives the sample sizes required to achieve 80% and 90% power for BVT trials obtained when we use the LRT, WT(γ^2), and Wilcoxon Signed Rank (WSR) test. The WSR sample sizes were reported in Sormani et al. [2001b]. We use c_0 as the critical value for LRT and simulated percentile for $WT(\gamma^2)$. Sample sizes are obtained for 20-50% reduction in the lesion rate during a follow-up period of 3 and 6 months and for a power of 80% and 90%. It can be seen that the sample sizes for

the LRT are around 25-60% smaller than those for the WSR test. As the treatment effect increases from 20% to 50%, the percentage reduction in sample sizes observed seems to increase from around 25% to 60%. Samples sizes for the Wald test are 10-30% smaller than those of the LRT and 35-65% smaller than the nonparametric sample sizes. Sample sizes for 90% power are 30-40% (35-50%) greater for LRT (Wald test) than those for 80% power. In contrast to the PG trials, doubling the period duration will reduce the sample size by almost half. This can be anticipated from the properties of Poisson process and the fact that the comparison takes place within a subject. Increasing μ from 5.9 to 13 (by a factor of 2.2) reduces the sample size required by almost 50%. This is true since, conditional on the subject the lesion counts every month are independent and following up a patient for twice as long results in twice the amount of 'information' gained regarding the parameters there by resulting in a significant reduction.

3.7 Discussion

In this chapter, we looked at several parametric tests for the treatment effect in BVT trials assuming a BNB model for the data. A UMPU test was proposed in Section 3.4.1 when the dispersion parameter α is assumed known. It was shown that the "Mid p- value" method did best according to having a higher power and also maintaining Type I error rate very well even for small sample sizes. When α is assumed unknown, several likelihood based tests such as LRT, RST and Wald tests were compared using both the asymptotic and exact percentiles. When an unbiased test is preferred, any of the three asymptotic tests, LRT, RST and WT($\log(\gamma)$) can be used. They perform equally well in terms of power and also maintain Type I error

Table 3.1: BVT Trial: Numbers of patients needed to achieve 80% and 90% power for $100(1-\gamma)\%$ treatment effect and follow-up period (t) of 3 and 6 months, and initial estimates $(\mu, \alpha) = (5.9, 0.49)$ for $t = 3$ and $(13, 0.52)$ for $t = 6$; level $\nu = 0.05$.

$1-\gamma$	t	80% power; 5% level			
		Test			Critical Value
		WSR	LRT	WT(γ^2)	WT(γ^2)
			Asymptotic/Exact	Exact	Exact
0.20	3	80	61	51	4.155
	6	48	29	25	4.068
0.30	3	42	27	21	4.428
	6	22	13	10	4.467
0.40	3	26	15	11	5.420
	6	14	8	6	5.253
0.50	3	18	10	7	6.769
	6	12	5	4	6.703
$1-\gamma$	t	90% power; 5% level			
		Test			Critical Value
		WSR	LRT	WT(γ^2)	WT(γ^2)
			Asymptotic/Exact	Exact	Exact
0.20	3	108	84	70	4.051
	6	60	39	34	4.013
0.30	3	50	35	31	4.260
	6	32	18	15	4.195
0.40	3	35	21	16	4.721
	6	20	11	8	4.699
0.50	3	25	13	10	5.621
	6	17	7	6	5.278

rates even for smaller sample sizes. So an *exact* percentile based test is not necessary in these cases. The sample sizes obtained using the LRT are 25-50% smaller than the corresponding sample sizes using WSR test.

When an unbiased test is not important, the WT(γ^2) has the highest power for $\gamma < 1$ and has the least power for $\gamma > 1$. With this test one needs to use the simulated percentiles to maintain the nominal levels. The *exact* WT(γ^2) reduces the sample sizes 10-30% when compared to the LRT based sample sizes. As discussed, this test is highly appropriate for clinical trials in MS as one is only interested in the region $\gamma < 1$ and needs to find a test that has the highest power in that region.

There are other examples for which the BNB distribution was shown to fit well. Appendix Section C.1.1 gives MRI lesion count data on 23 RRMS and SPMS patients. The number of new active lesions seen during 6 monthly follow-up scans from 31 RRMS patients is given in Appendix Section C.1.2. The number of epileptic seizures observed in 28 patients from the placebo group is given in Appendix Section C.2. In all these cases the BNB distribution has been shown to fit well. The methods developed in this chapter can also be applied to all areas where the BNB distribution provides a good fit and not just for clinical trials involving RRMS patients.

CHAPTER 4

Bivariate Negative Binomial Models for Parallel Group with Baseline Correction Trials

4.1 Introduction

Parallel group trials with a baseline correction (PGB) is a very important trial design in clinical trials involving count data. This is also a common design in multiple sclerosis clinical trials. The baseline correction scan is obtained to counter any significant differences among the two arms of the trial even before the treatment is initiated. Denoting the number of new enhancing lesions seen in the baseline scan as X and the number of new enhancing lesions seen in the follow-up period as Y, we propose in Section 4.2 a joint bivariate NB (BNB) distribution for modelling (X, Y) coming from RRMS PGB trials. This extends the univariate NB model seen in Chapter 2. In Sections 4.3 and 4.4, MLEs of the parameters are obtained and parametric tests are constructed to test for the treatment effect using the joint distribution of (X, Y). The power of these tests are compared in Section 4.5 and sample size estimates using the "best" tests are given in Section 4.6. In Section 4.7, we derive the distribution of $Y - X$ and show why this distribution is not suited for comparison between two groups. We finally conclude this chapter with some discussion in Section 4.8.

4.2 The Model

Let $(X_i, Y_i),\ i = 1, 2$ be the total number of *new* enhancing lesions seen during the baseline and follow-up period of a patient in the placebo and treatment groups respectively. Let $Z_i,\ i = 1, 2$ denote the random subject effects assumed to be iid Gamma(α, α^{-1}) rvs. We assume that

$$X_1|Z_1 = z_1 \sim \text{Poisson}(\mu z_1) \text{ independent of } Y_1|Z_1 = z_1 \sim \text{Poisson}(t\mu z_1) \text{ and}$$

$$X_2|Z_2 = z_2 \sim \text{Poisson}(\mu z_2) \text{ independent of } Y_2|Z_2 = z_2 \sim \text{Poisson}(t\gamma\mu z_2).$$

The joint marginal distribution of (X_1, Y_1) can be easily derived as follows:

$$\begin{aligned}
P_{X_1,Y_1}(x_1, y_1) =& \int_0^\infty P(X_1|Z_1 = z_1) \times P(Y_1|Z_1 = z_1) \times f_{Z_1}(z_1)dz_1 \\
=& \int_0^\infty \frac{e^{-\mu z_1}(-\mu z_1)^{x_1}}{x_1!} \times \frac{e^{-t\mu z_1}(t\mu z_1)^{y_1}}{y_1!} \times \frac{\alpha^\alpha}{\Gamma(\alpha)} z_1^{\alpha-1} e^{-\alpha z} dz_1 \\
=& \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\mu^{x_1}}{x_1!} \frac{(t\mu)^{y_1}}{y_1!} \int_0^\infty z_1^{(x_1+y_1+\alpha)-1} e^{-(\mu+t\mu+\alpha)z_1} dz_1 \\
=& \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\mu^{x_1}}{x_1!} \frac{(t\mu)^{y_1}}{y_1!} \frac{\Gamma(x_1 + y_1 + \alpha)}{(\mu + t\mu + \alpha)^{x_1+y_1+\alpha}} \\
& \times \int_0^\infty \frac{(\mu + t\mu + \alpha)^{x_1+y_1+\alpha}}{\Gamma(x_1 + y_1 + \alpha)} z_1^{(x_1+y_1+\alpha)-1} e^{-(\mu+t\mu+\alpha)z_1} dz_1 \\
=& \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\mu^{x_1}}{x_1!} \frac{(t\mu)^{y_1}}{y_1!} \frac{\Gamma(x_1 + y_1 + \alpha)}{(\mu + t\mu + \alpha)^{x_1+y_1+\alpha}}, \quad x_1, y_1 = 0, 1, 2 \ldots;\ \mu, \alpha > 0.
\end{aligned} \tag{4.2.1}$$

Thus (X_1, Y_1) follows a BNB distribution with parameters μ and α and we write (X_1, Y_1) is BNB$(\mu, t\mu, \alpha)$. It can be shown similarly that (X_2, Y_2) is BNB$(\mu, t\gamma\mu, \alpha)$ with pmf

$$P_{X_2,Y_2}(x_2, y_2) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\mu^{x_2}}{x_2!} \frac{(t\gamma\mu)^{y_2}}{y_2!} \frac{\Gamma(x_2 + y_2 + \alpha)}{(\mu + t\gamma\mu + \alpha)^{x_2+y_2+\alpha}}, \quad x_2, y_2 = 0, 1, 2 \ldots;$$
$$\gamma, \mu, \alpha > 0. \tag{4.2.2}$$

4.3 Estimation

In this section we obtain the MLEs of the BNB model parameters for the PGB design. Let there be n_1 observations in the placebo group and n_2 observations in the treatment group. If the observed data is given by $(\mathbf{x}, \mathbf{y}) = (\mathbf{x_1}, \mathbf{y_1}), (\mathbf{x_2}, \mathbf{y_2})$ where $(x_i, y_i) = (x_{i1}, y_{i1}), (x_{i2}, y_{i2}), \ldots, (x_{in_i}, y_{in_i}),\ i = 1, 2$ then the joint likelihood function for the above PGB model is

$$\begin{aligned}
L(\gamma, \mu, \alpha|(\mathbf{x}, \mathbf{y})) &= \prod_{i=1}^{n_1} P_{X_1,Y_1}(x_{1i}, y_{1i}) \times \prod_{j=1}^{n_2} P_{X_2,Y_2}(x_{2j}, y_{2j}) \\
&= \prod_{i=1}^{n_1} \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\mu^{x_{1i}}}{x_{1i}!} \frac{(t\mu)^{y_{1i}}}{y_{1i}!} \frac{\Gamma(x_{1i} + y_{1i} + \alpha)}{(\mu + t\mu + \alpha)^{x_{1i}+y_{1i}+\alpha}} \\
&\quad \times \prod_{j=1}^{n_2} \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\mu^{x_{2j}}}{x_{2j}!} \frac{(t\gamma\mu)^{y_{2j}}}{y_{2j}!} \frac{\Gamma(x_{2j} + y_{2j} + \alpha)}{(\mu + t\gamma\mu + \alpha)^{x_{2i}+y_{2i}+\alpha}} \\
&= \left\{ \frac{\alpha^\alpha}{\Gamma(\alpha)} \right\}^{n_1+n_2} \frac{\mu^{n_1(\bar{x}_1+\bar{y}_1)} t^{n_1\bar{y}_1}}{(\mu + t\mu + \alpha)^{n_1(\bar{x}_1+\bar{y}_1+\alpha)}} \frac{\prod_{i=1}^{n_1} \Gamma(x_{1i} + y_{1i} + \alpha)}{\prod_{i=1}^{n_1} x_{1i}! y_{1i}!} \\
&\quad \times \frac{\mu^{n_2(\bar{x}_2+\bar{y}_2)} (\gamma t)^{n_2\bar{y}_2}}{(\mu + t\gamma\mu + \alpha)^{n_2(\bar{x}_2+\bar{y}_2+\alpha)}} \frac{\prod_{j=1}^{n_2} \Gamma(x_{2j} + y_{2j} + \alpha)}{\prod_{j=1}^{n_2} x_{2j}! y_{2j}!}. \qquad (4.3.1)
\end{aligned}$$

The log-likelihood function is

$$\begin{aligned}
\ell(\gamma, \mu, \alpha) &= (n_1 + n_2) \{\alpha \log(\alpha) - \log \Gamma(\alpha)\} \\
&\quad + n_1(\bar{x}_1 + \bar{y}_1) \log \mu + n_1\bar{y}_1 \log t - n_1(\bar{x}_1 + \bar{y}_1 + \alpha) \log(\mu + t\mu + \alpha) \\
&\quad + n_2(\bar{x}_2 + \bar{y}_2) \log \mu + n_2\bar{y}_2 \log(t\gamma) - n_2(\bar{x}_2 + \bar{y}_2 + \alpha) \log(\mu + t\gamma\mu + \alpha) \\
&\quad + \sum \log \Gamma(x_{1i} + y_{1i} + \alpha) - \sum \log(x_{1i}!) - \sum \log(y_{1i}!) \\
&\quad + \sum \log \Gamma(x_{2j} + y_{2j} + \alpha) - \sum \log(x_{2j}!) - \sum \log(y_{2j}!). \qquad (4.3.2)
\end{aligned}$$

The components of the score vector for the log-likelihood function given in (4.3.2) are

$$\frac{\partial \ell(\gamma,\mu,\alpha)}{\partial \gamma} = \frac{n_2\bar{y}_2}{\gamma} - \frac{n_2(\bar{x}_2+\bar{y}_2+\alpha)}{\mu+t\gamma\mu+\alpha}(t\mu), \tag{4.3.3}$$

$$\begin{aligned}\frac{\partial \ell(\gamma,\mu,\alpha)}{\partial \mu} =& \frac{n_1(\bar{x}_1+\bar{y}_1)+n_2(\bar{x}_2+\bar{y}_2)}{\mu} - \frac{n_1(\bar{x}_1+\bar{y}_1+\alpha)}{\mu+t\mu+\alpha}(1+t) \\ & - \frac{n_2(\bar{x}_2+\bar{y}_2+\alpha)}{\mu+t\gamma\mu+\alpha}(1+t\gamma),\end{aligned} \tag{4.3.4}$$

$$\begin{aligned}\frac{\partial \ell(\gamma,\mu,\alpha)}{\partial \alpha} =& (n_1+n_2)\left[1+\log\alpha-\psi(\alpha)\right] - n_1\left[\frac{\bar{x}_1+\bar{y}_1+\alpha}{\mu+t\mu+\alpha}+\log(\mu+t\mu+\alpha)\right] \\ & + \sum_{i=1}^{n_1}\psi(x_{1i}+y_{1i}+\alpha) - n_2\left[\frac{\bar{x}_2+\bar{y}_2+\alpha}{\mu+t\gamma\mu+\alpha}+\log(\mu+t\gamma\mu+\alpha)\right] \\ & + \sum_{j=1}^{n_2}\psi(x_{2j}+y_{2j}+\alpha).\end{aligned} \tag{4.3.5}$$

Lemma 4.3.1. *The MLEs of μ, γ and α do not have a closed form expression and they need to be obtained by direct numerical maximization of the log-likelihood or by equating the score vector to zero and solving for the parameters.*

When $\gamma = \gamma_0$ is known the MLEs of μ and α can be obtained by equating (4.3.4) and (4.3.5) to 0 and solving for the parameters. For a general γ_0 no close form solutions for the parameters are available and numerical methods need to be employed to obtain the MLEs.

Lemma 4.3.2. *When $\gamma_0 = 1$, the MLE of μ, denoted by $\tilde{\mu}$, reduces to*

$$\tilde{\mu} = \frac{n_1(\bar{x}_1+\bar{y}_1)+n_2(\bar{x}_2+\bar{y}_2)}{(n_1+n_2)(1+t)}.$$

The MLE of α, $\tilde{\alpha}$, solves (4.3.5) set to 0 at $\gamma = 1$ and $\mu = \tilde{\mu}$. One can also numerically maximize $\ell(1,\tilde{\mu},\alpha)$ with respect to α to obtain $\tilde{\alpha}$.

The second order derivatives of the log-likelihood function in (4.3.2) are

$$
\begin{aligned}
\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \gamma^2} =& -\frac{n_2\bar{y}_2}{\gamma^2} + \frac{n_2(\bar{x}_2+\bar{y}_2+\alpha)(t\mu)^2}{(\mu+t\gamma\mu+\alpha)^2},\\
\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \gamma \partial \mu} =& -\frac{n_2(\bar{x}_2+\bar{y}_2+\alpha)(t\alpha)}{(\mu+t\gamma\mu+\alpha)^2} = \frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \mu \partial \gamma},\\
\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \gamma \partial \alpha} =& -n_2 t\mu \left\{ \frac{\mu+t\gamma\mu-\bar{x}_2-\bar{y}_2}{(\mu+t\gamma\mu+\alpha)^2} \right\} = \frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \alpha \partial \gamma},\\
\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \mu^2} =& \frac{n_1(\bar{x}_1+\bar{y}_1+\alpha)(1+t)^2}{(\mu+t\mu+\alpha)^2} + \frac{n_2(\bar{x}_2+\bar{y}_2+\alpha)(1+t\gamma)^2}{(\mu+t\gamma\mu+\alpha)^2}\\
& - \frac{n_1(\bar{x}_1+\bar{y}_1)+n_2(\bar{x}_2+\bar{y}_2)}{\mu^2},\\
\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \mu \partial \alpha} =& -\frac{n_1(1+t)[(\mu+t\mu)-(\bar{x}_1+\bar{y}_1)]}{(\mu+t\mu+\alpha)^2} - \frac{n_2(1+t\gamma)[(\mu+t\gamma\mu)-(\bar{x}_2+\bar{y}_2)]}{(\mu+t\gamma\mu+\alpha)^2}\\
=& \frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \alpha \partial \mu},\\
\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \alpha^2} =& \frac{n_1(\mu+t\mu)}{\alpha(\mu+t\mu+\alpha)} + \frac{n_2(\mu+t\gamma\mu)}{\alpha(\mu+t\gamma\mu+\alpha)} - (n_1+n_2)\psi'(\alpha)\\
& - \frac{n_1[(\mu+t\mu)-(\bar{x}_1+\bar{y}_1)]}{(\mu+t\mu+\alpha)^2} - \frac{n_2[(\mu+t\gamma\mu)-(\bar{x}_2+\bar{y}_2)]}{(\mu+t\gamma\mu+\alpha)^2}\\
& + \sum_{i=1}^{n_1} \psi'(x_{1i}+y_{1i}+\alpha) + \sum_{j=1}^{n_2} \psi'(x_{2j}+y_{2j}+\alpha).
\end{aligned}
\tag{4.3.6}
$$

Lemma 4.3.3. *The elements of the FIM* $\mathbf{I}(\boldsymbol{\theta})$*, where* $\big(I(\boldsymbol{\theta})\big)_{i,j} = -E\left\{\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j}\right\}$ *and parameter vector* $\boldsymbol{\theta} = (\gamma,\mu,\alpha)$ *are given below:*

$$
\begin{aligned}
I_{11}(\boldsymbol{\theta}) =& -E\left\{\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \gamma^2}\right\} = \frac{n_2 t\mu(\mu+\alpha)}{\gamma(\mu+t\gamma\mu+\alpha)},\\
I_{22}(\boldsymbol{\theta}) =& -E\left\{\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \mu^2}\right\} = \frac{n_1(1+t)+n_2(1+t\gamma)}{\mu} - \frac{n_1(1+t)^2}{\mu+t\mu+\alpha} - \frac{n_2(1+t\gamma)^2}{\mu+t\gamma\mu+\alpha},\\
I_{33}(\boldsymbol{\theta}) =& -E\left\{\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \alpha^2}\right\} = -\frac{n_1\mu(1+t)}{\alpha(\mu+t\mu+\alpha)} - \frac{n_2\mu(1+t\gamma)}{\alpha(\mu+t\gamma\mu+\alpha)} + (n_1+n_2)\psi'(\alpha)\\
& - \sum_{i=1}^{n_1} E\left[\psi'(x_{1i}+y_{1i}+\alpha)\right] - \sum_{j=1}^{n_2} E\left[\psi'(x_{2i}+y_{2i}+\alpha)\right],\\
I_{12}(\boldsymbol{\theta}) =& -E\left\{\frac{\partial^2 \ell(\gamma,\mu,\alpha)}{\partial \gamma \partial \mu}\right\} = \frac{n_2 t\alpha}{\gamma(\mu+t\gamma\mu+\alpha)} = I_{21}(\boldsymbol{\theta}), \quad \textit{and}\\
I_{ij}(\boldsymbol{\theta}) =& \ 0, \quad \textit{otherwise.}
\end{aligned}
$$

Lemma 4.3.4. *The asymptotic variance of $\hat{\gamma}$ is the first element of the inverse of the FIM given in Lemma 4.3.3. We have,*

$$\begin{aligned}\sigma^2_{\hat{\gamma}}(\boldsymbol{\theta}) =& I^{-1}_{1.2} = \left[I_{11} - I_{12}I^{-1}_{22}I_{21}\right]^{-1} \\ =& \frac{(\mu + t\gamma\mu + \alpha)}{n_2 t\mu} \\ & \times \left\{\frac{\mu+\alpha}{\gamma} - \frac{n_2 t\alpha(\mu + t\mu + \alpha)}{n_1(\mu + t\gamma\mu + \alpha)(1+t) + n_2(\mu + t\mu + \alpha)(1+t\gamma)}\right\}^{-1}.\end{aligned}$$

4.4 Testing for the Treatment Effect

In this section we propose likelihood based parametric tests (LRT, RST and WT) to test for a general $H_0 : \gamma = \gamma_0$ vs. $H_1 : \gamma \neq \gamma_0$. For RRMS clinical trials a test for no treatment effect would test $H_0 : \gamma = 1$ vs. $H_1 : \gamma \neq 1$.

Theorem 4.4.1. *(LRT). The LRT statistic to test $H_0 : \gamma = \gamma_0$ vs. $H_1 : \gamma \neq \gamma_0$ is*

$$LRT_{PGB} = -2(\ell(\gamma_0, \tilde{\mu}, \tilde{\alpha}) - \ell(\hat{\gamma}, \hat{\mu}, \hat{\alpha})),$$

where $\tilde{\mu}, \tilde{\alpha}$ are the MLEs under Θ_0 given in Lemma 4.3.2 and $\hat{\gamma}, \hat{\mu}, \hat{\alpha}$ are the MLEs under Θ given in Lemma 4.3.1.

Theorem 4.4.2. *(RST). The RST statistic to test $H_0 : \gamma = \gamma_0$ vs. $H_1 : \gamma \neq \gamma_0$ is*

$$\begin{aligned}RST_{PGB} =& \left[\frac{\partial \ell(\gamma, \mu, \alpha)}{\partial \gamma}\right]^2_{\theta=\tilde{\theta}} \times \sigma^2_{\hat{\gamma}}(\tilde{\theta}) \\ =& \left[\frac{n_2\bar{y}_2}{\gamma_0} - \frac{n_2(\bar{x}_2 + \bar{y}_2 + \tilde{\alpha})(t\tilde{\mu})}{\tilde{\mu} + t\gamma_0\tilde{\mu} + \tilde{\alpha}}\right]^2 \left[\frac{\tilde{\mu} + t\gamma_0\tilde{\mu} + \tilde{\alpha}}{n_2 t\tilde{\mu}}\right] \\ & \times \left\{\frac{\tilde{\mu}+\tilde{\alpha}}{\gamma_0} - \frac{n_2 t\tilde{\alpha}(\tilde{\mu} + t\tilde{\mu} + \tilde{\alpha})}{n_1(\tilde{\mu} + t\gamma_0\tilde{\mu} + \tilde{\alpha})(1+t) + n_2(\tilde{\mu} + t\tilde{\mu} + \tilde{\alpha})(1+t\gamma_0)}\right\}^{-1}.\end{aligned}$$

The MLEs $\tilde{\mu}$ and $\tilde{\alpha}$ are given in Lemma 4.3.2.

Theorem 4.4.3. *(WT) The WT statistic for testing $H_0 : g(\gamma) = g(\gamma_0)$ vs. $H_1 : g(\gamma) \neq g(\gamma_0)$ is given by*

$$WT_{PGB} = \left[\frac{g(\hat{\gamma}) - g(\gamma_0)}{g'(\hat{\gamma})\sigma_{\hat{\gamma}}(\hat{\boldsymbol{\theta}})}\right]^2$$

where $\hat{\gamma}$ is MLE of γ under Θ (Lemma 4.3.1) and $\sigma^2_{\hat{\gamma}}(\hat{\boldsymbol{\theta}})$ is the asymptotic variance of $\hat{\gamma}$ given in Lemma 4.3.4.

We consider the following functions $g(\gamma) : \gamma, \log(\gamma), \sqrt{\gamma}$ and γ^2. Each of the above statistic is asymptotically distributed as χ^2_1 rv under H_0 and an approximate level ν test rejects H_0 if the test statistic is $> \chi^2_1(1-\nu)$. This 5% critical value of 3.8415 is denoted, as before, by c_0.

A simulation study was done to evaluate the effect of changing n, μ and α on the *exact* 95$^{\text{th}}$ percentiles and actual Type I error levels when one uses c_0 as the critical value. Figures 4.1 and 4.2 give the simulated *exact* percentiles for asymptotic LRT and for the WT(γ^2). The empirical Type I error rates for these two tests are given in Figures 4.3 and 4.4. These were obtained for $n = 5, 10, 20, 50$ per group, $t\mu$ ranging from 1 to 20 and for α =0.20, 0.50, 1, 2 and 5.

For the LRT and WT(γ^2), the *exact* percentiles (Type I error rates) converge to c_0 (0.05) as $t\mu$ increases and the convergence is much quicker for higher sample sizes. Increasing α results in higher *exact* percentiles and Type I error rates for the LRT but these values reduced for WT(γ^2). For the LRT, the Type I error rates are well under 0.06 for sample sizes 20 or higher for all values of μ and α. Even for a sample size of 10 subjects per group, the error rates for the asymptotic LRT are close to 0.05 unless μ is very small or α is very large. Hence the asymptotic LRT can be used for PGB trials as long as the sample size is reasonably large for most values of μ and α. For the asymptotic WT(γ^2) the Type I error rates are much higher than the nominal level for most values of μ and α even for higher sample sizes. Hence for the WT(γ^2) the asymptotic approximation is not suggested unless the sample size is 50 or above, $t\mu \geq 10$ and $\alpha \geq 2$.

For the RST, the *exact* percentile estimates converge to c_0 as n increases and also as α increases for a fixed n (Figure not shown). They are relatively stable with respect to μ. The Type I error rates for the asymptotic RST are very close to the nominal level of 0.05 for all values of $n >= 10$, μ and α. Thus, unless the sample size is very small (< 10 subjects per group) and $\mu <= 5$ and $\alpha = 0.2$ it is very safe to use the asymptotic RST without compromising on Type I error rates. We did not consider the other tests since they were not the most powerful on either side of the null hypothesis as we will see in the next section.

4.5 Power Analysis

Sormani et al. [2001b] used Monte-carlo simulations assuming the BNB model but use nonparametric tests to obtain sample sizes for RRMS PGB trials. Here we use the BNB assumption for simulations and use the associated parametric tests described in Section 4.4 to obtain sample size estimates. We use the initial parameter estimates suggested by Sormani as done in the case of PG and BVT trials seen in Chapters 2 and 3 respectively. For the BNB model given in Section 4.2, μ represents the new enhancing lesions seen during one month (or on one scan) and hence we take $\mu = 5.9/3$ for $t = 3$ and $\mu = 13/6$ for $t = 6$ as our initial parameter estimates for our simulations. We generated $10,000$ trials (datasets) assuming the BNB model in Section 4.2 for several values of γ ranging from 0.50 to 2 and assuming $n_1 = n_2 = 10$ subjects for each group. The power of each test was estimated as the proportion of these 10,000 trials for which the null hypothesis is rejected.

The *exact* percentiles for the six tests are much higher than c_0 for small sample sizes (see Figure 4.5). Therefore using asymptotic approximations for WT(γ^2),

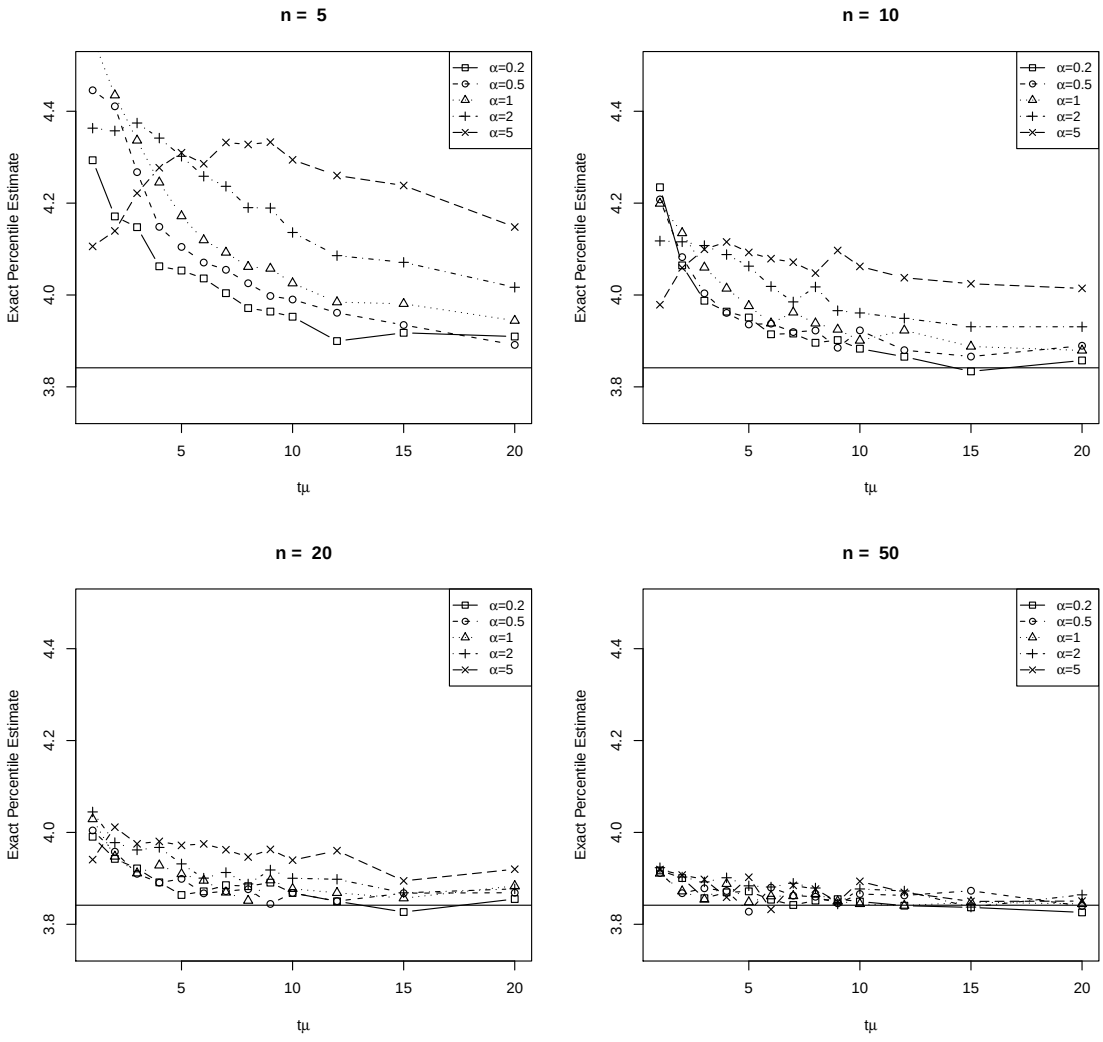

Figure 4.1: PGB Trial: Simulation based 95th percentile value for the null distribution of LRT statistic for different values of $t\mu$ and α and $n = 5, 10, 20, 50$ subjects per group; The solid horizontal line refers to the $c_0(= 3.8415 = \chi^2_1(0.95))$.

WT(γ) and WT(log(γ)) results in inflated error levels (Figure 4.6). Even though WT(γ^2) has the highest power for $\gamma < 1$, it also has the highest Type I error rate. Hence it is not advisable to use asymptotic approximations especially for WT(γ^2)

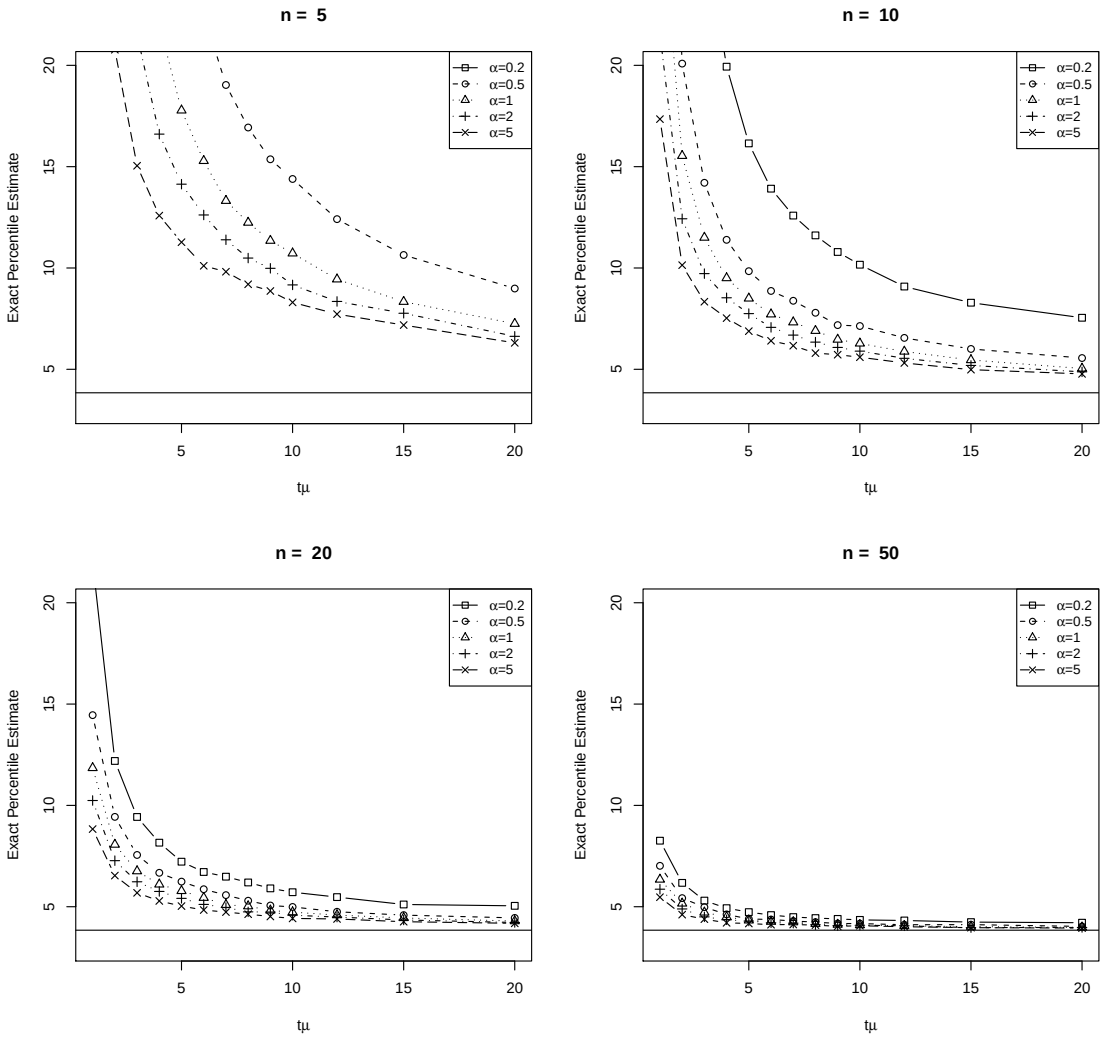

Figure 4.2: PGB Trial: Simulation based 95th percentile value for the null distribution of WT(γ^2) statistic for different values of $t\mu$ and α and $n = 5, 10, 20, 50$ subjects per group; The solid horizontal line refers to the $c_0 (= 3.8415 = \chi^2_1(0.95))$.

(and also for other Wald tests). The LRT and RST seem to maintain Type I error rates very well even for small sample sizes and hence the asymptotic approximation may be used.

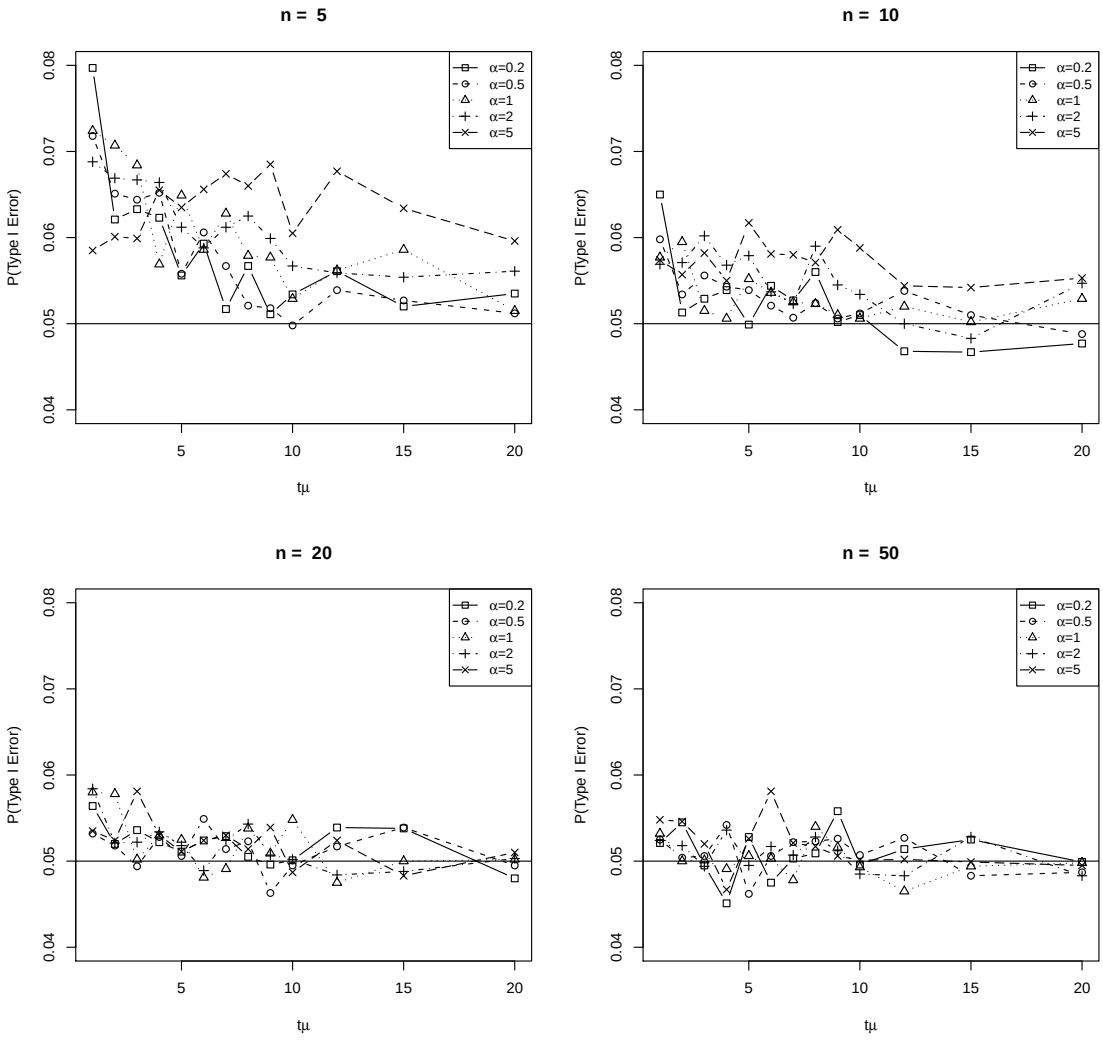

Figure 4.3: PGB Trial: Type I error rates for LRT with critical value c_0 for different values of $t\mu$ and α and $n = 5, 10, 20, 50$ subjects per group; The solid horizontal line refers to nominal level $\nu = 0.05$.

The power curves for the six *exact* tests are shown in Figure 4.7. All the tests (by definition) maintain significance levels. The WT(γ^2) has the highest power for $\gamma < 1$. For $\gamma > 1$ its power goes to 0. The LRT, RST and WT(log(γ)) are empirically

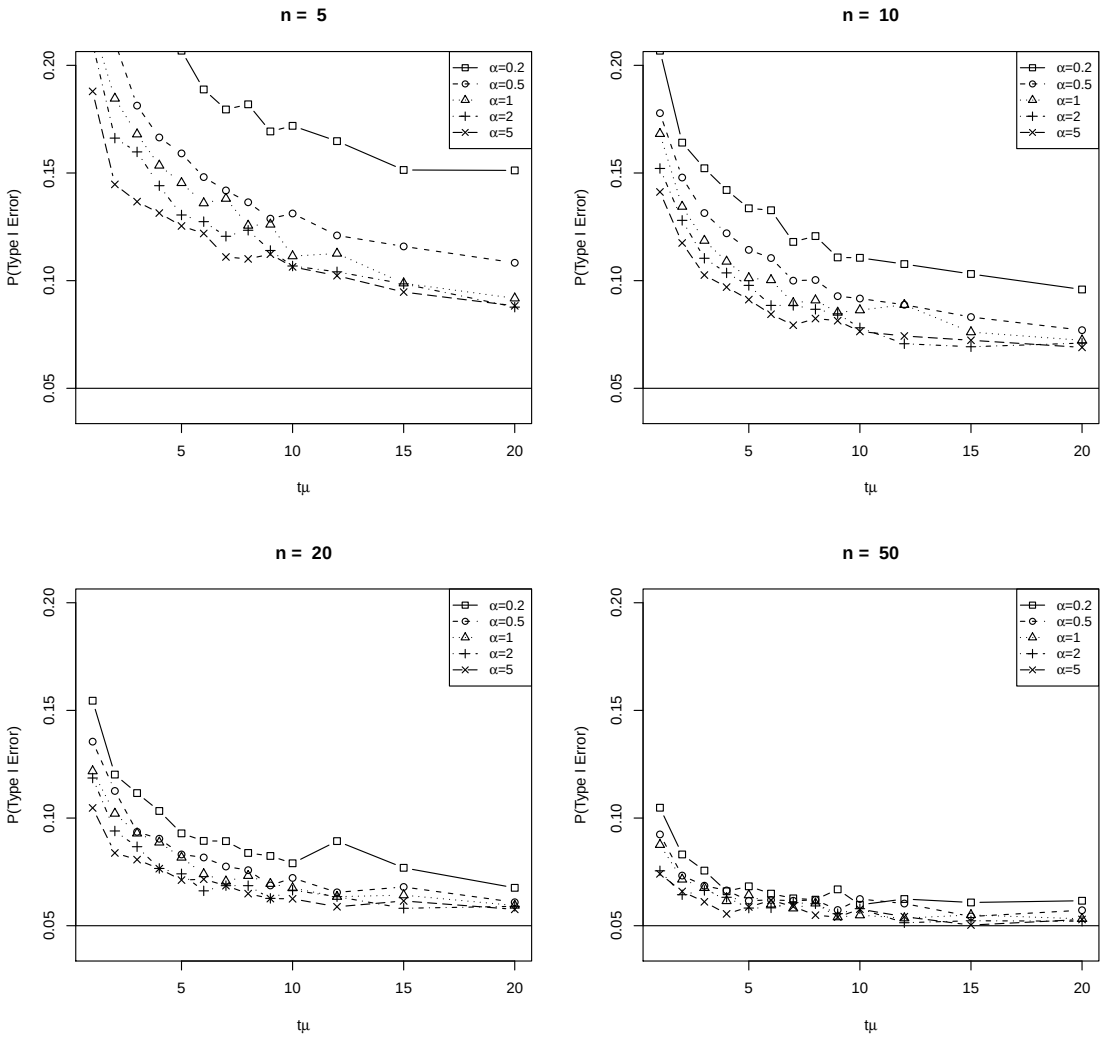

Figure 4.4: PGB Trial: Type I error rates for WT(γ^2) with critical value c_0 for different values of $t\mu$ and α and $n = 5, 10, 20, 50$ subjects per group; The solid horizontal line refers to nominal level $\nu = 0.05$.

unbiased. The WT(log(γ)) or the LRT seem to have similar power estimates for $\gamma < 1$ and their power is higher than the power corresponding to RST. Hence either the LRT or the WT(log(γ)) are recommended for $\gamma < 1$ if an unbiased test is desired.

For $\gamma > 1$, the RST has the highest power and is recommended. The WT(γ) and WT($\sqrt{\gamma}$) have consistently lesser power (although very slightly for WT(γ)) than the WT(γ^2) for $\gamma < 1$ and for $\gamma > 1$ they have lesser power than either of the three unbiased tests and thus are not considered further.

Based on the results given above it is very clear that a choice of the "best" test very much depends on the underlying research problem and the region of interest ($\gamma < 1$ vs. $\gamma > 1$). For RRMS PGB clinical trials where the interest is in reducing the number of lesions seen in the brain, a test which has the highest power to detect a reduction in the total number of brain lesions is desired. Consequently the *exact* WT(γ^2) is best suited. If on the other hand, the treatment is expected to increase the counts (of a certain response) then perhaps a RST which has the highest power for $\gamma > 1$ is more appropriate.

4.6 Sample Size Estimation

Appropriate tests based on the research hypothesis and the side of interest were suggested in the previous section. The LRT and WT(γ^2) are the two tests we are interested in the case of RRMS PGB trials (LRT is unbiased and WT(γ^2) has the highest power) and we obtain sample sizes for these only. For the LRT we estimate sample sizes based on the asymptotic approximation as well as the *exact* test and for WT(γ^2) we only use the *exact* test as the asymptotic approximation performs poorly as seen in the previous sections. The power for each sample size was computed using the method described in Section 4.5. Sample sizes (Table 4.1) to achieve 80 and 90% power are obtained for a follow-up period of $t = 3, 6$ months and for treatment effects $(1 - \gamma)$ ranging from 0.50 to 0.80. These estimates are compared to the

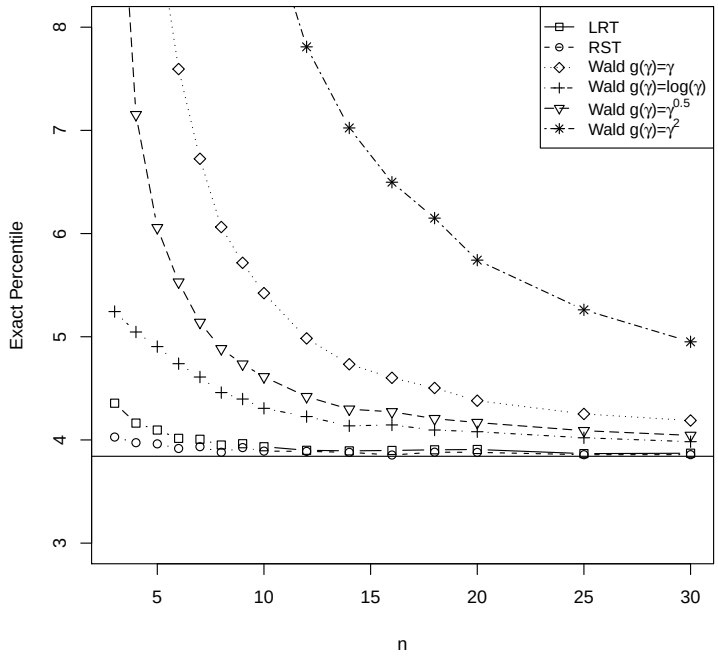

Figure 4.5: PGB Trial: Simulation based 95th percentile value for the null distribution of LRT, RST and WT statistics as a function of common sample size n; $\mu = 5.9$, $\alpha = 0.49$. The solid horizontal line refers to $c_0 (= 3.8415 = \chi^2_1(0.95))$.

sample sizes obtained by Sormani et al. [2001b] using nonparametric tests. Initial parameter estimates used was $t\mu = 5.9, \alpha = 0.49$ for a 3 month follow-up period and $t\mu = 13, \alpha = 0.52$ for a 6 month follow-up period. These parameter estimates were those obtained by Sormani et al. [2001b] by fitting a NB distribution to MRI lesion counts data obtained from a group of 66 untreated RRMS patients not selected for activity during the baseline scan. Hence we strongly recommend that the sample sizes given in this section be used only for RRMS clinical trials involving such a group of patients.

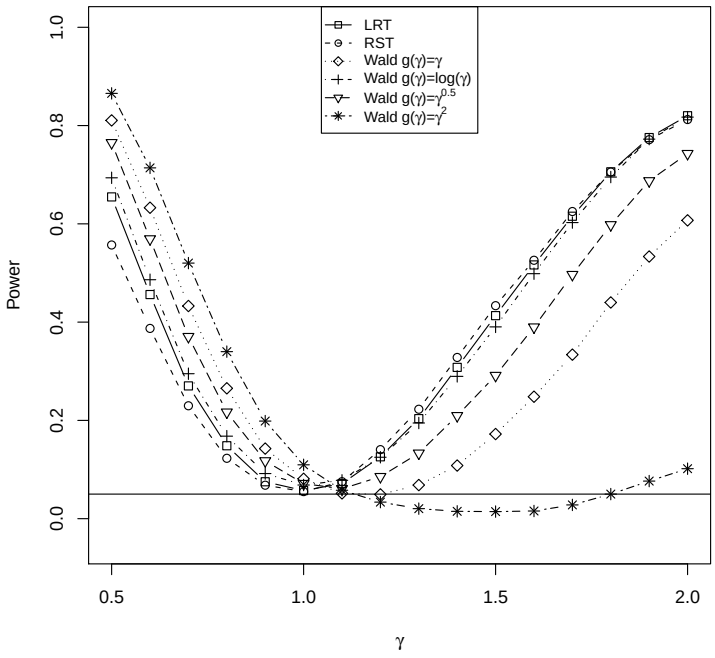

Figure 4.6: PGB trial: Power of asymptotic 5% level LRT, RST, and Wald tests for treatment effect, assuming initial parameter estimates $\mu = 5.9, \alpha = 0.49$, sample sizes $n_1 = n_2 = 10$. The solid horizontal line refers to the nominal level $\nu = 0.05$.

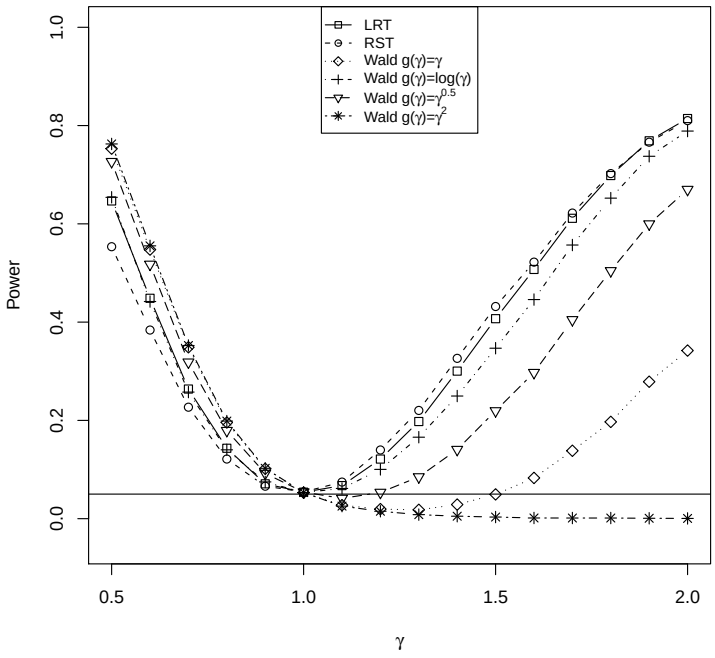

Figure 4.7: PGB trial: Power of *exact* 5% level LRT, RST, and Wald tests for treatment effect, assuming initial parameter estimates $\mu = 5.9, \alpha = 0.49$, sample sizes $n_1 = n_2 = 10$. The solid horizontal line refers to the nominal level $\nu = 0.05$.

The results show that for the LRT, the asymptotic and the *exact* sample sizes are very close in most cases and the *exact* sample sizes are higher than the asymptotic ones by at most one. This can be seen due to the fact that the asymptotic LRT maintains significance level reasonably well even for small sample sizes as seen from the discussion in the previous sections. The LRT sample sizes are around 70-80% smaller than the sample sizes obtained using WRS test. For a 3 month follow-up period assuming a 50% treatment effect the WRS Test estimates a sample of 57 subjects (minimum) per group to have 80% power where as the LRT (approximate or *exact*) estimates only 15 per group. Similar reduction in sample sizes are observed for other values of treatment effects, follow-up period and for 90% power. Also, the sample sizes using the WT(γ^2) are a further 25% smaller than the LRT sample sizes and upto 85% lesser than those obtained using WRS test. For the situation described above, the WT(γ^2) estimates only 12 per group which is a 79% reduction from the 57 estimated by WRS test.

The sample sizes for a 6 month follow-up period are 15 to 20% lesser than those for 3 months. That is if a patient is followed for approximately twice as long, we expect to see a reduction in sample sizes of around 20%. As seen in the case of PG trials, following patients for twice as long (in addition to the 1 month baseline period) does not add a whole lot of independent 'information' to the already existing knowledge we have about the parameters. Since these data are collected on the same patient the number of lesion counts seen in the patients for the second 3 months are likely very highly correlated with the number of lesions seen in the patients during the first 3 months of the follow-up period. This is the reason for the sample sizes reducing by only 20% and not by 50%. The initial estimate for $t\mu$ used in the calculations of these

sample sizes were 5.9 and 13 respectively for 3 and 6 months without much change in α. Thus we expect about a 20% reduction in sample sizes when μ is doubled without changing α. The sample sizes for achieving 90% power are 30-40% higher than those for 80% power.

As noted above, the reduction in sample sizes when using a parametric test such as LRT or WT(γ^2) is much greater for PGB trials assuming a BNB model. For PG (BVT) trials using the NB (BNB) models a 40-60% reduction was observed when parametric tests were used as opposed to nonparametric tests but for PGB trials the reduction was much greater at 80%. For PG trials, the minimum required sample size per group to detect a 50% treatment effect with 80% power and a 3 month follow-up period as estimated by LRT is 76 where as for PGB trials it is only 15 which is an 80% reduction. This suggests that adding a baseline correction scan just before the treatment is initiated results in much lesser sample sizes there by reducing trial costs. One needs to however ensure that appropriate models and appropriate tests are used.

4.7 Testing Using the Distribution of $Y - X$

When a baseline correction scan is obtained before the treatment period, the number of new enhancing lesions seen during this scan, X, may sometimes be subtracted from the total number of new enhancing lesions seen in the treatment period, Y. This method was adopted by Nauta et al. [1994], Sormani et al. [1999, 2001b]. Although nonparametric tests were employed by these authors, obtaining the distribution of $Y - X$ will allow us to employ parametric tests to test for the significance of the treatment effect. In this section we derive the distribution of $Y - X$ and show that

Table 4.1: PGB Trial: Sample sizes per group to achieve 80% and 90% power for $100(1-\gamma)\%$ treatment effect, baseline period of 1 month, and follow-up period (t) of 3 and 6 months, and initial estimates $(t\mu, \alpha) = (5.9, 0.49)$ for $t = 3$ and $(13, 0.52)$ for $t = 6$; level $\nu = 0.05$.

$1-\gamma$	t	80% power; 5% level					
		Test				Critical Value	
		WRS	LRT		WT(γ^2)	LRT	WT(γ^2)
		Sormani	Asymptotic	*Exact*	*Exact*	Simul.	Simul.
0.50	3	57	15	15	12	3.9013	7.8747
	6	50	12	12	9	3.9343	8.6777
0.60	3	42	10	10	8	3.9441	11.2759
	6	38	8	8	6	3.9570	13.1829
0.70	3	34	7	7	6	4.0040	16.3149
	6	29	6	6	4	4.0252	27.9923
0.80	3	25	5	5	4	4.0574	40.0743
	6	22	4	4	3	4.1886	75.3760
$1-\gamma$	t	90% power; 5% level					
		Test				Critical Value	
		WRS	LRT		WT(γ^2)	LRT	WT(γ^2)
		Sormani	Asymptotic	*Exact*	*Exact*	Simul.	Simul.
0.50	3	80	19	20	16	3.8904	6.5502
	6	70	16	16	12	3.8862	7.2002
0.60	3	58	13	14	10	3.8751	9.0683
	6	50	10	11	8	3.9241	9.7634
0.70	3	47	9	9	7	3.9448	13.5862
	6	38	7	8	6	3.9540	13.2785
0.80	3	35	7	7	5	3.9570	23.5071
	6	30	5	6	4	4.0322	28.1685

comparing the two groups based on this distribution is not a viable approach as the 'effect size' Δ does not always increase monotonically with longer follow-up period t.

Consider the model described in Section 4.2. Using the joint pmf of (X_1, Y_1) which is given in (4.2.1), the pmf of $Y_1 - X_1$ can be derived as

$$\begin{aligned} P(Y_1 - X_1 = k) &= \sum_{j=max(0,-k)}^{\infty} P(X_1 = j, Y_1 - X_1 = k) \\ &= \sum_{j=max(0,-k)}^{\infty} P(X_1 = j, Y_1 = k + j) \\ &= \sum_{j=max(0,-k)}^{\infty} \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\mu^j}{j!} \frac{(t\mu)^{k+j}}{(k+j)!} \frac{\Gamma(j + (k+j) + \alpha)}{(\mu + t\mu + \alpha)^{j+(k+j)+\alpha}} \\ &= \sum_{j=max(0,-k)}^{\infty} \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\mu^j}{j!} \frac{(t\mu)^{k+j}}{(k+j)!} \frac{\Gamma(2j + k + \alpha)}{(\mu + t\mu + \alpha)^{2j+k+\alpha}}, \quad k = 0, \pm 1, \ldots. \end{aligned} \tag{4.7.1}$$

Similarly, using the joint distribution of (X_2, Y_2) which is $\text{BNB}(\mu, t\gamma\mu, \nu)$, the distribution of $Y_2 - X_2$ can be found to be

$$P(Y_2 - X_2 = l) = \sum_{j=max(0,-l)}^{\infty} \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\mu^j}{j!} \frac{(t\gamma\mu)^{l+j}}{(l+j)!} \frac{\Gamma(2j + l + \alpha)}{(\mu + t\gamma\mu + \alpha)^{2j+l+\alpha}}, \quad l = 0, \pm 1, \ldots. \tag{4.7.2}$$

We note that marginally $X_1, X_2 \sim \text{NB}(\mu, \alpha)$, $Y_1 \sim \text{NB}(t\mu, \alpha)$ and $Y_2 \sim \text{NB}(t\gamma\mu, \alpha)$.

Lemma 4.7.1. *The expected values of $Y_1 - X_1$ and $Y_2 - X_2$ are*

$$\begin{aligned} E(Y_1 - X_1) &= E(Y_1) - E(X_1) = t\mu - \mu = \mu(t-1) \\ E(Y_2 - X_2) &= E(Y_2) - E(X_2) = t\gamma\mu - \mu = \mu(t\gamma - 1) \end{aligned}$$

Lemma 4.7.2. *We have*

$$Cov(X_1, Y_1) = \frac{t\mu^2}{\alpha}$$

Proof:

$$\begin{aligned}
Cov(X_1, Y_1) &= E(X_1Y_1) - E(X_1)E(Y_1) \\
&= E(E(X_1Y_1|Z = z)) - E(X_1)E(Y_1) \\
&= E(E(X_1|Z = z)E(Y_1|Z = z)) - E(X_1)E(Y_1) \\
&= E\left\{(\mu Z)(t\mu Z)\right\} - t\mu^2 \\
&= t\mu^2\{E(Z^2) - 1\} \\
&= t\mu^2\left\{\frac{1}{\alpha} + 1 - 1\right\} \quad (\text{ Since } Z \sim \text{Gamma}(\alpha, 1/\alpha)) \\
&= \frac{t\mu^2}{\alpha}
\end{aligned}$$

Similarly it can be shown that $Cov(X_2, Y_2) = t\gamma\mu^2/\alpha$. Using the fact that each of X_1, X_2, Y_1, Y_2 are marginally distributed as NB random variables and from Lemma 4.7.2 we have the following result:

Lemma 4.7.3.

$$Var(Y_1 - X_1) = \mu(1 + t) + \frac{\mu^2}{\alpha}(1 - t)^2$$

and

$$Var(Y_2 - X_2) = \mu(1 + t\gamma) + \frac{\mu^2}{\alpha}(1 - t\gamma)^2.$$

In Section 4.2, we used the joint (X, Y) distribution to construct the model for PGB trials. Similarly, we can construct a model for PGB trials using the distribution of $Y - X$ and come up with parametric tests. Parameter estimates for the model can be obtained using the maximum likelihood method and parametric tests such as LRT, RST and Wald tests can be constructed to test for the treatment effect. Although we do not provide the details of estimation and testing here, we estimated the sample sizes required to detect a significant treatment effect based on the distribution of

$Y - X$. We used the same initial parameter estimates as we used for the (X, Y) model. For example, for the LRT, to detect a 50% treatment effect with 80% power, the required sample size is 25 subjects per group for a follow-up period of 3 months and 38 per group for a follow-up period of 6 months. This is intriguing since the required sample size increases when the patients are followed for longer time periods. That is, the power of the test decreased as the follow-up period increased. The same was true even for the other parametric tests and also for the nonparametric tests. We attempt here to give a plausible explanation for this anomaly.

When comparing two population means μ_1 and μ_2, one often uses the *effect size* $\Delta = \frac{\mu_1 - \mu_2}{\sigma}$ (Cohen [1988]) as a standardized measure of difference. Here σ is the standard deviation of either population when they are assumed equal. When the standard deviations of the two populations are different, say $\sigma_1 \neq \sigma_2$, Cohen suggests to use the square root of the average of the two variances σ' in place of σ. That is

$$\sigma' = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}.$$

For comparing means of two normal populations with unequal variances (using the same notations as above), there is a direct connection between the power of the test $1 - \beta$, where $\beta = \text{P(Type II error)}$, the significance level ν, the sample size n and the effect size Δ as can be seen from the formula for the sample size;

$$\begin{aligned} n &= \frac{(\sigma_1^2 + \sigma_2^2)(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2}{(\mu_1 - \mu_2)^2} \\ &= \frac{(Z_{1-\nu/2} + Z_{1-\beta})^2}{\Delta^2}. \end{aligned} \tag{4.7.3}$$

It is clear that for a fixed sample size n, the power increases as the effect size increases. Although this is true for normal populations, the relationship between the power and effect size is not clear in our case. To estimate the *effect size* for the BNB model

above, we need to define the standard deviation σ. Since for our BNB model, the standard deviation depends on the mean, the standard deviation of the lesion counts for the two groups are different. Hence, we use a pooled standard deviation σ' as defined above. For our model,

$$\sigma' = \sqrt{\frac{Var(Y_1 - X_1) + Var(Y_2 - X_2)}{2}}. \tag{4.7.4}$$

Using Lemmas 4.7.1 and 4.7.3, the *effect size* is

$$\begin{aligned} \Delta &= \frac{E(Y_1 - X_1) - E(Y_2 - X_2)}{\sigma'} \\ &= \frac{t\mu(1-\gamma)}{\sqrt{\frac{\mu(2+t(1+\gamma))+\frac{\mu^2}{\alpha}((1-t)^2+(1-t\gamma)^2)}{2}}} \end{aligned} \tag{4.7.5}$$

To study the behaviour of Δ as the follow-up period t varies, we first consider

$$\begin{aligned} \frac{1}{\Delta^2} &= \frac{\mu(1+t) + \frac{\mu^2}{\alpha}(1-t)^2 + \mu(1+t\gamma) + \frac{\mu^2}{\alpha}(1-t\gamma)^2}{2\left[t\mu(1-\gamma)\right]^2} \\ &= \frac{\mu[2+t(1+\gamma)]}{2t^2\mu^2(1-\gamma)^2} + \frac{\mu^2[(1-t)^2+(1-t\gamma)^2]}{2\alpha t^2\mu^2(1-\gamma)^2} \\ &= \frac{\alpha+\mu}{t^2\alpha\mu(1-\gamma)^2} + \frac{1+\gamma^2}{2\alpha(1-\gamma)^2} + \frac{(1+\gamma)(\alpha-2\mu)}{2\mu\alpha t(1-\gamma)^2}. \end{aligned} \tag{4.7.6}$$

Note that the above expression decreases as t increases for all $t > 0$ when $\alpha - 2\mu \geq 0$. That is the effect size Δ increases as t increases when $\alpha \geq 2\mu$.

To determine the behavior of Δ when $\alpha < 2\mu$, we consider the first order derivative of (4.7.6) with respect to t.

$$\frac{\partial(1/\Delta^2)}{\partial t} = \frac{(\alpha+\mu)}{\alpha\mu(1-\gamma)^2}\left[\frac{-2}{t^3}\right] + \frac{(1+\gamma)(\alpha-2\mu)}{2\mu\alpha(1-\gamma)^2}\left[\frac{-1}{t^2}\right].$$

Equating the above expression to 0 and solving for t we get

$$t_0 = \frac{-4(\alpha+\mu)}{(1+\gamma)(\alpha-2\mu)}. \tag{4.7.7}$$

The second order derivative of (4.7.6) with respect to t is

$$\frac{\partial^2(1/\Delta^2)}{\partial t^2} = \frac{(\alpha+\mu)}{\alpha\mu(1-\gamma)^2}\left[\frac{6}{t^4}\right] + \frac{(1+\gamma)(\alpha-2\mu)}{2\mu\alpha(1-\gamma)^2}\left[\frac{2}{t^3}\right]$$

Then

$$\left.\frac{\partial^2(1/\Delta^2)}{\partial t^2}\right|_{t=t_0} = \frac{(1+\gamma)^4(\alpha-2\mu)^4}{128(\alpha\mu)(1-\gamma)^2(\alpha+\mu)^3}$$

which is positive when $\alpha < 2\mu$. Therefore t_0 given in (4.7.7) is the value at which $\frac{1}{\Delta^2}$ attains its minimum or Δ attains its maximum. Thus, for $\alpha < 2\mu$, the effect size Δ reaches a maximum at t_0 after which it starts to decrease. For example, for $\mu = 5.9/3$ and $\alpha = 0.49$ (the parameter estimates so far considered in thesis), Δ has a maximum at $t = 1.9025$. So for a follow-up period of 2 months or more, the effect size Δ decreases with increasing t. For example, for $(\gamma, \mu, \alpha) = (0.50, 5.9/3, 0.49)$ the effect size $\Delta = 0.61$ for $t = 3$ and 0.53 for $t = 6$. This is a bad property to have for a test and hence using the distribution of $Y - X$ to compare between two groups is not recommended when (X, Y) are jointly distributed according to a BNB distribution.

4.8 Discussion

In this chapter we proposed a BNB model to produce sample sizes for PGB trials in RRMS clinical trials. Since the initial estimates in our simulation studies were obtained by fitting the NB model to RRMS patients not selected for MRI activity at baseline, we suggest that the sample sizes reported in this chapter be used only for clinical trials involving a similar group of patients.

We propose likelihood based parametric tests such as LRT, RST and several Wald tests and compare their performances when using both asymptotic and *exact* percentiles. We recommend appropriate tests based on the properties of the test and

the region of interest desired. If an unbiased test is preferred and $\gamma < 1$ is the region of interest, then the asymptotic LRT is well suited. On the other hand, if $\gamma > 1$ is the region of importance, then the asymptotic RST is most powerful and is hence recommended. The WT(γ^2) has the highest power among all tests when $\gamma < 1$ but it is not unbiased. We suggest the use of WT(γ^2) for clinical trials in MS since the interest is in finding a test which has the highest power to detect reductions in mean lesion counts in the treatment group.

The sample sizes obtained using the LRT are 70-80% smaller than those based on nonparametric tests that are given in Sormani et al. [2001b]. The WT(γ^2) provides a further 25% reduction in required sample sizes. We considered as initial estimates, a mean ($t\mu$) of 5.9 and 13 for a follow-up period of 3 and 6 months respectively with little change in α. This doubling of the mean reduced the sample sizes by 15-20%.

We also considered simulation studies to examine the effect of μ and α on the simulated percentiles of the LRT and WT(γ^2) and the Type I error rates when one uses c_0 as the critical value. For both these tests, increasing μ reduced the simulated percentiles and brought them closer to c_0 and consequently the Type I error rates came closer to the nominal level of 0.05. Increasing α actually resulted in higher (smaller) simulated percentiles and the Type I error rates for the LRT (WT(γ^2)).

We derived the distribution of the difference $Y - X$ when (X, Y) is assumed to be distributed according to a BNB distribution and showed that the effect size for comparing between the groups based on the difference in lesion counts seen in the follow-up period to the baseline scan does not always increase monotonically when the follow-up period t increases.

The results of this chapter show that when a parametric BNB distribution is assumed, using appropriate parametric tests results in significantly higher power compared to nonparametric tests and provides vast reduction in sample sizes. These methods thus reduce the costs of clinical trials in RRMS that are prone to be very expensive. There are other situations where the BNB distribution is appropriate. For example, it fits well to number of eplipetic seizures observed in a group of 28 patients in a placebo arm of a trial (Appendix Section C.2). In such cases the methods developed in this chapter can be used as well.

CHAPTER 5

Poisson-Inverse Gaussian Distribution for Parallel Group Trials

5.1 Introduction

The Poisson-Inverse Gaussian (P-IG) distribution has been discussed by numerous authors in the literature as a possible alternative to the negative binomial distribution in modelling overdispersed count data. Willmot [1987] studied the P-IG distribution and showed its mathematical and statistical properties are very similar to those of the NB distribution. Hence it is only natural to think of P-IG as a possible alternative to the NB model for modelling MRI count data arising out of RRMS clinical trials. Sormani et al. [2001a] show that the P-IG distribution provides a marginally better fit than the NB distribution for MRI lesion counts in 115 RRMS patients selected for having at least one enhancing lesion on the baseline scan. These data are given in Appendix Section C.1.3.

In this chapter we use the P-IG distribution as a model for the MRI count data in PG trials involving RRMS patients who are selected for activity at baseline and derive parametric tests for the treatment effect. In Section 5.2.2, we use the P-IG

distribution to model MRI count data in such trials. The parameters of the model are estimated in Section 5.3 and likelihood based parametric tests for the treatment effect are developed in Section 5.4. In Section 5.5, the performance of the proposed tests are compared and sample size estimates for RRMS PG trials are estimated in Section 5.6. We conclude with some discussion in Section 5.7.

5.2 The Poisson-Inverse Gaussian (P-IG) Model

5.2.1 The Basic Model

Tweedie [1957] discusses the properties of the Inverse Gaussian (IG) distribution with density function given by

$$f_Z(z|\mu,\lambda) = \left[\frac{\lambda}{2\pi z^3}\right]^{\frac{1}{2}} \exp\left\{\frac{-\lambda(z-\mu)^2}{2\mu^2 z}\right\}, \quad z>0,\ \mu,\lambda>0. \tag{5.2.1}$$

We use the notation IG(μ,λ) to denote IG distribution with the above density function. The mean of an IG(μ,λ) random variable is μ and its variance is μ^3/λ. The parameter λ is seen as a measure of relative precision. Tweedie shows that the MLE of μ is the sample mean and that the distributions of the MLE of μ and λ are stochastically independent.

Holla [1971] first derived the P-IG distribution by assuming that the parameter of the Poisson distribution follows an IG distribution and discussed its applications to accident statistics. Ord and Whitmore [1986] discuss the P-IG distribution as a model for species abundance.

Assume $Y|Z=z \sim$ Poisson(z) and that Z is further distributed according to IG(μ,λ) with density in (5.2.1). As noted in Section 1.3.1, the marginal distribution

of Y is then P-IG(μ, λ) with pmf

$$P(Y = y) = p_y = \int_0^\infty \frac{e^{-z} z^y}{y!} f_Z(z|\mu, \lambda) dz$$
$$= \frac{\tau^y}{y!} \left(\frac{2\theta}{\pi}\right)^{\frac{1}{2}} \exp\left(\frac{\lambda}{\mu}\right) K_{y-\frac{1}{2}}(\theta), \qquad y = 0, 1, \ldots, \tag{5.2.2}$$

Here $\tau = \left[\frac{1}{\mu^2} + \frac{2}{\lambda}\right]^{-\frac{1}{2}}$, $\theta = \frac{\lambda}{\tau}$ and $K(\cdot)$ is the modified Bessel function of the third kind. It is defined as

$$K_\nu(z) = \frac{\pi}{2} \cdot \frac{I_{-\nu}(z) - I_\nu(z)}{\sin \nu\pi} \tag{5.2.3}$$

where $I(\cdot)$ is the modified Bessel function of the first kind defined as

$$I_\nu(z) = \sum_{m=0}^{\infty} \frac{(\frac{z}{2})^{\nu+2m}}{m!\Gamma(m + \nu + 1)}. \tag{5.2.4}$$

It can be seen from (5.2.2) that

$$p_0 = \exp\left\{\frac{\lambda}{\mu} - \frac{\lambda}{\tau}\right\} \text{ and } p_1 = \tau p_0. \tag{5.2.5}$$

Also, the P-IG probabilities satisfy the following recurrence relation:

$$p_y = \tau^2 \left\{\frac{p_{y-2}}{y(y-1)} + \frac{2y-3}{\lambda y} p_{y-1}\right\}, \qquad y = 2, 3, \ldots. \tag{5.2.6}$$

To compute the P-IG probabilities using the closed form expression given in (5.2.2) we need to evaluate the Bessel function of the third kind $K(\cdot)$. This can be computed easily for small values of y, but for large values, (5.2.3) and/or (5.2.2) may return infinite values in which case the probabilities cannot be computed directly. To avoid this problem, one can first check if (5.2.2) returns a finite value for the maximum value of y in the observed data. If yes, the direct expression in (5.2.2) can be used; otherwise, the recursive relation (5.2.6) needs to be used. This additional comparison operation for each data led to a 15% increase in computational times when compared

to using the recurrence relation alone to obtain the probabilities. Thus the use of recurrence relation given in (5.2.6) is suggested for computing the probabilities very efficiently.

The mean and variance of P-IG(μ, λ) are μ and $\mu + \mu^3/\lambda$ respectively. We also observe here that the variance of the P-IG distribution exceeds the variance of the Poisson distribution by μ^3/λ, which is the variance of the mixing IG distribution.[3]

The P-IG distribution is unimodal and right skewed as the NB model and hence can be used as a possible model for overdispersed count data. The MLE of the mean of the P-IG distribution is the sample mean itself (Stein et al. [1987]) as in the case of the NB model. Hence it is only natural to think of P-IG as a possible alternative to the NB model for modelling MRI count data arising out of RRMS clinical trials.

Remark: Sichel [1971] derived a generalized P-IG distribution by mixing Poisson distribution with a generalized IG distribution. The resulting distribution with three shape parameters has pmf

$$P(X = x) = \frac{(1-\theta)^{\gamma/2}}{K_\gamma(\alpha\sqrt{1-\theta})} \frac{\alpha\theta/2}{x!} K_{x+\gamma}(\alpha), \quad \alpha > 0,\ 0 < \theta < 1,\ -\infty < \gamma < \infty. \tag{5.2.7}$$

This distribution is called the Sichel distribution and we denote it as P-IG(α, θ, γ). When $\gamma = -\frac{1}{2}$ is known, this becomes the two parameter P-IG distribution (although under a different parametrization) for which Sichel derives the MLEs. This parametrization adopted by Sichel leads to the MLEs of α and θ being strongly correlated resulting in an unstable estimation process. Stein et al. [1987] propose a parametrization in terms of the population mean where the MLEs of the two parameters are asymptotically independent.

[3]This is true for any compound Poisson distribution.

5.2.2 The P-IG Model for PG Trials

Suppose there are n_1 subjects in the placebo group and n_2 subjects in the treatment group. Let $Y_i,\ i = 1, 2$ denote the total number of new enhancing lesions seen for the patients in the placebo group and the treatment group, respectively. Let $Z_i,\ i = 1, 2$ denote the random subject effects seen for two groups. We assume that

$$Y_1|Z_1 = z_1 \ \sim \text{Poisson}(z_1) \text{ and } Y_2|Z_2 = z_2 \ \sim \text{Poisson}(z_2). \tag{5.2.8}$$

We also assume that Z_1 and Z_2 follow IG(μ, λ) and IG$(\gamma\mu, \lambda)$ respectively. Here $1 - \gamma$ is the measure of the treatment effect which can be viewed as the percentage reduction in the mean lesion counts seen in the treatment group. Then, it can be seen that Y_1 and Y_2 are independently and marginally distributed as P-IG(μ, λ) and P-IG$(\gamma\mu, \lambda)$ respectively. Using the pmf for P-IG given in (5.2.2), the marginal pmfs of Y_1 and Y_2 can be seen to be

$$\begin{aligned} P(Y_1 = y_1) &= \frac{\tau_1^{y_1}}{y_1!}\left(\frac{2\theta_1}{\pi}\right)^{\frac{1}{2}} \exp\left\{\frac{\lambda}{\mu}\right\} K_{y_1-\frac{1}{2}}(\theta_1) \\ P(Y_2 = y_2) &= \frac{\tau_2^{y_2}}{y_2!}\left(\frac{2\theta_2}{\pi}\right)^{\frac{1}{2}} \exp\left\{\frac{\lambda}{\gamma\mu}\right\} K_{y_2-\frac{1}{2}}(\theta_2) \end{aligned} \tag{5.2.9}$$

where

$$\begin{aligned} \tau_1 &= \left[\frac{1}{\mu^2} + \frac{2}{\lambda}\right]^{-\frac{1}{2}} \\ \tau_2 &= \left[\frac{1}{\gamma^2\mu^2} + \frac{2}{\lambda}\right]^{-\frac{1}{2}} \text{ and} \\ \theta_i &= \frac{\lambda}{\tau_i},\ i = 1, 2. \end{aligned}$$

Then $E(Y_1) = \mu$, $Var(Y_1) = \mu + \mu^3/\lambda$, $E(Y_2) = \gamma\mu$ and $Var(Y_2) = \gamma\mu + (\gamma\mu)^3/\lambda$. Note that we assume the same shape parameter λ for the two groups.

The following lemma presents some useful results needed to obtain the score vector and the matrix of the second order derivatives.

Lemma 5.2.1.

$$\frac{\partial\tau_1}{\partial\gamma}=0; \qquad \frac{\partial\tau_1}{\partial\mu}=\frac{\tau_1^3}{\mu^3}; \qquad \frac{\partial\tau_1}{\partial\lambda}=\frac{\tau_1^3}{\lambda^2}.$$
$$\frac{\partial\tau_2}{\partial\gamma}=\frac{\tau_2^3}{\gamma^3\mu^2}; \quad \frac{\partial\tau_2}{\partial\mu}=\frac{\tau_2^3}{\gamma^2\mu^3}; \quad \frac{\partial\tau_2}{\partial\lambda}=\frac{\tau_2^3}{\lambda^2}.$$

Using the above results we further obtain

$$\frac{\partial\theta_1}{\partial\gamma}=0; \qquad \frac{\partial\theta_1}{\partial\mu}=-\frac{\lambda\tau_1}{\mu^3}; \quad \frac{\partial\theta_1}{\partial\lambda}=\frac{\lambda-\tau_1^2}{\lambda\tau_1}$$
$$\frac{\partial\theta_2}{\partial\gamma}=-\frac{\lambda\tau_2}{\gamma^3\mu^2}; \quad \frac{\partial\theta_2}{\partial\mu}=-\frac{\lambda\tau_2}{\gamma^2\mu^3}; \quad \frac{\partial\theta_2}{\partial\lambda}=\frac{\lambda-\tau_2^2}{\lambda\tau_2}.$$

5.3 Estimation

The parameters for the P-IG model given in the previous section can be estimated using the method of maximum likelihood. Suppose there are n_1 subjects in the placebo group and n_2 subjects in the treatment group. For observed counts $\mathbf{y}_1 = (y_{11}, y_{12}, \ldots, y_{1n_1})$ and $\mathbf{y}_2 = (y_{21}, y_{22}, \ldots, y_{2n_2})$, the likelihood function for the model in (5.2.8) is given by

$$\begin{aligned} L(\gamma,\mu,\lambda|\mathbf{y}_1,\mathbf{y}_2) &= \prod_{i=1}^{n_1} P(Y_1=y_{1i}) \times \prod_{j=1}^{n_2} P(Y_2=y_{2j}) \\ &= \frac{(\tau_1)^{n_1\bar{y}_1}}{\prod y_{1i}!}\left(\frac{2\lambda}{\pi\tau_1}\right)^{\frac{n_1}{2}} \exp\left\{\frac{n_1\lambda}{\mu}\right\} \prod_{i=1}^{n_1} K_{y_{1i}-\frac{1}{2}}(\theta_1) \\ &\quad \times \frac{(\tau_2)^{n_2\bar{y}_2}}{\prod y_{2j}!}\left(\frac{2\lambda}{\pi\tau_2}\right)^{\frac{n_2}{2}} \exp\left\{\frac{n_2\lambda}{\gamma\mu}\right\} \prod_{j=1}^{n_2} K_{y_{2j}-\frac{1}{2}}(\theta_2) \qquad (5.3.1) \end{aligned}$$

Then the log-likelihood function is

$$\begin{aligned}\ell(\gamma,\mu,\lambda|\mathbf{y}_1,\mathbf{y}_2) =& n_1\bar{y}_1\log(\tau_1)+\frac{n_1}{2}\log\left(\frac{2\lambda}{\pi\tau_1}\right)+\frac{n_1\lambda}{\mu}+\sum_{i=1}^{n_1}\log K_{y_{1i}-\frac{1}{2}}(\theta_1)\\ &+n_2\bar{y}_2\log(\tau_2)+\frac{n_2}{2}\log\left(\frac{2\lambda}{\pi\tau_2}\right)+\frac{n_2\lambda}{\gamma\mu}+\sum_{j=1}^{n_2}\log K_{y_{2j}-\frac{1}{2}}(\theta_2)\\ &+\sum_{i=1}^{n_1}\log y_{1i}!+\sum_{j=1}^{n_2}\log y_{2j}!. \end{aligned} \tag{5.3.2}$$

The following lemma provides useful results involving modified Bessel functions that simplify the derivation of the score vector and the second order derivatives.

Lemma 5.3.1. *(Modified Bessel function of the third kind (See Section 9.6, Abramowitz and Stegun [1970])).*

The following relations hold for the modified Bessel function of the third kind $K_\nu(z)$*:*

$$\begin{aligned} K_{-\nu}(z) &= K_\nu(z)\\ K_{-\frac{1}{2}}(z) &= K_{\frac{1}{2}}(z)=\sqrt{\frac{\pi}{2z}}e^{-z}\\ K_{\nu+1}(z) &= K_{\nu-1}(z)+\frac{2\nu}{z}K_\nu(z)\\ \frac{\partial}{\partial z}K_\nu(z) &= K'_\nu(z)=-K_{\nu+1}(z)+\frac{\nu}{z}K_\nu(z). \end{aligned} \tag{5.3.3}$$

The ratio of modified Bessel functions $R_\nu(z)=\frac{K_{\nu+1}(z)}{K_\nu(z)}$*, satisfies the following relations:*

$$\begin{aligned} R_{-\frac{1}{2}}(z) &= 1;\\ R_\nu(z) &= \frac{2\nu}{z}+\frac{1}{R_{\nu-1}(z)},\quad \nu=\frac{1}{2},\frac{3}{2},\frac{5}{2},\ldots;\\ \frac{\partial}{\partial z}R_\nu(z) &= R'_\nu(z)=R^2_\nu(z)-\frac{2(\nu+1/2)}{z}R_\nu(z)-1. \end{aligned} \tag{5.3.4}$$

The score vector components for the log-likelihood function given in (5.3.2) are

$$\frac{\partial \ell(\gamma,\mu,\lambda)}{\partial \gamma} = \frac{\lambda\tau_2}{\gamma^3\mu^2}\sum_{j=1}^{n_2} R_{y_{2j}-\frac{1}{2}}(\theta_2) - \frac{n_2\lambda}{\gamma^2\mu}, \tag{5.3.5}$$

$$\frac{\partial \ell(\gamma,\mu,\lambda)}{\partial \mu} = \frac{\lambda\tau_1}{\mu^3}\sum_{i=1}^{n_1} R_{y_{1i}-\frac{1}{2}}(\theta_1) + \frac{\lambda\tau_2}{\gamma^2\mu^3}\sum_{j=1}^{n_2} R_{y_{2j}-\frac{1}{2}}(\theta_2) - \frac{\lambda(n_1\gamma+n_2)}{\gamma\mu^2}, \tag{5.3.6}$$

$$\begin{aligned}\frac{\partial \ell(\gamma,\mu,\lambda)}{\partial \lambda} =& \frac{n_1\bar{y}_1+n_2\bar{y}_2}{\lambda} + \frac{n_1\gamma+n_2}{\gamma\mu} - \left(\frac{\lambda-\tau_1^2}{\lambda\tau_1}\right)\sum_{i=1}^{n_1} R_{y_{1i}-\frac{1}{2}}(\theta_1) \\ & - \left(\frac{\lambda-\tau_2^2}{\lambda\tau_2}\right)\sum_{j=1}^{n_2} R_{y_{2j}-\frac{1}{2}}(\theta_2).\end{aligned} \tag{5.3.7}$$

The MLEs of the parameters γ, μ and λ can be obtained by equating the score vector equations (5.3.5) - (5.3.7) to zero and by solving simultaneously for the three parameters. Equating the first two components to 0, we readily obtain

$$\sum_{i=1}^{n_1} R_{y_{1i}-\frac{1}{2}}(\theta_1) = \frac{n_1\mu}{\tau_1} \quad \text{and} \quad \sum_{j=1}^{n_2} R_{y_{2j}-\frac{1}{2}}(\theta_2) = \frac{n_2\gamma\mu}{\tau_2}.$$

Using these results in the third score vector component equated to 0, and using the fact that

$$\left(\frac{\lambda-\tau_1^2}{\tau_1^2}\right) = \frac{\lambda+\mu^2}{\mu^2} \quad \text{and} \quad \left(\frac{\lambda-\tau_2^2}{\tau_2^2}\right) = \frac{\lambda+\gamma^2\mu^2}{\gamma^2\mu^2},$$

we get

$$n_1(\bar{y}_1-\mu) + n_2(\bar{y}_2-\gamma\mu) = 0. \tag{5.3.8}$$

One solution for equation (5.3.8) is that $\hat{\mu} = \bar{y}_1$ and $\hat{\gamma}\hat{\mu} = \bar{y}_2$ which is a local maxima. We examined the surface of the log-likelihood as a function of γ and μ for selected λ values (figures not shown). This is a smooth concave function for the values of the parameters considered with a unique maxima attained at $\hat{\mu} = \bar{y}_1$ and $\hat{\gamma} = \bar{y}_2/\bar{y}_1$, where λ maximized the profile log-likelihood $\ell(\bar{y}_2/\bar{y}_1, \bar{y}_1, \lambda)$. Thus the local maxima obtained as a solution to the score vector equations ((5.3.5)-(5.3.7)) can be argued to be the MLEs of the parameters which leads us to the following conclusion:

Lemma 5.3.2. *The MLEs of μ and γ are*

$$\hat{\mu} = \bar{y}_1 \quad \text{and} \quad \hat{\gamma} = \frac{\bar{y}_2}{\bar{y}_1}.$$

The MLE of λ, $\hat{\lambda}$ can be obtained as a solution to the equation

$$\left(\frac{\hat{\lambda} - \hat{\tau}_1^2}{\lambda\hat{\tau}_1}\right)\sum_{i=1}^{n_1} R_{y_{1i}-\frac{1}{2}}(\hat{\theta}_1) + \left(\frac{\hat{\lambda} - \hat{\tau}_2^2}{\hat{\lambda}\hat{\tau}_2}\right)\sum_{j=1}^{n_2} R_{y_{2j}-\frac{1}{2}}(\hat{\theta}_2) = \frac{n_1\bar{y}_1 + n_2\bar{y}_2}{\hat{\lambda}} + \frac{n_1\hat{\gamma} + n_2}{\hat{\gamma}\hat{\mu}},$$

where $\hat{\theta}_i = \hat{\lambda}/\hat{\tau}_i,\ i = 1, 2$ and

$$\hat{\tau}_1 = \left[\frac{1}{\hat{\mu}^2} + \frac{2}{\hat{\lambda}}\right]^{-\frac{1}{2}} \quad \text{and} \quad \hat{\tau}_2 = \left[\frac{1}{\hat{\gamma}^2\hat{\mu}^2} + \frac{2}{\hat{\lambda}}\right]^{-\frac{1}{2}}.$$

Alternatively the MLE of λ can also be obtained by numerically maximizing the profile log-likelihood function $\ell(\hat{\gamma}, \hat{\mu}, \lambda)$ with respect to λ.

When $\gamma = \gamma_0$ is assumed known, the MLEs of μ and λ can be obtained by setting the score equations (5.3.6) and (5.3.7) to 0 with $\gamma = \gamma_0$ and simultaneously solving for the two parameters. For a general γ_0 the MLEs are not available in closed form and numerical methods seen previously must be employed. However, for $\gamma_0 = 1$, we have $\tau_1 = \tau_2$ and $\theta_1 = \theta_2$ using which we have the following result:

Lemma 5.3.3. *When $\gamma = 1$, the MLE of μ is*

$$\tilde{\mu} = \frac{n_1\bar{y}_1 + n_2\bar{y}_2}{n_1 + n_2}.$$

The MLE of λ, $\tilde{\lambda}$ solves the equation

$$\left(\frac{\lambda - \tilde{\tau}_1^2}{\lambda\tilde{\tau}_1}\right)\sum_{i=1}^{n_1} R_{y_{1i}+\frac{1}{2}}(\tilde{\theta}_1) + \left(\frac{\tilde{\lambda} - \tilde{\tau}_2^2}{\tilde{\lambda}\tilde{\tau}_2}\right)\sum_{j=1}^{n_2} R_{y_{2j}-\frac{1}{2}}(\tilde{\theta}_2) = \frac{n_1\bar{y}_1 + n_2\bar{y}_2}{\tilde{\lambda}} + \frac{n_1 + n_2}{\tilde{\mu}}$$

where $\tilde{\theta}_i = \tilde{\lambda}/\tilde{\tau}_i,\ i = 1, 2$ and

$$\tilde{\tau}_1 = \left[\frac{1}{\tilde{\mu}^2} + \frac{2}{\tilde{\lambda}}\right]^{-\frac{1}{2}} \quad \text{and} \quad \tilde{\tau}_2 = \left[\frac{1}{\tilde{\mu}^2} + \frac{2}{\tilde{\lambda}}\right]^{-\frac{1}{2}}.$$

The MLE of λ can also be obtained by numerically maximizing the profile log-likelihood function $\ell(1, \tilde{\mu}, \lambda)$ with respect to λ.

The second order derivatives of the log-likelihood function are

$$
\begin{aligned}
\frac{\partial^2 \ell(\gamma,\mu,\lambda)}{\partial \gamma^2} =& -\frac{\lambda^2\tau_2^2}{\gamma^6\mu^4}\sum_{j=1}^{n_2} R'_{y_{2j}-\frac{1}{2}}(\theta_2) + \frac{\lambda\tau_2}{\gamma^6\mu^4}(\tau_2^2-3\gamma^2\mu^2)\sum_{j=1}^{n_2} R_{y_{2j}-\frac{1}{2}}(\theta_2) + \frac{2n_2\lambda}{\gamma^3\mu},\\
\frac{\partial^2 \ell(\gamma,\mu,\lambda)}{\partial \gamma\partial\mu} =& -\frac{\lambda^2\tau_2^2}{\gamma^5\mu^5}\sum_{j=1}^{n_2} R'_{y_{2j}-\frac{1}{2}}(\theta_2) + \frac{\lambda\tau_2^3}{\gamma^5\mu^5}(\tau_2^2-2\gamma^2\mu^2)\sum_{j=1}^{n_2} R_{y_{2j}-\frac{1}{2}}(\theta_2) + \frac{n_2\lambda}{\gamma^2\mu^2},\\
\frac{\partial^2 \ell(\gamma,\mu,\lambda)}{\partial \gamma\partial\lambda} =& \left(\frac{\lambda-\tau_2^2}{\gamma^3\mu^2}\right)\sum_{j=1}^{n_2} R'_{y_{2j}-\frac{1}{2}}(\theta_2) + \frac{\tau_2}{\gamma^3\mu^2\lambda}(\tau_2^2+\lambda)\sum_{j=1}^{n_2} R_{y_{2j}-\frac{1}{2}}(\theta_2) - \frac{n_2}{\gamma^2\mu},
\end{aligned}
\tag{5.3.9}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\gamma,\mu,\lambda)}{\partial \mu^2} =& -\frac{\lambda^2\tau_1^2}{\mu^6}\sum_{i=1}^{n_1} R'_{y_{1i}-\frac{1}{2}}(\theta_1) + \frac{\lambda\tau_1}{\mu^6}(\tau_1^2-3\mu^2)\sum_{i=1}^{n_1} R_{y_{1i}-\frac{1}{2}}(\theta_1)\\
& -\frac{\lambda^2\tau_2^2}{\gamma^4\mu^6}\sum_{j=1}^{n_2} R'_{y_{2j}-\frac{1}{2}}(\theta_2) + \frac{\lambda\tau_2}{\gamma^4\mu^6}(\tau_2^2-3\gamma^2\mu^2)\sum_{j=1}^{n_2} R_{y_{2j}-\frac{1}{2}}(\theta_2)\\
& + \frac{2\lambda(n_1\gamma+n_2)}{\gamma\mu^3},\\
\frac{\partial^2 \ell(\gamma,\mu,\lambda)}{\partial \mu\partial\lambda} =& \left(\frac{\lambda-\tau_1^2}{\mu^3}\right)\sum_{i=1}^{n_1} R'_{y_{1i}-\frac{1}{2}}(\theta_1) + \frac{\tau_1}{\lambda\mu^3}(\tau_1^2+\lambda)\sum_{i=1}^{n_1} R_{y_{1i}-\frac{1}{2}}(\theta_1)\\
& + \left(\frac{\lambda-\tau_2^2}{\gamma^2\mu^3}\right)\sum_{j=1}^{n_2} R'_{y_{2j}-\frac{1}{2}}(\theta_2) + \frac{\tau_2}{\lambda\gamma^2\mu^3}(\tau_2^2+\lambda)\sum_{j=1}^{n_2} R_{y_{2j}-\frac{1}{2}}(\theta_2)\\
& - \frac{n_1\gamma+n_2}{\gamma\mu^2},\\
\frac{\partial^2 \ell(\gamma,\mu,\lambda)}{\partial \lambda^2} =& -\frac{n_1\bar{y}_1+n_2\bar{y}_2}{\lambda^2} - \left(\frac{\lambda-\tau_1^2}{\lambda\tau_1}\right)^2\sum_{i=1}^{n_1} R'_{y_{1i}-\frac{1}{2}}(\theta_1) + \frac{\tau_1^3}{\lambda^3}\sum_{i=1}^{n_1} R_{y_{1i}-\frac{1}{2}}(\theta_1)\\
& - \left(\frac{\lambda-\tau_2^2}{\lambda\tau_2}\right)^2\sum_{j=1}^{n_2} R'_{y_{2j}-\frac{1}{2}}(\theta_2) + \frac{\tau_2^3}{\lambda^3}\sum_{j=1}^{n_2} R_{y_{2j}-\frac{1}{2}}(\theta_2).
\end{aligned}
$$

Lemma 5.3.4. *Using the fact that $E(\bar{Y}_1)=\mu$ and $E(\bar{Y}_2)=\gamma\mu$, and from the fact that $Y_{1i},\ i=1,ldots,n_1$ and $Y_{2j},\ j=1,\ldots,n_2$ are respectively identically distributed, the elements of the FIM $\mathbf{I}(\boldsymbol{\theta})$, where $\big(I(\boldsymbol{\theta})\big)_{i,j} = -E\left\{\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\right\}$ and parameter vector*

$\boldsymbol{\theta} = (\gamma, \mu, \lambda)$ *are:*

$$
\begin{aligned}
I_{11}(\boldsymbol{\theta}) =& \frac{n_2\lambda^2\tau_2^2}{\gamma^6\mu^4} E\left\{R'_{Y_2-\frac{1}{2}}(\theta_2)\right\} - \frac{n_2\lambda\tau_2}{\gamma^6\mu^4}(\tau_2^2 - 3\gamma^2\mu^2) E\left\{R_{Y_2-\frac{1}{2}}(\theta_2)\right\} - \frac{2n_2\lambda}{\gamma^3\mu},\\
I_{12}(\boldsymbol{\theta}) =& \frac{n_2\lambda^2\tau_2^2}{\gamma^5\mu^5} E\left\{R'_{Y_2-\frac{1}{2}}(\theta_2)\right\} - \frac{n_2\lambda\tau_2^3}{\gamma^5\mu^5}(\tau_2^2 - 2\gamma^2\mu^2) E\left\{R_{Y_2-\frac{1}{2}}(\theta_2)\right\} - \frac{n_2\lambda}{\gamma^2\mu^2},\\
I_{13}(\boldsymbol{\theta}) =& -n_2\left(\frac{\lambda-\tau_2^2}{\gamma^3\mu^2}\right) E\left\{R'_{Y_2-\frac{1}{2}}(\theta_2)\right\} - \frac{n_2\tau_2}{\gamma^3\mu^2\lambda}(\tau_2^2+\lambda) E\left\{R_{Y_2-\frac{1}{2}}(\theta_2)\right\} + \frac{n_2}{\gamma^2\mu},\\
I_{22}(\boldsymbol{\theta}) =& \frac{n_1\lambda^2\tau_1^2}{\mu^6} E\left\{R'_{Y_1-\frac{1}{2}}(\theta_1)\right\} - \frac{n_1\lambda\tau_1}{\mu^6}(\tau_1^2 - 3\mu^2) E\left\{R_{Y_1-\frac{1}{2}}(\theta_1)\right\} - \frac{2\lambda(n_1\gamma+n_2)}{\gamma\mu^3}\\
&+ \frac{n_2\lambda^2\tau_2^2}{\gamma^4\mu^6} E\left\{R'_{Y_2-\frac{1}{2}}(\theta_2)\right\} - \frac{n_2\lambda\tau_2}{\gamma^4\mu^6}(\tau_2^2 - 3\gamma^2\mu^2) E\left\{R_{Y_2-\frac{1}{2}}(\theta_2)\right\},\\
I_{23}(\boldsymbol{\theta}) =& -n_1\left(\frac{\lambda-\tau_1^2}{\mu^3}\right) E\left\{R'_{Y_1-\frac{1}{2}}(\theta_1)\right\} - \frac{n_1\tau_1}{\lambda\mu^3}(\tau_1^2+\lambda) E\left\{R_{Y_1-\frac{1}{2}}(\theta_1)\right\} + \frac{n_1\gamma+n_2}{\gamma\mu^2}\\
&- n_2\left(\frac{\lambda-\tau_2^2}{\gamma^2\mu^3}\right) E\left\{R'_{Y_2-\frac{1}{2}}(\theta_2)\right\} - \frac{n_2\tau_2}{\lambda\gamma^2\mu^3}(\tau_2^2+\lambda) E\left\{R_{Y_2-\frac{1}{2}}(\theta_2)\right\},\\
I_{33}(\boldsymbol{\theta}) =& \frac{n_1\mu+n_2\gamma\mu}{\lambda^2} + n_1\left(\frac{\lambda-\tau_1^2}{\lambda\tau_1}\right)^2 E\left\{R'_{Y_1-\frac{1}{2}}(\theta_1)\right\} - \frac{n_1\tau_1^3}{\lambda^3} E\left\{R_{Y_1-\frac{1}{2}}(\theta_1)\right\}\\
&+ n_2\left(\frac{\lambda-\tau_2^2}{\lambda\tau_2}\right)^2 E\left\{R'_{Y_2-\frac{1}{2}}(\theta_2)\right\} - \frac{n_2\tau_2^3}{\lambda^3} E\left\{R_{Y_2-\frac{1}{2}}(\theta_2)\right\}.
\end{aligned}
$$

and $I_{ij} = I_{ji}$ *for* $i \neq j$.

Lemma 5.3.5. *The asymptotic variance of* $\hat{\gamma}$ *for the P-IG model is the first element of the inverse of the FIM. That is,* $\sigma^2_{\hat{\gamma}}(\boldsymbol{\theta}) = I^{-1}_{1.2} = [I_{11} - I_{12}I_{22}^{-1}I_{21}]^{-1}$ *where the elements of the FIM are as in Lemma 5.3.4.*

5.4 Hypothesis Testing

In this section we propose parametric tests such as the LRT, RST and WT to test for the treatment effect for a general $H_0 : \gamma = \gamma_0$ vs. $H_1 : \gamma \neq \gamma_0$. For RRMS clinical trials, a test for no treatment effect would test $H_0 : \gamma = 1$ vs. $H_1 : \gamma \neq 1$. In the results that follow, $\hat{\gamma}$, $\hat{\mu}$, $\hat{\lambda}$ denote the MLEs with no restrictions on the parameters and are given in Lemma 5.3.2. Further, $\tilde{\mu}$, $\tilde{\lambda}$ denote the MLEs of the parameters under H_0 and are given in Lemma 5.3.3.

Theorem 5.4.1. *(LRT). The LRT statistic to test* $H_0 : \gamma = \gamma_0$ *vs.* $H_1 : \gamma \neq \gamma_0$ *is*

$$LRT = -2(\ell(\gamma_0, \tilde{\mu}, \tilde{\alpha}) - \ell(\hat{\gamma}, \hat{\mu}, \hat{\alpha})).$$

Theorem 5.4.2. *(RST). The RST statistic to test* $H_0 : \gamma = \gamma_0$ *vs.* $H_1 : \gamma \neq \gamma_0$ *is*

$$\begin{aligned} RST &= \left[\frac{\partial \ell(\gamma, \mu, \alpha)}{\partial \gamma}\right]^2_{\theta=\tilde{\theta}} \times \sigma^2_{\hat{\gamma}}(\tilde{\boldsymbol{\theta}}) \\ &= \left[\frac{\tilde{\lambda}\tilde{\tau}_2}{\gamma_0^3 \tilde{\mu}^2} \sum_{j=1}^{n_2} R_{y_{2j}-\frac{1}{2}}(\tilde{\theta}_2) - \frac{n_2 \tilde{\lambda}}{\gamma_0^2 \tilde{\mu}}\right]^2 \times \sigma^2_{\hat{\gamma}}(\tilde{\boldsymbol{\theta}}) \end{aligned}$$

where $\sigma^2_{\hat{\gamma}}(\tilde{\boldsymbol{\theta}})$ *is the asymptotic variance of* $\hat{\gamma}$ *(Lemma 5.3.5) evaluated at the MLEs under* H_0*.*

To compute $\sigma^2_{\hat{\gamma}}$, one needs to be able to compute the Fisher information matrix. Since this is not available in a closed form, the observed information evaluated at the MLEs under H_0 is used in its place. A major problem here is that the observed information can generate negative variance estimates which leads to negative score test statistics. Further it may produce an inconsistent test. Thus an RST using the observed information may not produce a valid chi-square test in this case. Freedman [2007] gives a detailed discussion on the anomalies of the score test when using the observed information matrix. Morgan et al. [2007] give an example involving a zero-inflated Poisson distribution where the score test statistic using the observed information is negative.

Theorem 5.4.3. *(WT) The WT statistic for testing* $H_0 : g(\gamma) = g(\gamma_0)$ *vs.* $H_1 : g(\gamma) \neq g(\gamma_0)$ *is given by*

$$WT = \left[\frac{g(\hat{\gamma}) - g(\gamma_0)}{g'(\hat{\gamma})\sigma_{\hat{\gamma}}}\right]^2$$

where $\sigma^2_{\hat{\gamma}}$ *is the asymptotic variance of* $\hat{\gamma}$ *given in Lemma 5.3.5.*

To evaluate $\sigma^2_{\hat{\gamma}}$, one needs to obtain the inverse of the observed information matrix (negative of the Hessian matrix) evaluated at the unrestricted MLEs. At the unrestricted MLE, the observed information will be positive definite ensuring consistency of the test.

We consider the following functions $g(\gamma) : \gamma, \log(\gamma), \sqrt{\gamma}$ and γ^2. Each of the above statistic is asymptotically distributed as a χ^2_1 rv under H_0 and an approximate level ν test rejects H_0 if the test statistic is $> \chi^2_1(1-\nu)$. This approximation does not always work well for small sample sizes as seen in previous chapters. Figure 5.1 shows the simulation based *exact* 95^{th} percentiles for the six asymptotic tests as a function of sample size n. These percentiles, especially for the Wald tests, can be much higher than c_0 even for sample sizes as high as 100 for each group.

We did a simulation study to evaluate further, the effect of changing sample size n, μ and λ on the simulated levels and *exact* percentiles. We computed the *exact* percentiles for these test statistics by simulating $M = 10,000$ datasets from the null distribution (under H_0) and then picking the 95^{th} percentile for a 5% test.[4] Four different sample sizes were considered: $n = 10, 20, 50, 100$ with μ ranging from 1 to 20 and $\lambda = 0.5, 1, 25, 10$. The simulated *exact* 95^{th} percentiles of the LRT and the WT(γ) statistics are given in Figures 5.2 and 5.3 respectively. Increasing μ, has no effect on the *exact* percentiles for the LRT . However, when μ increases the exact percentiles for WT(γ) move away from c_0. For both the LRT and the WT(γ), increasing α brings the exact percentiles closer to c_0. As expected, the exact percentiles converge to c_0 as the sample size n increases.

[4]Computing P-IG probabilities is very time consuming and hence we had to restrict ourselves to using $M = 10,000$. Increasing M to at least 100,000 will produce more precise estimates of the *exact* percentiles.

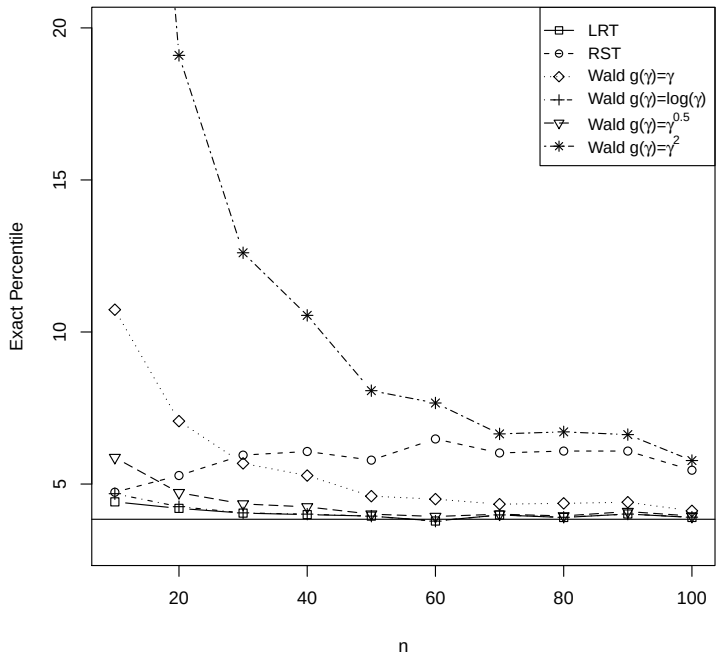

Figure 5.1: P-IG Model for PG Trial: Simulation based 95th percentile value for the null distribution of LRT, RST and WT statistics as a function of common sample size n; $\mu = 16.8$, $\lambda = 6.56$. The solid horizontal line refers to $c_0 = 3.8415 (= \chi^2_1(0.95))$.

The Type I error rates for the asymptotic LRT and WT(γ) (Figures 5.4 and 5.5) converge to the nominal level of 0.05 as n increases and also as λ increases for a fixed n. Changing μ does not seem to have an effect on the empirical error levels for the asymptotic LRT. However, increasing μ seems to increase the Type I error levels although not very significantly. Also the error rates for the asymptotic WT(γ) seem to be higher than those for the LRT. For the asymptotic LRT, the error rates are very close to the nominal level for sample sizes 50 and above per group for all values of μ and λ. For smaller sample sizes however the error rates can be much higher than

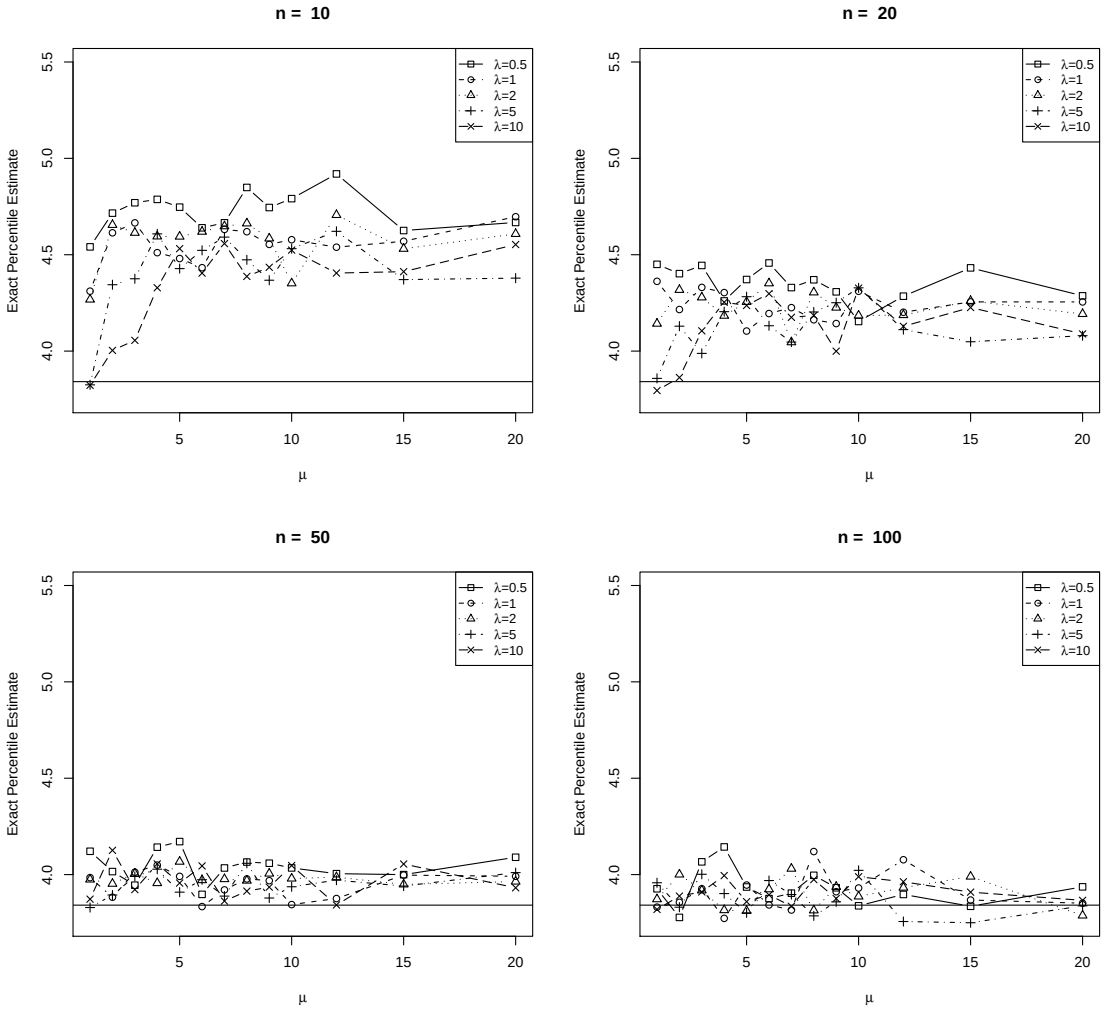

Figure 5.2: P-IG Model for PG Trial: Simulation based 95th percentile value for the null distribution of LRT statistic for different values of μ and λ and $n_1 = n_2 = 10, 20, 50, 100$, the solid horizontal line refers to $c_0 (= 3.8415 = \chi^2_1(0.95))$.

0.05 and hence the asymptotic LRT is not suggested. For the asymptotic WT(γ), the empirical Type I error rates are very high for all values of μ and γ for sample sizes less than or equal to 20 per group. Even for very large sample sizes per group ($n =$ 50, 100), the error rates can be very high for λ (≤ 5). Thus, we suggest the use of

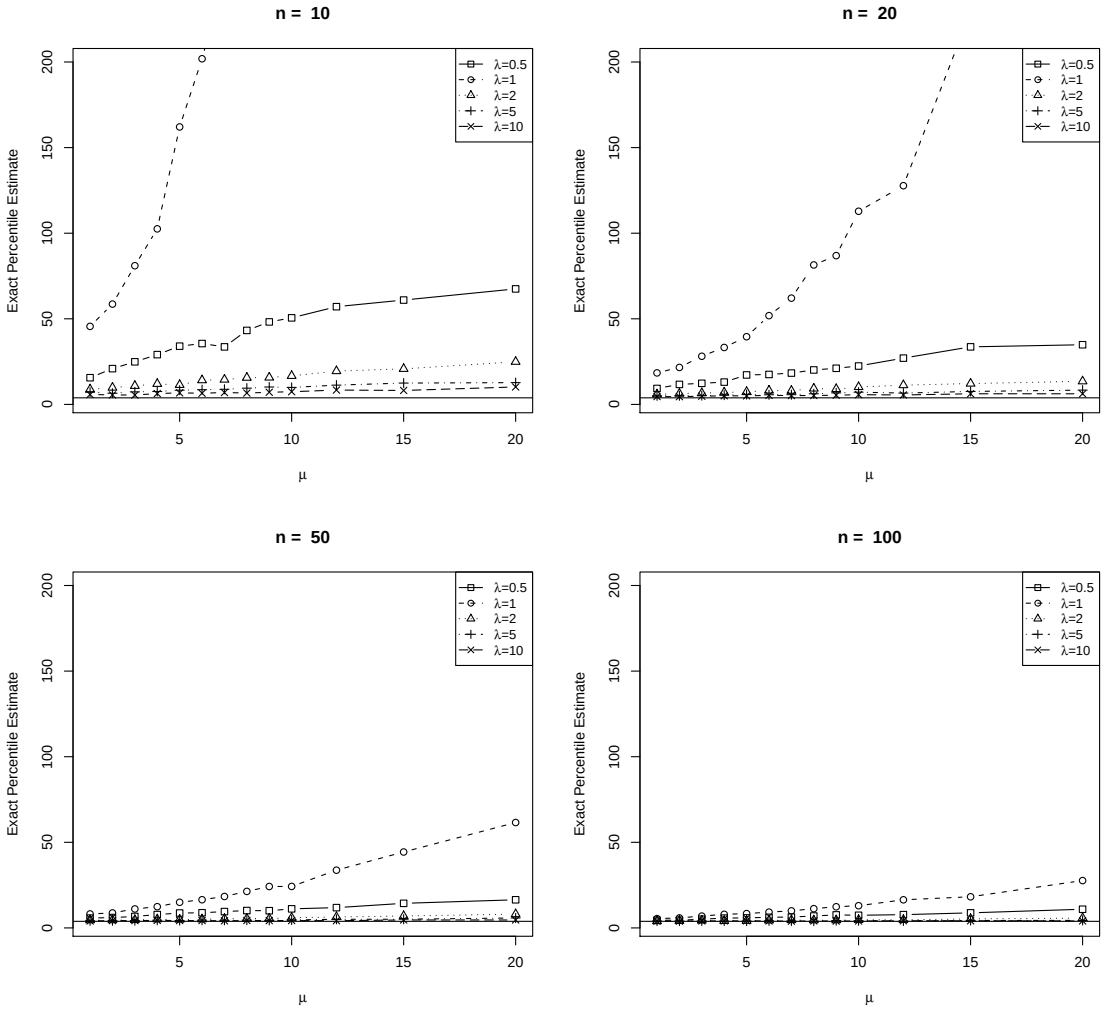

Figure 5.3: P-IG Model for PG Trial: Simulation based 95th percentile value for the null distribution of WT(γ) statistic for different values of μ and λ and $n = 10, 20, 50, 100$, the solid horizontal line refers to $c_0 (= 3.8415 = \chi^2_1(0.95))$.

the *exact* percentile based WT(γ) unless the sample size is 100 or above per group and λ is 10 or higher.

The exact percentile estimates and the Type I error rates RST and WT($\sqrt{\gamma}$) as they were not the 'best' tests as we will see in the next section (see Figure 5.7 of

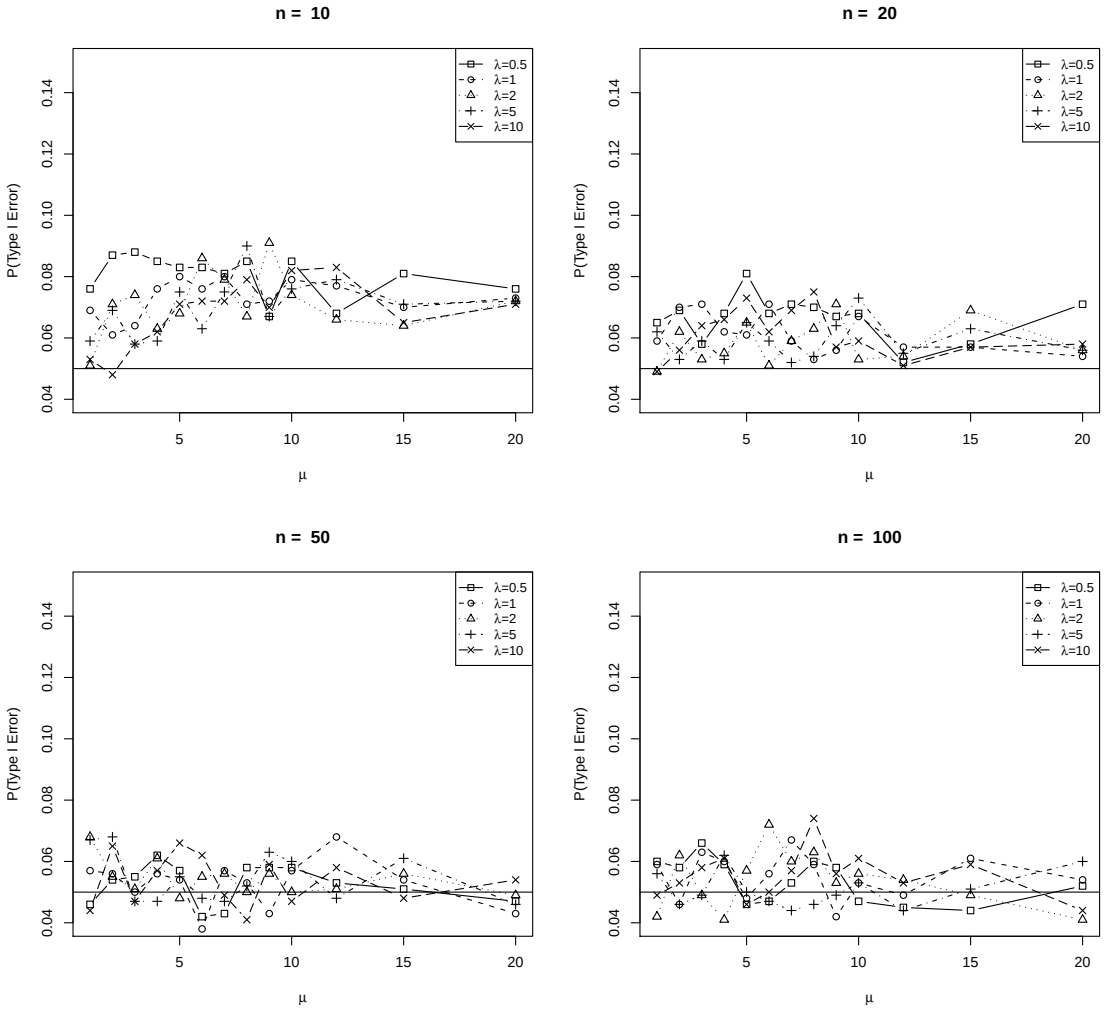

Figure 5.4: P-IG Model for PG Trial: Type I error rates for LRT with critical value c_0 for different values of μ and λ and $n = 10, 20, 50, 100$ subjects per group, the solid horizontal line refers to nominal level $\nu = 0.05$.

Section 5.5). The exact percentiles and the Type I error rates for WT(log(γ)) are very similar to those for the LRT (Figure not shown). For the WT(γ^2), the exact percentiles and simulated levels are much higher than those for $WT(\gamma)$ but the effect of n, μ and λ is very similar to that of $WT(\gamma)$ (Figure not shown).

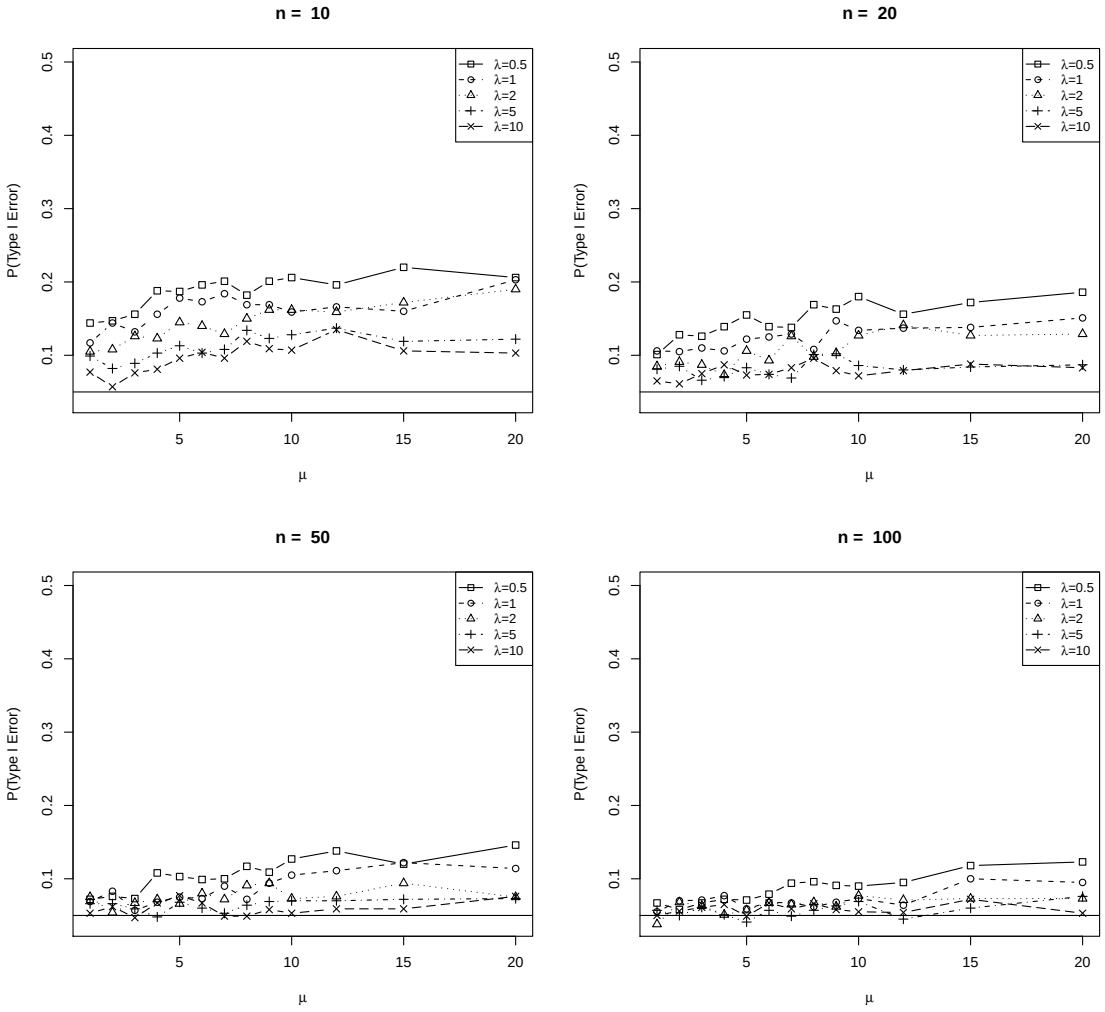

Figure 5.5: P-IG Model for PG Trial: Type I error rates for WT(γ) with critical value c_0 for different values of μ and α and $n = 10, 20, 50, 100$ subjects per group, the solid horizontal line refers to nominal level $\nu = 0.05$.

5.5 Power Analysis

We conducted a simulation study to compare the performance of the asymptotic and the *exact* percentile based LRT, RST and the Wald tests given in Section 5.4. The

exact percentiles for these test statistics are calculated as discussed in the previous section using $M = 10,000$. Equal sample sizes of 50 per group was considered. Nominal level was set to be $\nu = 0.05$. The power for each γ is then calculated as the proportion of trials for which the null hypothesis H_0 is rejected using either the asymptotic or the *exact* percentile based test. The initial parameter estimates for the simulation study were $\mu = 16.8$ and $\lambda = 6.56$. These were estimated by Sormani et al. [2001b] by fitting a P-IG distribution (by the method of maximum likelihood) to MRI lesion count data from RRMS patients who were selected for lesion activity at baseline and followed for a period of 6 months.

The power curves for the six asymptotic tests for the P-IG model are shown in Figure 5.6. The LRT and WT($\log(\gamma)$) are unbiased tests and have very similar power estimates for all values of γ. These two tests also maintain Type I error rates very well (close to the nominal level 0.05 under $H_0 : \gamma = 1$) for a sample size of 50 per group. The WT(γ^2) has the highest power for $\gamma < 1$ but also has the highest Type I error rate, much higher than the nominal 5%. The asymptotic WT(γ) has lesser power than the WT(γ^2) for $\gamma < 1$ and has slightly higher power for $\gamma > 1$. The simulated error rate for WT(γ) is still higher than the nominal 5%. The asymptotic WT($\sqrt{\gamma}$) is not considered further as it is not the most powerful test on either side of the null hypothesis.

Figure 5.7 shows the power curves for the six *exact* tests. One can see that all the tests maintain the Type I error rates very precisely. The LRT and the WT($\log(\gamma)$) are unbiased and have very similar power estimates. WT(γ) and WT(γ^2) have similar power estimates and have the highest power for $\gamma < 1$. For $\gamma > 1$, their power quickly goes to 0 and hence either test is best suited for RRMS clinical trials as discussed in

previous chapters. We therefore suggest the use of WT(γ) for RRMS clinical trials. The WT($\sqrt{\gamma}$) is neither unbiased nor has the highest power on either sides of the null hypothesis and is not considered further.

Thus if an unbiased test is preferred, either the LRT or WT(log(γ)) can be used. For these two tests, asymptotic approximation can be used for sample sizes 50 or above but for smaller sample sizes the *exact* percentile based tests are recommended. For RRMS clinical trials, we are interested in the region $\gamma < 1$ and hence the *exact* WT(γ) is recommended. We do not recommend the asymptotic WT(γ) even for very large sample sizes.

The RST statistic was computed using the observed information matrix evaluated at the MLEs under the null hypothesis. This sometimes led to the observed information matrix not being positive definite which resulted in a negative RST statistic and an inconsistent test. The power of the RST using the observerd information may also be non monotonic. Moving away from H_0 does not necessarily result in an increase in power (see Figures 5.6 and 5.7). Thus, the RST is not considered for the P-IG model for PG trials.

5.6 Sample Size Estimation

The LRT was the preferred unbiased test and the WT(γ) had the highest power for $\gamma < 1$ and was hence chosen as the best test for RRMS clinical trials even though it was not unbiased. In this section, we present sample size estimates for PG trials assuming the P-IG model for these two tests. We present sample sizes using both the asymptotic and the *exact* LRT and only for the *exact* WT(γ) as the asymptotic WT(γ) was very liberal. The *exact* percentile and the power estimates were calculated

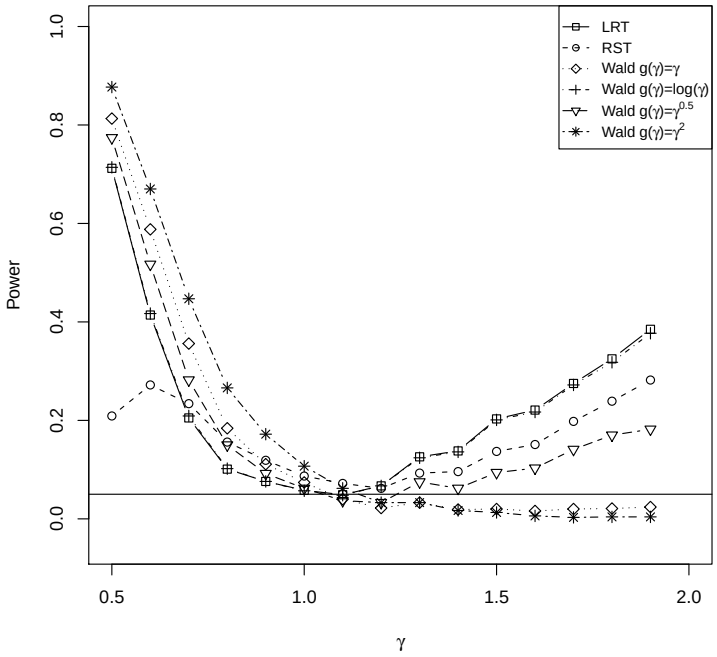

Figure 5.6: P-IG Model for PG Trial: Power of asymptotic 5% level LRT, RST, and WTs for treatment effect, assuming initial parameter estimates $\mu = 16.8, \lambda = 6.56$, sample sizes $n_1 = n_2 = 50$. The solid horizontal line refers to the nominal level $\nu = 0.05$.

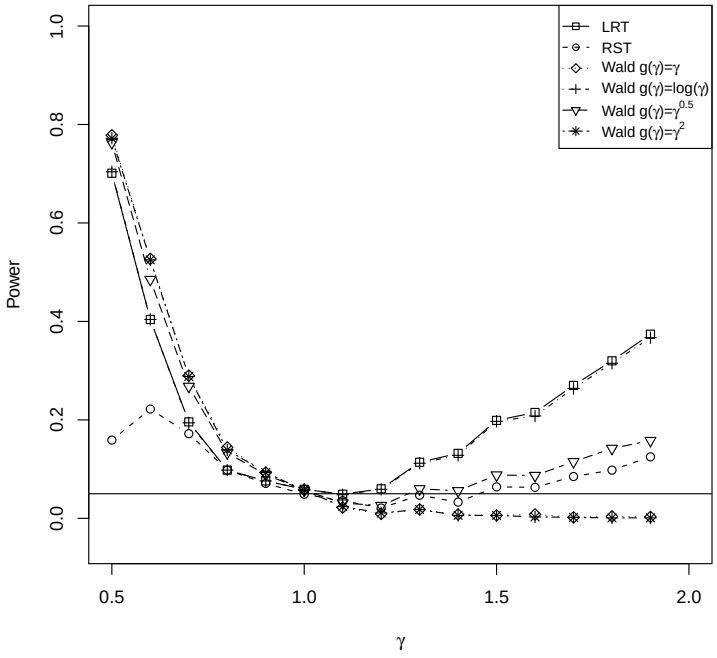

Figure 5.7: P-IG Model for PG Trial: Power of exact 5% level LRT, RST, and WTs for treatment effect, assuming initial parameter estimates $\mu = 16.8, \lambda = 6.56$, sample sizes $n_1 = n_2 = 50$. The solid horizontal line refers to the nominal level $\nu = 0.05$.

using the method described in Section 5.4. We also present the sample sizes obtained using the Wilcoxon Rank Sum test as well.

Sormani et al. [2001a] fit the P-IG model (using alternate parametrization μ and $\beta = \mu^2/\lambda$) to new MRI lesion counts in 115 RRMS patients selected for having at least 1 enhancing lesion on the baseline scan. They estimate, using the method of maximum likelihood, that $\hat{\mu} = 16.8$ and $\hat{\beta} = 43$ for a 6 month follow-up period. Thus, in our parametrization, we have $\hat{\mu} = 16.8$ and $\hat{\lambda} = 6.56$. We use these estimates for μ and λ as our initial estimates for the simulation study. Since parameter estimates are not available for a 3 month follow-up period, we do not present sample sizes for that case.

Sample sizes (Table 5.1) for each group are given for 80 and 90% power, follow-up period of 6 months and treatment effect $1 - \gamma$ ranging from 0.50 to 0.80. The LRT sample sizes are approximately 30-52% smaller than the WRS sample sizes. For example, for a 50% treatment effect and 80% power, WRS estimates a sample size of 134 per group where as the LRT (asymptotic or *exact*) estimates only 66 per group, a reduction of 51%. The *exact* LRT sample sizes are higher than the asymptotic LRT sample sizes by at most 1. This difference is seen only for higher values of treatment effects which yield smaller sample sizes. Sample sizes using the *exact* WT(γ) are 7-18% smaller than the sample sizes using LRT and around 40-57% smaller than the WRS sample sizes. Sample sizes to achieve 90% power are 20-40% higher in general than the sample sizes required for 80% power.

Table 5.1: P-IG Model for PG Trial: Sample sizes per group to achieve 80% and 90% power for $100(1-\gamma)\%$ treatment effect, baseline period of 1 month, and follow-up period (t) of 6 months, and initial estimates $(\mu, \lambda) = (16.8, 6.56)$; level $\nu = 0.05$.

$1-\gamma$	80% power; 5% level					
	Test				Critical Value	
	WRS	LRT		WT(γ)	LRT	WT(γ)
		Asymptotic	*Exact*	*Exact*	Simul.	Simul.
0.50	134	66	66	58	3.8402	3.9483
0.60	69	36	36	30	3.8846	4.0164
0.70	34	19	20	17	4.0961	4.3876
0.80	17	11	12	10	4.2762	4.7157
$1-\gamma$	90% power; 5% level					
	Test				Critical Value	
	WRS	LRT		WT(γ)	LRT	WT(γ)
		Asymptotic	*Exact*	*Exact*	Simul.	Simul.
0.50	179	87	87	78	3.9009	4.0004
0.60	88	48	49	40	3.8530	4.0088
0.70	45	26	26	24	4.0377	4.0382
0.80	23	13	13	12	4.2129	4.4614

5.7 Discussion

In this chapter we proposed likelihood based parametric tests such as LRT, RST and several Wald tests for the P-IG model for PG trials. We compared their performance using the asymptotic and *exact* percentiles and proposed sample size estimates based on appropriate tests. The initial parameter estimates used in the simulation study were obtained by fitting the P-IG model to MRI lesion count data in RRMS patients selected for baseline activity. Therefore, we suggest the sample sizes reported in this chapter be used only in clinical trials involving such a group of patients.

The recommendations are based on the properties of the test and the research hypothesis in question. We do not consider the RST due to the issue of getting negative test statistics. For RRMS clinical trials, though not unbiased, the *exact* WT(γ) for all sample sizes. When an unbiased test is desired irrespective of whether one is interested in $\gamma < 1$ or $\gamma > 1$, the LRT or the WT(log(γ)) is preferred. Asymptotic approximation (for the LRT and WT(log(γ)) can be used for sample sizes over 50 but for lesser sample sizes the *exact* test needs to be used to ensure the Type I error levels are close to nominal values. When compared to the WRS sample sizes, the ones obtained using LRT are around 30-50% smaller and the ones using WT(γ) are around 40-57% smaller. Thus, when the P-IG model is assumed, using the parametric tests proposed in this chapter provide a way of significantly reducing costs in RRMS clinical trials.

Several other areas of application of the P-IG model have been discussed in the literature. Willmot [1987] showed that the model provides an extremely good fit to automobile claim frequency data and also showed that the P-IG model fits better than the negative binomial model in most cases. Ord and Whitmore [1986] discuss

the P-IG distribution as a model for species abundance. Sankaran [1968] illustrates the applicability of the P-IG model to larvae counts on corn bean plants. In all these cases the methods developed in this chapter can be used.

CHAPTER 6

Conclusion and Further Work

Although the NB distribution has been shown to fit well to MRI lesion count data in RRMS clinical trials, only nonparametric tests are being used in current clinical trials to test for the treatment effect. The main focus of this thesis was to develop parametric tests for NB models with applications to RRMS clinical trials. Three important trial designs that are very common in MS trials were considered. In Chapter 2, we assumed the NB model for PG trials and proposed likelihood based parametric tests such as LRT, RST and WTs to test for the treatment effect. Using BNB models parametric tests for BVT trials and PGB trials using BNB models were proposed in Chapters 3 and 4, respectively. Likelihood based tests were also derived for PG trials assuming the P-IG model in Chapter 5. These tests were then compared using simulation studies and appropriate tests were chosen based on the properties of the test and the region of interest for each trial design and each parametric model. Sample size estimates for RRMS clinical trials were obtained using the chosen parametric tests and were compared to those obtained using nonparametric tests.

The choice of the test depends on the parametric model being considered, the property of the test itself and the research hypothesis in question. The LRT, RST

and WT($\log(\gamma)$) are empirically unbiased for all trial designs and parametric models considered. If unbiasedness of a test is a property that is strictly desired, usually one of the three tests depending on the region of interest can be used for PG, BVT and PGB trials that assume NB models. For the NB models in Chapters 2 to 4, the WT(γ^2), though not unbiased, had the highest power for the region $\gamma < 1$ and lowest power for $\gamma > 1$ and hence best suited for clinical trials in MS. For the P-IG model for PG trials in Chapter 5, the WT(γ) is the preferred test for RRMS clinical trials.

Asymptotic percentile based tests may be a poor approximation for smaller sample sizes and may result in a very liberal or a very conservative test. We developed *exact* percentile based tests that maintain Type I error rates even for very small sample sizes. For PG trials using the NB and the P-IG model, the asymptotic LRT can be used in general for sample sizes over 50 but for lesser sample sizes the *exact* LRT is preferred. For the BVT trials and the PGB trials assuming BNB models, the asymptotic LRT maintains Type I error rates reasonably well even for very small sample sizes and hence the *exact* LRT is not required. These recommendations hold true even for the other two unbiased tests RST and WT($\log(\gamma)$) when their use is appropriate. The asymptotic WT(γ^2) and the WT(γ) are not suggested even for very large sample sizes and the *exact* percentile based tests are suggested always for all trial designs and models considered.

When compared to the nonparametric tests, the preferred parametric tests provided significant reductions in sample sizes. For the NB models considered, the LRT provided a reduction of 30-45% for PG trials, 25-50% for BVT trials and 70-80% for PGB trials. The *exact* WT(γ^2) provided a further 10-20% reduction in sample sizes when compared to the LRT. For the P-IG model for PG trials, the LRT provided a

reduction of 30-50% in sample sizes when compared to the WRS test and the *exact* WT(γ) provided a further reduction of 7-16%.

The RST statistic using the observed information can sometimes be negative (Section 5.4). This is because the observed information matrix evaluated at the restricted MLEs may sometimes generate negative variance estimates. For the NB models in Chapters 2 to 4, closed form expressions for the Fisher information were available and hence observed information was not required. For the P-IG model for PG trials however, the Fisher information was not available in closed form and hence the observed information was used. This resulted in negative RST statistics and hence the RST using the observed information is not suited for this model. The same issue arose when the observed information was used for the PG trials assuming NB model. Thus one has to be look out for such issues when using the observed information for the RST.

We have developed likelihood based parametric tests for the treatment effect for PG, BVT and PGB trials assuming NB models and for PG trials assuming a P-IG model. For the P-IG model for PG trials we have only considered the case of equal λ for the two groups. As done for the NB model in Chapter 2, the parametric tests developed in this chapter can be extended to case of unequal λ as well. A robustness study needs to be done to study the performance of these tests for deviations from distributional assumptions and also for deviations from the equal λ assumption.

Similar methods can be developed for BVT and PGB trials assuming a bivariate P-IG (BPIG) model. Also, since cross-over trials are also commonly used in RRMS clinical trials, one can develop similar tests for such a design assuming NB and P-IG distributions. One can develop similar tests to test for the treatment effect in the

presence of covariates. For a NB model for example, a generalized linear model can be used to adjust for covariates. MLEs of the parameters can be obtained using a Fisher scoring algorithm and similar tests for the treatment effect can be developed in those cases as well.

BIBLIOGRAPHY

I. Aban, G. R. Cutter, and N. Mavinga. Inferences and power analysis concering two negative binomial distributions with an application to MRI lesion counts data. *Computational Statistics and Data Analysis*, 53:820–833, 2009.

M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions*. Dover Publications, Inc., New York, 1970.

A. G. Arbous and J. E. Kerrich. Accident statistics and the concept of accident proneness. *Biometrics*, 7:643–656, 1951.

G. E. Bates and J. Neyman. Contributions to the theory of accident proneness, I: An optimistic model of correlation between light and severe accidents. *University of California Publications in Statistics*, 1:215–254, 1952.

J. Cohen. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Inc., Publishers, 2nd edition, 1988.

F. Downton. Bivariate exponential distributions in reliability theory. *Journal of the Royal Statistical Society, B*, 32:408–417, 1970.

C. B. Edwards and J. Gurland. A class of distributions applicable to accidents. *Journal of the American Statistical Association*, 56:503–517, 1961.

D. A Freedman. How can the score test be inconsistent? *The American Statistician*, 61(4):291–295, 2007.

M. Greenwood and G. U. Yule. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attack of disease or of repeated accidents. *Journal of the Royal Statistical Society*, 83:255–279, 1920.

A. Guldberg. On discontinuous frequency distributions of two variables. *Scandinavisk Aktuarietidskrift*, 17:89–117, 1934.

M. S. Holla. Canonical expansion of the compounded correlated bivariate poisson distribution. *The American Statistician*, 23:32–33, 1971.

N. L. Johnson, A. W. Kemp, and S. Kotz. *Univariate Discrete Distributions*, chapter 5. John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.

J. F. Kurtzke. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*, 33:1444–1452, 1983.

E. L. Lehmann. *Testing Statistical Hypotheses, Second Edition.* Wadsworth & Brooks/Cole, Pacific Grove, CA., 1986.

B. J. T. Morgan, K. J. Palmer, and M. S. Ridout. Score test oddities: Negative score test statistic. *The American Statistician*, 61(4):285–288, 2007.

J. J. P. Nauta, A. J. Thompson, F. Barkhof, and D. H. Miller. Magnetic resonance imaging in monitoring the treatment of multiple sclerosis patients: statistical power of parallel-groups and crossover designs. *Journal of Neurological Sciences*, 122:6–14, 1994.

J. K. Ord and K. A. Whitmore. The poisson-inverese gaussian distribution as a model for species abundance. *Communications in Statistics - Theory and Methods*, 15(3): 853–871, 1986.

G. P. Patil. Certain properties of the generalized power series distribution. *Annals of Institute of Statistical Mathematics Tokyo*, 14:179–182, 1962.

C. R. Rao. Large sample tests of statistical hypotheses concerning several parameters with application to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, 44:50–57, 1948.

C. R. Rao. *Advances in Ranking and Selection, Multiple Comparisons, and Reliability*, chapter Score Test: Historical Review and Recent Developments, pages 3–20. Birkhäuser, Boston, 2005.

M Sankaran. Mixtures by the inverse gaussian distribution. *Sankhya B*, 30:455–458, 1968.

H. S. Sichel. On a family of discrete distributions particularly suited to represent long-tailed frequency data. In South Africa: Council for Scientific N. F. Laubscher, Pretoria and Industrial Research, editors, *Proceedings of the Third Symposium on Mathematical Statistics*, pages 51–97, 1971.

M. P. Sormani, P. Bruzzi, D. H. Miller, C. Gasperini, F. Barkhof, and M. Filippi. Modelling MRI enhancing lesion counts in multiple sclerosis using a negative binomial model: implications for clinical trials. *Journal of the Neurological Sciences*, 163:74–80, 1999.

M. P. Sormani, P. Bruzzi, M. Rovaris, F. Barkhof, G. Comi, D. H. Miller, G. R. Cutter, and M. Filippi. Modelling new enhancing MRI lesion counts in multiple sclerosis. *Multiple Sclerosis*, 7:298–304, 2001a.

M. P. Sormani, D. H. Miller, G. Comi, F. Barkhof, M. Rovaris, P. Bruzzi, and M. Filippi. Clinical trials of multiple sclerosis monitored with enhanced MRI: new sample size calculations based on large data sets. *Journal of Neurology Neurosurgery and Psychiatry*, 70:494–499, 2001b.

G. Z. Stein, W. Zucchini, and J. M. Juritz. Parameter estimation for the sichel distribution and its multivariate distribution. *Journal of the American Statistical Association*, 82(399):938–944, 1987.

K. Subramaniam. A test for 'intrinsic correlation' in the theory of accident proneness. *Journal of the Royal Statistical Society, B*, 28:180–189, 1966.

K. Subramaniam and K. Subramaniam. *Bivariate Discrete Distributions*, chapter 5. Marcel Dekker, Inc, 1992.

P. F. Thall and S. C. Vail. Some covariate models for longitudinal count data with overdispersion. *Biometrics*, 46:657–671, 1990.

N. Tubridy, H. J. Ader, F. Barkhof, A. J. Thompson, and D. H. Miller. Exploratory treatment trials in multiple sclerosis using mri: sample size calculations for relapsing-remitting and secondary progressive subgroups using placebo controlled parallel groups. *Journal of Neurology Neurosurgery and Psychiatry*, 64: 50–55, 1998.

M. C. K. Tweedie. Statistical properties of inverse gaussian distributions. i. *Annals of Mathematical Statistics*, 28:362–377, 1957.

A. Wald. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54:426–482, 1943.

Y. Wang, L. J. Young, and D. E. Johnson. A UMPU test for comparing means for two negative binomial distributions. *Communications in Statistics - Simulation and Computation*, 30:1053–1075, 2001.

G. Willmot. The poisson-inverse gaussian distribution as an alternative to the negative binomial. *Scandinavian Actuarial Journal*, 87:113–127, 1987.

Appendices

APPENDIX A

Notations and Abbreviations

A.1 Abbreviations

BNB	Bivariate negative binomial
BVT	Baseline vs. Treatment
EDSS	Expanded Disability Status Scale
FIM	Fisher Information Matrix
LRT	Likelihood ratio test
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
NB	Negative binomial
PG	Parallel Group
PGB	Parallel Group with baseline correction
P-IG	Poisson-Inverse Gaussian
PPMS	Primary Progressive Multiple Sclerosis
PRMS	Progressive Relapsing Multiple Sclerosis
RRMS	Relapsing Remitting Multiple Sclerosis
RST	Rao's score test
SPMS	Secondary Progressive Multiple Sclerosis
WRS	Wilcoxon Rank Sum test
WSR	Wilcoxon Signed Rank test
WT	Wald test
UMPU	Uniformly Most Powerful Unbiased

A.2 Symbols

L	Likelihood function given the data
ℓ	$\log(L)$
$P_X(x)$	Probability mass function (pmf) of a discrete random variable X
$E(X)$	Expected value of a random variable X
$f_X(x)$	Probability density function (pdf) of a continuous random variable X
$G(s)$	Probability generating function of a discrete random variable
$I(\boldsymbol{\theta})$	Fisher information of parameter vector $\boldsymbol{\theta}$
$M(t)$	Moment generating function
ν	Significance level
Θ	Unrestricted parameter space
Θ_0	Restricted parameter space

A.3 Distributions

Gamma (α, β)	Gamma distribution with pdf $f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, \quad x > 0.$
Poisson(λ)	Poisson distribution with pmf $P_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots.$
NB(μ, α)	Negative binomial distribution with mean μ and dispersion parameter α and with pmf $P_X(x) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\mu^x}{x!} \frac{\Gamma(x+\alpha)}{(\mu+\alpha)^{x+\alpha}}, \quad x = 0, 1, 2, \ldots.$
BNB(μ_1, μ_2, α)	Bivariate negative binomial distribution with pmf $P_{X,Y}(x, y) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\mu_1^x}{x!} \frac{\mu_2^y}{y!} \frac{\Gamma(x+y+\alpha)}{(\mu_1+\mu_2+\alpha)^{x+y+\alpha}}, \quad x, y = 0, 1, 2, \ldots.$
IG(μ, λ)	Inverse Gaussian distribution with pdf $f_X(x) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}, \quad x > 0.$
P-IG(μ, λ)	Poisson-Inverse Gaussian distribution with mean μ and scale parameter λ. with pmf $P_X(x) = \frac{\tau^x}{x!} \left(\frac{2\theta}{\pi}\right)^{\frac{1}{2}} \exp\left(\frac{\lambda}{\mu}\right) K_{x-\frac{1}{2}}(\theta), \qquad y = 0, 1, \ldots,$ with $\tau = \left[\frac{1}{\mu^2} + \frac{2}{\lambda}\right]^{-\frac{1}{2}}$ and $\theta = \frac{\lambda}{\tau}$.

APPENDIX B

Simulation Method

The parameters of the NB distribution can be estimated from an already existing dataset. These estimated parameters $\hat{\mu}$ and $\hat{\alpha}$ can then be used to simulate the data. Here are the steps to simulate data using the NB model for PG trials and BNB model for BVT trials. The code to do this in R is given in Appendix Chapter D.

B.1 PG Trial

Suppose that there are n_1 patients in the placebo group and n_2 patients in the treatment group.

If the underlying distribution is NB, the placebo group is generated by drawing n_1 observations at random from NB(μ, α). The n_2 observations in the treatment group are randomly generated from NB$(\gamma\mu, \alpha)$. For the unequal dispersion parameter case the observations in the placebo group are randomly generated from NB(μ, α_1) and the observations in the treatment group are generated from NB$(\gamma\mu, \alpha_2)$.

If the underlying distribution is P-IG distribution, the data are generated by simply drawing n_1 random observations from P-IG(μ, λ) and n_2 random obsevations from P-IG$(\gamma\mu, \lambda)$.

B.2 BVT Trial - BNB model

Random observations $(x_i, y_i),\ i = 1, \ldots, n$ from the BNB(μ_1, μ_2, α) distribution can be generated by following the steps below: For each i,

1. z_i is randomly sampled from Gamma(α, α^{-1}).

2. x_i is generated by randomly sampling one observation from Poisson$(\mu_1 z_i)$.

3. y_i is generated by randomly sampling from Poisson$(\mu_2 z_i)$.

To generate data for BVT trials assuming the BNB model in Section 3.2, n random observations from a BNB$(\mu, \gamma\mu, \alpha)$ can be generated using the above method.

B.3 PGB Trial - BNB model

To generate data from PGB trials assuming the BNB model in Section 4.2, we need to generate n_1 random observations from BNB$(\mu, t\mu, \alpha)$ and n_2 random observations from BNB$(\mu, t\gamma\mu, \alpha)$. These can be generated using the method described in Section B.2.

APPENDIX C

Datasets

In this chapter we look at some of the datasets from the literature for which the NB and BNB fit well. Pearson's χ^2 goodness-of-fit (GOF) statistics along with the degrees of freedom (df) and P value are given.

C.1 MRI lesion counts

C.1.1 Nauta

Table C.1 (Nauta et al. [1994]) gives the number of MRI active lesion counts seen each month for 6 months from 23 unselected patients who remained untreated during the course of the study. Active lesions were defined as either new enhancing lesions, new non-enhancing lesions or enlarging non-enhancing lesions. We assumed the number of lesions seen during the first month as the baseline value (x) and the total number of lesions seen during the other 5 months (y) as the follow-up value. The BNB distribution (1.3.2) was fit to this data using ML method (χ^2 GOF statistic $= 4.68$, df $= 4$, $P = 0.32$). The estimates of the parameters μ_1, μ_2 and α are 0.91, 6.39 and 0.65 respectively.

Patient	Month of Scanning						Total
	1	2	3	4	5	6	
1	0	1	0	0	0	0	1
2	5	10	2	2	1	2	22
3	0	3	2	2	1	1	9
4	2	1	1	0	3	1	8
5	0	0	0	0	0	0	0
6	1	0	0	1	0	0	2
7	0	0	0	0	0	0	0
8	0	0	0	0	0	1	1
9	4	0	0	0	0	0	4
10	3	0	3	0	2	0	8
11	4	0	4	1	0	1	10
12	0	0	0	0	0	0	0
13	0	1	0	0	0	1	2
14	0	0	0	0	0	1	1
15	0	0	1	0	0	1	2
16	1	0	0	0	0	0	1
17	1	2	0	0	0	0	3
18	0	4	0	0	1	0	5
19	0	0	0	0	1	1	2
20	0	1	3	0	2	2	8
21	0	0	0	2	2	1	5
22	0	3	3	2	0	3	11
23	0	16	9	13	10	15	63

Table C.1: MRI lesions (Nauta): Number of active lesions seen in 6 monthly follow-up MRI scans for 23 RRMS and SPMS patients.

C.1.2 Tubridy

Table C.2 (taken from Tubridy et al. [1998]) gives the number of new active lesions seen in 6 monthly MRI scans for 31 RRMS patients. New active lesions are defined as

in Appendix Section C.1.1. We fit the BNB distribution given in (1.3.2) to the joint data on the total number of new active lesions seen during the first 3 monthly scans x and the total number of new active lesions seen during the last 3 monthly scans y. The χ^2 GOF statistic was 13.31 on 7 degrees of freedom with $P = 0.065$ suggesting the appropriateness of the BNB distribution.

C.1.3 Sormani

Sormani et al. [2001a] give the data on total MRI lesion counts seen during 6 monthly scans of the brain in 66 RRMS patients not selected for the presence of MRI activity on the baseline scan (Group A) and 115 RRMS patients selected for having at least one enhancing lesion seen during the baseline scan (Group B). Observed and expected frequency counts obtained by fitting the NB (2.1.1) and P-IG model (1.3.3) to the data are given in Table C.3. They estimate the parameters of the NB model to be $(\hat{\mu}, \hat{\alpha}) = (13, 0.52)$ for Group A and $(16.8, 0.74)$ for Group B. The P-IG model parameters were estimated as $(\hat{\mu}, \hat{\lambda}) = (13, 2.96)$ for Group A and $(16.8, 6.56)$ for Group B. The NB model (χ^2 GOF statistic = 1.21, df = 2, $P = 0.55$) fit better than the P-IG model (χ^2 GOF statistic = 5.25, df = 2, $P = 0.07$) for Group A. The P-IG model (χ^2 GOF statistic = 5.08, df = 2, $P = 0.08$) provided a more satisfactory fit than the NB model (χ^2 GOF statistic = 18.7, df = 2, $P < 0.0001$) for Group B.

C.2 Epileptic Seizures

The data given below arise from a clinical trial involving 59 epileptic patients (Thall and Vail [1990]). Patients suffering from simple or complex partial seizures were randomized to receive either the antiepileptic drug progabide or a placebo, as an

adjuvant to standard chemotherapy. At each of four successive post randomization clinic visits, the number of seizures occurring over the previous 2 weeks was reported. Baseline values (x) are the number of seizures seen in each patient during 8 weeks of the baseline period. Let y be the total number of seizures seen in the follow-up period. The BNB distribution with pmf (1.3.2) was fit to the observations in the placebo and treatment group.

The parameters were estimated using the method of maximum likelihood (ML). For the placebo group, μ_1, μ_2 and α were estimated to be 27.638, 34.386 and 1.951 respectively. For the treatment group, the parameters μ_1, μ_2 and α were estimated to be 31.606, 31.7658 and 1.5 respectively. The BNB distribution provided a good fit to the observations in the placebo group (χ^2 GOF statistic 14.54, df=10, P = 0.15) but not for the treatment group (χ^2 GOF statistic 29.69, df=12, $P = 0.003$).

Patient	Month of Scanning						x	y	Total
	1	2	3	4	5	6			
1	0	0	0	0	0	0	0	0	0
2	0	4	3	5	10	17	7	32	39
3	0	0	2	0	13	6	2	19	21
4	2	1	2	7	4	4	5	15	20
5	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0
7	1	1	0	0	1	0	2	1	3
8	0	0	0	0	0	0	0	0	0
9	4	2	2	0	0	1	8	1	9
10	0	1	1	3	1	0	2	4	6
11	6	4	4	7	9	7	14	23	37
12	0	0	0	0	0	0	0	0	0
13	0	0	0	1	0	0	0	1	1
14	2	1	0	0	0	0	3	0	3
15	2	2	1	2	4	1	5	7	12
16	4	1	4	0	0	0	9	0	9
17	1	0	0	0	1	0	1	1	2
18	0	0	0	0	0	0	0	0	0
19	2	0	0	0	0	0	2	0	2
20	0	2	2	2	0	0	4	2	6
21	0	5	1	1	1	0	6	2	8
22	1	0	7	5	11	12	8	28	36
23	0	0	1	2	0	0	1	2	3
24	1	1	0	1	0	1	2	2	4
25	4	0	1	0	2	0	5	2	7
26	2	0	0	0	0	0	2	0	2
27	11	5	3	2	1	0	19	3	22
28	1	1	1	0	0	0	3	0	3
29	3	0	0	1	1	2	3	4	7
30	2	0	0	0	1	1	2	2	4
31	1	1	1	3	0	1	3	4	7

Table C.2: MRI lesions (Tubridy): Number of new active lesions seen in 6 monthly follow-up MRI scans for 31 patients.

	Group A			Group B		
Lesion Counts	Observed	NB	P-IG	Observed	NB	P-IG
0	11	12.12	7.19	8	20.62	4.41
1	6	6.06	8.71	8	9.62	7.97
2-4	13	11.04	17.77	18	17.19	25.96
5-7	10	7.05	8.79	19	11.01	18.17
8-10	4	5.17	5.10	14	8.22	12.00
11-15	6	6.23	5.03	16	10.18	12.65
16-20	5	4.38	3.02	5	7.46	7.87
21-30	1	5.59	3.48	9	10.17	9.19
31-40	5	3.21	1.95	5	6.44	5.11
41-50	2	1.93	1.23	1	4.26	3.19
51-60	1	1.18	0.84	3	2.90	2.13
61-70	0	0.74	0.44	3	2.00	1.49
71-80	1	0.47	0.00	2	1.40	1.08
81-90	0	0.30	0.00	3	0.99	0.80
91+	1	0.19	0.00	1	0.71	0.61

Table C.3: MRI lesions (Sormani): Observed and expected frequency (NB and P-IG) of lesion counts for two groups of patients. Group A: 66 RRMS patients not selected for the presence of MRI activity on the baseline scan. Group B: 115 RRMS patients selected for having at least one Gd-enhancing lesion on the baseline scan.

Patient ID	Baseline	$y1$	$y2$	$y3$	$y4$
104	11	5	3	3	3
106	11	3	5	3	3
107	6	2	4	0	5
114	8	4	4	1	4
116	66	7	18	9	21
118	27	5	2	8	7
123	12	6	4	0	2
126	52	40	20	23	12
130	23	5	6	6	5
135	10	14	13	6	0
141	52	26	12	6	22
145	33	12	6	8	4
201	18	4	4	6	2
202	111	7	9	12	14
205	18	16	24	10	9
206	20	11	0	0	5
210	12	0	0	3	3
213	9	37	29	28	29
215	17	3	5	2	5
217	28	3	0	6	7
219	55	3	4	3	4
220	9	3	4	3	4
222	17	2	3	3	5
226	28	8	12	2	8
227	55	18	24	76	25
230	9	2	1	2	1
234	10	3	1	4	2
238	47	13	15	13	12

Table C.4: Epileptic Seizures: Bi-weekly seizure counts for 28 patients in the placebo group; baseline counts are observed over an 8 week period.

Patient ID	Baseline	$y1$	$y2$	$y3$	$y4$
101	76	11	14	9	8
102	38	8	7	9	4
103	19	0	4	3	0
108	10	3	6	1	3
110	19	2	6	7	4
111	24	4	3	1	3
112	31	22	17	19	16
113	14	5	4	7	4
117	11	2	4	0	4
121	67	3	7	7	7
122	41	4	18	2	5
124	7	2	1	1	0
128	22	0	2	4	0
129	13	5	4	0	3
137	46	11	14	25	15
139	36	10	5	3	8
143	38	19	7	6	7
147	7	1	1	2	3
203	36	6	10	8	8
204	11	2	1	0	0
207	151	102	65	72	63
208	22	4	3	2	4
209	41	8	6	5	7
211	32	1	3	1	5
214	56	18	11	28	13
218	24	6	3	4	0
221	16	3	5	4	3
225	22	1	21	19	8
228	25	2	3	0	1
232	13	0	0	0	0
236	12	1	4	3	2

Table C.5: Epileptic Seizures: Bi-weekly seizure counts for 31 patients in the treatment group.

APPENDIX D

R Code

The following is the R code used for the simulations in this thesis. Only the code for the NB model for PG trials (Chapter 2) is given. The code for the other chapters are similar and hence not provided for brevity.

D.1 PG Trial - NB Model

```
## MLE under the restricted hypothesis
nbsimMLEnull<-function(init=3,n1,n2,y1=y1,y2=y2,maxit=maxit,method=method)
{
    muhat<-(sum(y1)+sum(y2))/(n1+n2);
    loglikenull <- function(alpha)
    {
        if (alpha>0)
        {
            num<-sum(log(dnbinom(y1,mu=muhat,size=alpha)))
            + sum(log(dnbinom(y2,mu=muhat,size=alpha)));
            return(-num);
        }
        else
        {
            return(1e+08);
        }
    }
    mlenull<-optim(par=init,fn=loglikenull,method=method,
    control=list(maxit=maxit));
    return(c(-mlenull$value,muhat,mlenull$par));
}

## MLE under the unrestricted hypothesis
```

```
nbsimMLEalt<-function(init,n1,n2,y1=y1,y2=y2)
{
    mu1hat<-mean(y1);mu2hat<-mean(y2);
    #loglikealt <- function(par)
    loglikealt <- function(alpha)
    {
        if (alpha>0)
        {
            den<- sum(log(dnbinom(y1,mu=mu1hat,size=alpha)))
            + sum(log(dnbinom(y2,mu=mu2hat,size=alpha)));
            return(-den);
        }
        else
        {
            return(1e+08);
        }
    }
mlealt<-optim(par=init,fn=loglikealt, method="Nelder-Mead");
return(c(-mlealt$value,mu1hat,mu2hat,mlealt$par));
}

## First order score vector equations
firstorder<-function(y1,y2,mu1,mu2,alpha)
{
    n1<-length(y1);n2<-length(y2);
    domu1<- sum(y1)/mu1 - (sum(y1)+ n1*alpha)/(mu1+alpha);
    domu2<- sum(y2)/mu2 - (sum(y2)+ n2*alpha)/(mu2+alpha);
    doalpha<- (n1+n2)*(1+log(alpha)-digamma(alpha))
    - (sum(y1)+ n1*alpha)/(mu1+alpha) - (sum(y2)+ n2*alpha)/(mu2+alpha)
    - n1*log(mu1+alpha) - n2*log(mu2+alpha)
    + sum(digamma(y1+alpha)) + sum(digamma(y2+alpha));
    return(matrix(c(domu1,domu2,doalpha),ncol=1));
}

## 04/28/2009
## Matrix of the second order partial derivatives
## 10/22/2009 - Updated to expected second order derivatives
secondorder<-function(y1,y2,mu1,mu2,alpha)
{
    n1<-length(y1);n2<-length(y2);
    domu1.2 <- -n1*alpha/(mu1*(mu1+alpha));
    domu1mu2<- 0;
    domu1alpha<-0;
    domu2mu1<-domu1mu2;
```

```
    domu2.2<- -n2*alpha/(mu2*(mu2+alpha));
    domu2alpha<-0;
    doalphamu1<-domu1alpha;
    doalphamu2<-domu2alpha;
    doalpha.2 <- (n1+n2)*(1/alpha - trigamma(alpha)) - n1/(mu1+alpha)
    - n2/(mu2+alpha) + sum(trigamma(y1+alpha)) + sum(trigamma(y2+alpha));
    return(matrix(c(domu1.2,domu1mu2,domu1alpha,domu2mu1,domu2.2,domu2alpha,
    doalphamu1,doalphamu2,doalpha.2),nrow=3,byrow=T));
}

## Function to compute the power
nbpower<-function (n1,n2,mu1,alpha,alpha1=0,delta=0.50,B=1000,M=1000,
lrt.cutoff,rst.cutoff,wal1.cutoff,wal2.cutoff,wal3.cutoff,wal4.cutoff,
cutoff.sim=1,method="Nelder-Mead",maxit=500,nu=0.05)
{
    cat("mu = ",mu1,"Alpha = ",alpha,"Alpha1 = ",alpha1,"Gamma = ",(1-delta),
    "n1=",n1,"n2=",n2,sep=" ","\n");
    lrt.power.chisq<-0;lrt.power.exact<-0;rst.power.chisq<-0;
    rst.power.exact<-wal.power.chisq<-wal.power.exact<-0;

    ## variables for the Other Wald Chisq and Exact tests
    wal1.power.chisq<-wal2.power.chisq<-wal3.power.chisq<-wal4.power.chisq
    <-wal5.power.chisq<-0;
    wal1.power.exact<-wal2.power.exact<-wal3.power.exact<-wal4.power.exact
    <-wal5.power.exact<-0;

    ## To calculate the exact p-value, the quantile is obtained by simulating
    ## from the null distribution and calculating the LRT statistic

    lrt<-null<-alt<-rst<-wal<-wal1<-wal2<-wal3<-wal4<-wal5<-array(dim=M);

    ## Toggle variable to get the exact power or not for all tests
    toggle.lrt<-toggle.rst<-toggle.wal1<-toggle.wal2<-toggle.wal3<-
    toggle.wal4<-0;

    ## Chisq cutoff value
    cutoff.chisq<-qchisq(nu,1,lower.tail=F);
    ## Updated 12/04/2009
    ## Get simulated cutoff values only if asked to; if cutoff.sim=1,
    ## then get simulated cutoff
    ## Otherwise calculate using specified cutoff values
    if (cutoff.sim==1)
    {
        for (i in 1:M)
        {
```

```
y1 <- rnbinom(n=n1, mu=mu1,size = alpha);
## 11/26/2009 To allow for generating data from NB populations
## with different dispersion parameters
if (alpha1==0)
{
y2 <- rnbinom(n=n2, mu=mu1,size = alpha);
}
else
{
y2 <- rnbinom(n=n2, mu=mu1,size = alpha1);
}
y1bar<-mean(y1);y2bar<-mean(y2);

while (y2bar==0)
{
y1 <- rnbinom(n=n1, mu=mu1,size = alpha);
if (alpha1==0)
{
y2 <- rnbinom(n=n2, mu=mu1,size = alpha);
}
else
{
y2 <- rnbinom(n=n2, mu=mu1,size = alpha1);
}
y2bar=mean(y2);
}

## Likelihood Ratio Test
null.0<-nbsimMLEnull(init=alpha,n1=n1,n2=n2,y1,y2,maxit=maxit,
method=method);
alt.0<-nbsimMLEalt(init=alpha,n1=n2,n2=n2,y1,y2);
null[i]<-null.0[1];alt[i]<-alt.0[1];
lrt[i]<- -2* (null[i]-alt[i]);

## Rao's Score Test: Need MLEs of mu and alpha in the restricted case
mu.0<-null.0[2];alpha.0<-null.0[3];
first<-firstorder(y1,y2,mu.0,mu.0,alpha.0);
hessian_score<-secondorder(y1,y2,mu.0,mu.0,alpha.0);
fisherinf.inv_score<-diag(3);
fisherinf.inv_score[1,1]<- -1/hessian_score[1,1];
fisherinf.inv_score[2,2]<- -1/hessian_score[2,2];
fisherinf.inv_score[3,3]<- -1/hessian_score[3,3];
rst[i]<- t(first)%*%fisherinf.inv_score%*%first;
```

```
## Wald Test: Need MLEs of parameters under the unrestricted case
## MLE of mu1, mu2 and alpha in the unrestricted case
mu1.1<-alt.0[2];mu2.1<-alt.0[3];alpha.1<-alt.0[4];

## Other Wald Tests constructed using the delta method for variance

## Aban et. al Wald Statistic: (using  \hat{gamma})
gamma.hat<-mu2.1/mu1.1;gamma.0<-1;
sigma2.gamma.hat.1<-(gamma.hat*(n1*(alpha.1+gamma.hat*mu1.1)
+n2*gamma.hat*(alpha.1+mu1.1)))/(n1*n2*alpha.1*mu1.1);
wal1[i]<-(gamma.hat-gamma.0)^2/sigma2.gamma.hat.1;

## Wald Statistic: (using log transformation of gamma hat)
## g(gamma.hat) = log(gamma.hat); So g'(gamma.hat) = 1/gamma.hat
sigma2.gamma.hat.2<-(1/gamma.hat)^2*sigma2.gamma.hat.1;
wal2[i]<-(log(gamma.hat)-log(gamma.0))^2/((1/gamma.hat)
^2*sigma2.gamma.hat.1);

## Wald Statistic: using square root transformation;
## g(gamma.hat) = (gamma.hat)^(0.5); g'(gamma.hat) =
## 1/(2*sqrt(gamma.hat))
sigma2.gamma.hat.3<- (1/(2*sqrt(gamma.hat)))^2*sigma2.gamma.hat.1;
wal3[i]<-(sqrt(gamma.hat)-sqrt(gamma.0))^2/sigma2.gamma.hat.3;

## Wald Statistic: using square transformation
## H0: gamma^2=1; g(gamma.hat) = gamma.hat^2 => g'(gamma.hat)
## = 2*gamma.hat;
sigma2.gamma.hat.4<- (2*gamma.hat)^2*sigma2.gamma.hat.1;
wal4[i]<-(gamma.hat^2 - gamma.0^2)^2/sigma2.gamma.hat.4;

## 11/09/2009
## Wald Statistic: testing H0:mu1^2=mu2^2
#wal5[i]<- (mu1.1^2-mu2.1^2)^2/wald5_var(n1,n2,mu1.1,mu2.1,alpha.1);
}
like<-data.frame(null=null,alt=alt,lrt=lrt);
like$lrt[like$lrt<0]<-0;

## Get the exact cutoff for the lrt and the rst and wald tests
lrt.cutoff.exact<-sort(lrt)[M*(1-nu)];
rst.cutoff.exact<- sort(rst)[M*(1-nu)];
#     wal.cutoff.exact<- sort(wal)[M*(1-nu)];
wal1.cutoff.exact<-sort(wal1)[M*(1-nu)];
wal2.cutoff.exact<-sort(wal2)[M*(1-nu)];
wal3.cutoff.exact<-sort(wal3)[M*(1-nu)];
wal4.cutoff.exact<-sort(wal4)[M*(1-nu)];
```

```
    #      wal5.cutoff.exact<-sort(wal5)[M*(1-nu)];
}
else if (cutoff.sim==0)
{
    ## Get power using exact cutoff values given
    if (lrt.cutoff<0)
    {
    "LRT cutoff value must be positive. Cannot compute exact power for
    LRT";
    toggle.lrt<-1;
    }
    if (rst.cutoff<0)
    {
    "RST cutoff value must be positive. Cannot compute exact power for
    RST";
    toggle.rst<-1;
    }
    if (wal1.cutoff<0)
    {
    "Wald 1 cutoff value must be positive. Cannot compute exact power for
    Wald 1";
    toggle.wal1<-1;
    }
    if (wal2.cutoff<0)
    {
    "Wald 2 cutoff value must be positive. Cannot compute exact power for
     Wald 2";
    toggle.wal2<-1;
    }
    if (wal3.cutoff<0)
    {
    "Wald 3 cutoff value must be positive. Cannot compute exact power for
    Wald 3";
    toggle.wal3<-1;
    }
    if (wal4.cutoff<0)
    {
    "Wald 4 cutoff value must be positive. Cannot compute exact power for
    Wald 4";
    toggle.wal4<-1;
    }
    lrt.cutoff.exact<-lrt.cutoff;
    rst.cutoff.exact<-rst.cutoff;
    wal1.cutoff.exact<-wal1.cutoff;
    wal2.cutoff.exact<-wal2.cutoff;
```

```
    wal3.cutoff.exact<-wal3.cutoff;
    wal4.cutoff.exact<-wal4.cutoff;
}

## B simulations to calculate the power
for (i in 1:B)
{
    mu2<-mu1*(1-delta);
    y1 <- rnbinom(n=n1, mu=mu1,size = alpha);
    ## 11/26/2009 To allow for generating data from NB populations
    with different dispersion parameters
    if (alpha1==0)
    {
    y2 <- rnbinom(n=n2, mu=mu2,size = alpha);
    }
    else
    {
    y2 <- rnbinom(n=n2, mu=mu2,size = alpha1);
    }
    y1bar=mean(y1);y2bar=mean(y2);

    ## Get the MLEs under the restricted and unrestricted case
    null<-nbsimMLEnull(init=alpha,n1=n1,n2=n2,y1,y2,maxit=maxit,
    method=method);
    alt<-nbsimMLEalt(init=alpha,n1=n2,n2=n2,y1,y2);
    lrt<- -2* (null[1]-alt[1]);

    ## Likelihood Ratio Test

    ## Rao's Score Test (Need the MLEs under the restricted case)
    mu.hat.0<-null[2];alpha.hat.0<-null[3];
    first<-firstorder(y1=y1,y2=y2,mu1=mu.hat.0,mu2=mu.hat.0,
    alpha=alpha.hat.0);
    hessian<-secondorder(y1=y1,y2=y2,mu1=mu.hat.0,mu2=mu.hat.0,
    alpha=alpha.hat.0);
    fisherinf.inv<-diag(3);
    fisherinf.inv[1,1]<- -1/hessian[1,1];fisherinf.inv[2,2]<-
    -1/hessian[2,2];
    fisherinf.inv[3,3]<- -1/hessian[3,3];
    rst<- t(first)%*%(fisherinf.inv)%*%first;

    # Aban et. al Wald Statistic: (using  \hat{gamma})
    if (y2bar>0)
    {
    ## Get the MLEs under the restricted and unrestricted case
```

```
null<-nbsimMLEnull(init=alpha,n1=n1,n2=n2,y1,y2,maxit=maxit,
method=method);
alt<-nbsimMLEalt(init=alpha,n1=n2,n2=n2,y1,y2);

## Likelihood Ratio Test
lrt<- -2* (null[1]-alt[1]);

## Rao's Score Test (Need the MLEs under the restricted case)
mu.hat.0<-null[2];alpha.hat.0<-null[3];
first<-firstorder(y1=y1,y2=y2,mu1=mu.hat.0,mu2=mu.hat.0,
alpha=alpha.hat.0);
hessian<-secondorder(y1=y1,y2=y2,mu1=mu.hat.0,mu2=mu.hat.0,
alpha=alpha.hat.0);
fisherinf.inv<-diag(3);
fisherinf.inv[1,1]<- -1/hessian[1,1];fisherinf.inv[2,2]<-
-1/hessian[2,2];
fisherinf.inv[3,3]<- -1/hessian[3,3];
rst<- t(first)%*%(fisherinf.inv)%*%first;


## Wald Tests: Need MLEs of the parameters in the unrestricted case
mu1.hat.1<-alt[2];mu2.hat.1<-alt[3];alpha.hat.1<-alt[4];
gamma.hat<-mu2.hat.1/mu1.hat.1;gamma.0<-1;
sigma2.gamma.hat.1<-(gamma.hat*(n1*(alpha.hat.1+gamma.hat*mu1.hat.1)
+n2*gamma.hat*(alpha.hat.1+mu1.hat.1)))/(n1*n2*alpha.hat.1*mu1.hat.1);
wal1<-(gamma.hat-gamma.0)^2/sigma2.gamma.hat.1;

# Wald Statistic: (using log transformation of gamma hat)
# g(gamma.hat) = log(gamma.hat); So g'(gamma.hat) = 1/gamma.hat
sigma2.gamma.hat.2<-(1/gamma.hat)^2*sigma2.gamma.hat.1;
wal2<-(log(gamma.hat)-log(gamma.0))^2/( (1/gamma.hat)^2
 *sigma2.gamma.hat.1);

# Wald Statistic: using square root transformation;
# g(gamma.hat) = (gamma.hat)^(0.5); g'(gamma.hat) = 1/(2*sqrt(gamma.hat))
sigma2.gamma.hat.3<- (1/(2*sqrt(gamma.hat)))^2*sigma2.gamma.hat.1;
wal3<-(sqrt(gamma.hat)-sqrt(gamma.0))^2/sigma2.gamma.hat.3;

# Wald Statistic: using square transformation;
# g(gamma.hat) = gamma.hat^2 => g'(gamma.hat) = 2*gamma.hat;
sigma2.gamma.hat.4<- (2*gamma.hat)^2*sigma2.gamma.hat.1;
wal4<-(gamma.hat^2 - gamma.0^2)^2/sigma2.gamma.hat.4;

## Power for LRT, RST and Wald using chisq cutoff value
if (lrt > cutoff.chisq) {lrt.power.chisq<- lrt.power.chisq + 1;}
```

```
if (rst > cutoff.chisq) {rst.power.chisq<-rst.power.chisq+1;}
if (wal1 > cutoff.chisq) {wal1.power.chisq<-wal1.power.chisq+1;}
if (wal2 > cutoff.chisq) {wal2.power.chisq<-wal2.power.chisq+1;}
if (wal3 > cutoff.chisq) {wal3.power.chisq<-wal3.power.chisq+1;}
if (wal4 > cutoff.chisq) {wal4.power.chisq<-wal4.power.chisq+1;}
#if (wal5 > cutoff.chisq) {wal5.power.chisq<-wal5.power.chisq+1;}

## Power for LRT, RST and Wald tests using exact cutoff values
## Only if the cutoff values are valid
if (toggle.lrt==0)
{
if (lrt > lrt.cutoff.exact) { lrt.power.exact<- lrt.power.exact + 1;}
}
if (toggle.rst==0)
{
if (rst > rst.cutoff.exact) {rst.power.exact<-rst.power.exact+1;}
}
if (toggle.wal1==0)
{
if (wal1 > wal1.cutoff.exact) {wal1.power.exact<-wal1.power.exact+1;}
}
if (toggle.wal2==0)
{
if (wal2 > wal2.cutoff.exact) {wal2.power.exact<-wal2.power.exact+1;}
}
if (toggle.wal3==0)
{
if (wal3 > wal3.cutoff.exact) {wal3.power.exact<-wal3.power.exact+1;}
}
if (toggle.wal4==0)
{
if (wal4 > wal4.cutoff.exact) {wal4.power.exact<-wal4.power.exact+1;}
}
}
else if (y2bar==0) ## Assume the treatment is effective
{
 lrt.power.chisq<-lrt.power.chisq+1;
 rst.power.chisq<-rst.power.chisq+1;
 wal1.power.chisq<-wal1.power.chisq+1;
 wal2.power.chisq<-wal2.power.chisq+1;
 wal3.power.chisq<-wal3.power.chisq+1;
 wal4.power.chisq<-wal4.power.chisq+1;

 if (toggle.lrt==0) {lrt.power.exact<-lrt.power.exact+1;}
 if (toggle.rst==0) {rst.power.exact<-rst.power.exact+1;}
```

```
     if (toggle.wal1==0) {wal1.power.exact<-wal1.power.exact+1;}
     if (toggle.wal2==0) {wal2.power.exact<-wal2.power.exact+1;}
     if (toggle.wal3==0) {wal3.power.exact<-wal3.power.exact+1;}
     if (toggle.wal4==0) {wal4.power.exact<-wal4.power.exact+1;}
    }
}

lrt.power.chisq<-round(lrt.power.chisq/B,4);
lrt.power.exact<-round(lrt.power.exact/B,4);
## go one standard error below and see if it contains 0.80 or 0.90
lrt.chisq.lb<-lrt.power.chisq-1*sqrt(lrt.power.chisq*(1-lrt.power.chisq)/B);
lrt.exact.lb<-lrt.power.exact-1*sqrt(lrt.power.exact*(1-lrt.power.exact)/B);

rst.power.chisq<-round(rst.power.chisq/B,4);
rst.power.exact<-round(rst.power.exact/B,4);

rst.chisq.lb<-rst.power.chisq-1*sqrt(rst.power.chisq*(1-rst.power.chisq)/B);
rst.exact.lb<-rst.power.exact-1*sqrt(rst.power.exact*(1-rst.power.exact)/B);

## Chisq power for the other Wald Tests
wal1.power.chisq<-round(wal1.power.chisq/B,4);
wal2.power.chisq<-round(wal2.power.chisq/B,4);
wal3.power.chisq<-round(wal3.power.chisq/B,4);
wal4.power.chisq<-round(wal4.power.chisq/B,4);

## Exact power for the other Wald Tests
wal1.power.exact<-round(wal1.power.exact/B,4);
wal2.power.exact<-round(wal2.power.exact/B,4);
wal3.power.exact<-round(wal3.power.exact/B,4);
wal4.power.exact<-round(wal4.power.exact/B,4);

wal4.chisq.lb<-wal4.power.chisq-1*sqrt(wal4.power.chisq
*(1-wal4.power.chisq)/B);
wal4.exact.lb<-wal4.power.exact-1*sqrt(wal4.power.exact
*(1-wal4.power.exact)/B);

cat("LRT    : Chisq: ", lrt.power.chisq, "Exact: ",lrt.power.exact,
"Exact Cutoff: ",round(lrt.cutoff.exact,4),"\n");
cat("RST    : Chisq: ", rst.power.chisq, "Exact: ",rst.power.exact,
"Exact Cutoff: ",round(rst.cutoff.exact,4),"\n");
cat("WALD 1 : Chisq: ", wal1.power.chisq,"Exact: ",wal1.power.exact,
"Exact Cutoff: ", round(wal1.cutoff.exact,4), "\n");
cat("WALD 2 : Chisq: ", wal2.power.chisq,"Exact: ",wal2.power.exact,
"Exact Cutoff: ", round(wal2.cutoff.exact,4), "\n");
cat("WALD 3 : Chisq: ", wal3.power.chisq,"Exact: ",wal3.power.exact,
```

```
    "Exact Cutoff: ", round(wal3.cutoff.exact,4), "\n");
    cat("WALD 4 : Chisq: ", wal4.power.chisq,"Exact: ",wal4.power.exact,
    "Exact Cutoff: ", round(wal4.cutoff.exact,4), "\n");
    power<-c(lrt.chisq.lb,lrt.exact.lb,lrt.cutoff.exact,rst.chisq.lb,
    rst.exact.lb,rst.cutoff.exact, wal4.chisq.lb,wal4.exact.lb,
    wal4.cutoff.exact);
    return(power);
}
```