ADJUSTING FOR COVARIATES IN ZERO-INFLATED GAMMA AND ZERO-INFLATED LOG-NORMAL MODELS FOR SEMICONTINUOUS DATA

by

Elizabeth Dastrup Mills

An Abstract

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Biostatistics in the
Graduate College of The
University of Iowa

May 2013

Thesis Supervisor: Professor Jeffrey D. Dawson

ABSTRACT

Semicontinuous data consist of a combination of a point-mass at zero and a positive skewed distribution. This type of non-negative data distribution is found in data from many fields, but presents unique challenges for analysis. Specifically, these data cannot be analyzed using positive distributions, but distributions that are unbounded are also likely a poor fit. Two-part models incorporate both the zero values from semicontinuous data and the positive continuous values. In this dissertation, we compare zero-inflated gamma (ZIG) and zero-inflated log-normal (ZILN) two-part models. For both of these models, the probability that an outcome is non-zero is modeled via logistic regression. Then the distribution of the non-zero outcomes is modeled via gamma regression with a log-link for ZIG regression and via log-normal regression for ZILN.

In this dissertation we propose tests which combine the two parts of the ZIG and ZILN models in meaningful ways for performing a two group comparison. Then we compare these tests in terms of observed Type 1 error rates and power levels under both correctly specified and misspecified ZIG and ZILN models. Tests falling under two main hypotheses are examined. First, we look at two-part tests which come from a two-part hypothesis of no difference between the two groups in terms of the probability of non-zero values and in terms of the the mean of the non-zero values. The second type of tests are mean-based tests. These combine the two parts of the model in ways related to the overall group means of the semicontinuous variable. When not adjusting for covariates, two tests are developed based on a difference of means (DM) and a ratio of means (RM). When adjusting for covariates, tests using mean-based hypotheses are developed which marginalize over the values of the adjusting covariates. Under the adjusting framework, two ratio of means statistics are proposed and examined, an average of the subject specific ratio of means (RM_{SS})

and a ratio of the marginal group means (RM_{MAR}). Simulations are used to compare Type 1 error and power for these tests and standard two group comparison tests.

Simulation results show that when ZIG and ZILN models are misspecified and the coefficient of variation (CoV) and/or sample size is large, there are differences in Type 1 error and power results between the misspecified and correctly specified models. Specifically, when ZILN data with high CoV or sample size are analyzed as ZIG, Type 1 error rates are prohibitively high. On the other hand, when ZIG data are analyzed as ZILN under these scenarios, power levels are much lower for ZILN analyses than for ZIG analyses. Examination of Q-Q plots show, however, that in these settings, distinguishing between ZIG and ZILN data can be relatively straightforward. When the coefficient of variation is small it is harder to distinguish between ZIG and ZILN models, but the differences between Type 1 error rates and power levels for misspecified or correctly specified models is also slight.

Finally, we use the proposed methods to analyze a data set involving Parkinson's disease (PD) and driving. A number of these methods show that PD subjects exhibit poorer lane keeping ability than control subjects.

Abstract Approved: ______________________________
Thesis Supervisor

____
Title and Department

____
Date

ADJUSTING FOR COVARIATES IN ZERO-INFLATED GAMMA AND ZERO-INFLATED LOG-NORMAL MODELS FOR SEMICONTINUOUS DATA

by

Elizabeth Dastrup Mills

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Biostatistics in the
Graduate College of The
University of Iowa

May 2013

Thesis Supervisor: Professor Jeffrey D. Dawson

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Elizabeth Dastrup Mills

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Biostatistics at the May 2013 graduation.

Thesis Committee:

Jeffrey D. Dawson, Thesis Supervisor

Michael P. Jones

Brian J. Smith

Gideon K. Zamba

Ergun Y. Uc

To my dear husband Andrew who has been my strength and support from my comprehensive exams to my dissertation defense.

ACKNOWLEDGMENTS

I would like to acknowledge and thank my advisor, Dr. Jeffrey Dawson, for his help and counsel during the dissertation process. Our meetings have been invaluable to in the development and improvement of the content of this dissertation. I am particularly grateful for his patience and flexibility when I began working long distance. I am also thankful for his advice and mentoring throughout my time at the University of Iowa. He has truly helped me to become a better statistician and collaborator.

I would also like to acknowledge the other members of my dissertation committee; Dr. Michael Jones, Dr. Brian Smith, Dr. Gideon Zamba, and Dr. Ergun Uc. Our meetings were enlightening and helped me recognize areas where greater detail and clarity were needed.

Thanks are due to Dale McLerran for his email correspondence about ZIG regression.

I greatly appreciate all of those whom I worked with in the Division of Neuroergonomics, especially Dr. Ergun Uc and Dr. Matthew Rizzo. I also thank them for the data set used for this dissertation.

I'm thankful for the faculty and students in the Department of Biostatistics at the University of Iowa. I have found it to be a place where all root for each other and are on the same team. I am particularly grateful for Dr. Jacob Olsen for his support in the early years as my academic advisor, and for Terry Kirk who has always been a valuable resource.

I would like to thank my family and friends for their support, encouragement, and confidence - including my friends who have let me stay at their homes when I've returned to town for meetings etc. In addition, I'd like to thank my mother, Claudia Dastrup help with editing. And finally, I thank my husband, Andrew Mills, for his support and patience as we navigated the challenges of long-distance schooling.

ABSTRACT

Semicontinuous data consist of a combination of a point-mass at zero and a positive skewed distribution. This type of non-negative data distribution is found in data from many fields, but presents unique challenges for analysis. Specifically, these data cannot be analyzed using positive distributions, but distributions that are unbounded are also likely a poor fit. Two-part models incorporate both the zero values from semicontinuous data and the positive continuous values. In this dissertation, we compare zero-inflated gamma (ZIG) and zero-inflated log-normal (ZILN) two-part models. For both of these models, the probability that an outcome is non-zero is modeled via logistic regression. Then the distribution of the non-zero outcomes is modeled via gamma regression with a log-link for ZIG regression and via log-normal regression for ZILN.

In this dissertation we propose tests which combine the two parts of the ZIG and ZILN models in meaningful ways for performing a two group comparison. Then we compare these tests in terms of observed Type 1 error rates and power levels under both correctly specified and misspecified ZIG and ZILN models. Tests falling under two main hypotheses are examined. First, we look at two-part tests which come from a two-part hypothesis of no difference between the two groups in terms of the probability of non-zero values and in terms of the the mean of the non-zero values. The second type of tests are mean-based tests. These combine the two parts of the model in ways related to the overall group means of the semicontinuous variable. When not adjusting for covariates, two tests are developed based on a difference of means (DM) and a ratio of means (RM). When adjusting for covariates, tests using mean-based hypotheses are developed which marginalize over the values of the adjusting covariates. Under the adjusting framework, two ratio of means statistics are proposed and examined, an average of the subject specific ratio of means (RM_{SS})

and a ratio of the marginal group means (RM_{MAR}). Simulations are used to compare Type 1 error and power for these tests and standard two group comparison tests.

Simulation results show that when ZIG and ZILN models are misspecified and the coefficient of variation (CoV) and/or sample size is large, there are differences in Type 1 error and power results between the misspecified and correctly specified models. Specifically, when ZILN data with high CoV or sample size are analyzed as ZIG, Type 1 error rates are prohibitively high. On the other hand, when ZIG data are analyzed as ZILN under these scenarios, power levels are much lower for ZILN analyses than for ZIG analyses. Examination of Q-Q plots show, however, that in these settings, distinguishing between ZIG and ZILN data can be relatively straightforward. When the coefficient of variation is small it is harder to distinguish between ZIG and ZILN models, but the differences between Type 1 error rates and power levels for misspecified or correctly specified models is also slight.

Finally, we use the proposed methods to analyze a data set involving Parkinson's disease (PD) and driving. A number of these methods show that PD subjects exhibit poorer lane keeping ability than control subjects.

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1
INTRODUCTION

1.1 Motivation

Researchers in many fields often encounter nonnegative data that are right skewed and contain zero values. Because these outcomes are primarily continuous but have a point mass at zero, they are sometimes called semicontinuous data, point-mass mixture data or zero inflated data. The unique features of this type of data need to be accounted for when performing analyses. In this dissertation, a variety of tests employing two-part models are used to analyze semicontinuous data. Specifically, we develop tests for two group comparisons of semicontinuous data both with and without covariate adjustment.

1.2 Semicontinuous Data Examples

Examples of semicontinuous data outcomes are many and varied. In economics studies the amount an individual or household spends on a certain category during the study period is semicontinuous.[1, 2] For example, Duan et al. studied medical spending under various insurance plans.[3] Some individuals file no claims within a given year and others have some positive claims.

Semicontinuous data also arise in studies involving the amount of exercise an individual performs, as a zero value is included when no exercise is performed.[4] Feuerverger suggested analysis of cloud seeding data through a semicontinuous method; a zero indicating no rain and the positive values being the amount of the rain seen when there is rain.[5]

In medicine, examples include the Health Utility Index following an event with high mortality where death has a value of zero,[6] and Coronary Artery Calcification (CAC) scores where a zero value indicates either no calcification or calcification under a limit of detection.[7] In microbiology such nonnegative data could come from assays, virus titers, or metabolomic and proteomic data.[8]

A motivating example for this dissertation comes from studies of subjects in a driving simulator at the University of Iowa. While several measures may be examined to determine vehicle control, one measure that we had previously proposed is an extension of the commonly-used metric of lane deviation count. [9] A subject who tends to drive close to the lane line may have several lane deviations but may actually be safer than a subject who has one lane deviation but remains far outside of their lane for an extended period of time. Because of this, we use for our motivating example a measure of the extent to which a subject is driving outside of the driving lane over the course of a segment of interest; subjects who remain in their lane have a value of zero while subjects who exit their lane have an outcome that is positive and related to the amount of time and the amount of the vehicle that exited the lane. We will call this semicontinuous outcome the Lane Departure Severity Score (LDSS).

Figure 1.1 shows the distribution of this outcome from a data set where subjects with Parkinson's disease and elderly non-diseased control subjects drove in a simulation of foggy conditions. This is shown in histograms of relative frequencies. The first bin, shown in a lighter gray, contains only zero values, showing the proportion of the data that are observed to be zero. Then the remainder of histogram, shown in gray, includes the relative frequencies of the non-zero LDSS scores. It is clear from Figure 1.1 that LDSS is a semicontinuous outcome with a point-mass at zero and positive right skewed distribution for the remaining outcomes. We will return to this example in Chapter 5 after discussing methods to analyze such data in Chapters 2, 3, and 4.

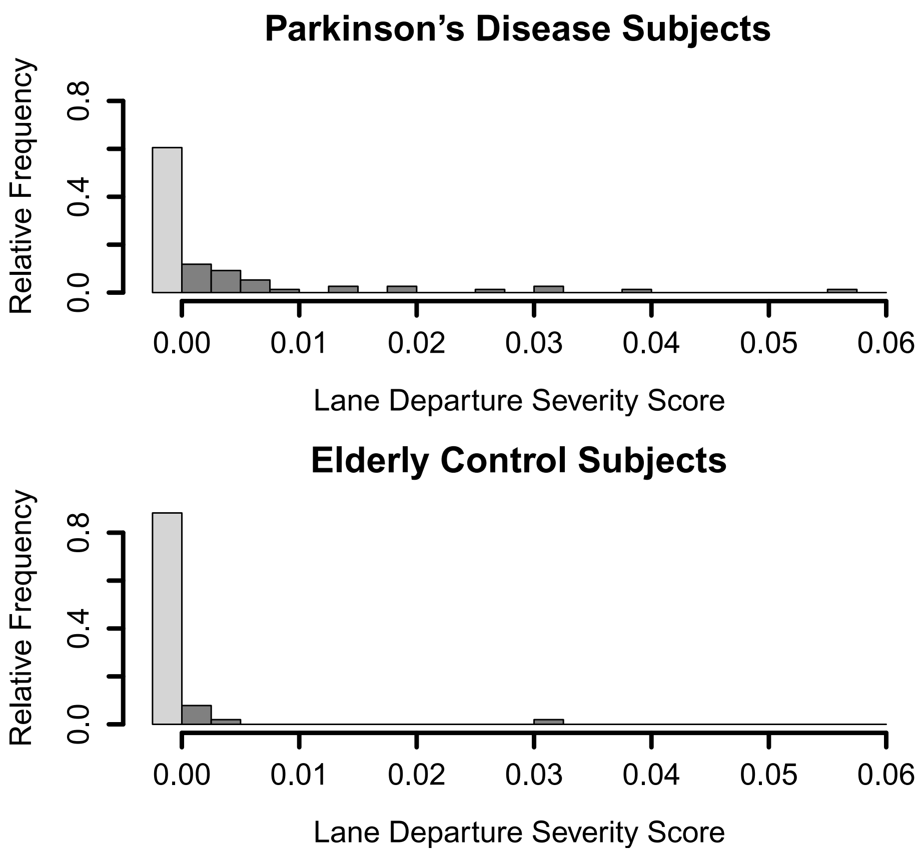

Figure 1.1: Distribution of Lane Departure Severity Score. The first bin, shown in light gray, illustrates the proportion of zero values (no lane deviation); the remainder of the gray histogram indicates the frequency of the other outcome values.

1.3 Semicontinuous Data Modeling Questions

When encountering continuous non-negative outcomes containing zeros, a researcher should address several questions in order to choose an appropriate model. Some of these questions include: Are the zeros observed true zeros or censored zeros? What distributional assumptions are reasonable for the positive values? and What is the hypothesis of interest?

1.3.1 True Zeros vs Censoring

When otherwise positive continuous data contain a point-mass at zero it is important to contemplate whether the zeros encountered represent censored values due to a limit of detection, true zeros due to actual observations not due to censoring, or a mixture of true zeros and censored values.

In medical spending examples, the zero spending values observed are true zeros because it was observed directly that the individuals or household had no medical claims during the study term.[1, 2, 3] Similarly in the driving simulation, a subject can have no lane departures for a given drive segment. The Coronary Artery Calcification scores[7, 8] and the microbiology examples, on the other hand, include limits of detection and therefore it is likely that some or all of the observed zero values are likely to be censored zeroes, not true zeros. In other situations the censoring vs. true zeros question is not clear from the context. To deal with such cases, formal likelihood methods have been proposed to test for the presence of censoring.[10]

Semicontinuous point-mass mixture data are data where the zeros observed are true zeros. Conceptually a true zero is different from a censored zero; a true zero typically represents something not happening, whereas a censored zero only indicates that its occurrence was below a certain threshold. This dissertation focuses on semicontinuous data without censored zeros.

1.3.2 Distributional Choices

Semicontinuous data cannot be analyzed through standard parametric techniques because standard positive continuous distributions such as the gamma, log-normal, log-logistic and beta have null probabilities for zero values. Also, under some conditions, standard nonparametric tests have low power resulting from the high number of ties.[6] Therefore, it is important to consider parametric techniques which can accommodate zero values.

Modelling semicontinuous data with true zero is often done in two steps; such models have been called two-part models. The first part models the probability that an outcome is non-zero and the second models the value of an outcome given it is greater than zero. Duan et al. used a probit regression to model the probability of a non-zero outcome and a log-normal model to model the value of the non-zero outcome.[3] Zhou and Tu used logistic regression for the probability portion, and log-normal models for the continuous portion.[11] McLerran proposed a similar model using a logistic model for the probability portion and and gamma model with a log link for modeling the value of the non-zero outcome.[12] This dissertation describes modelling the probability of a non-zero outcome via logistic regression, and compares models where the non-zero outcome is assumed to be distributed as either gamma or log-normal. These models are explained in more detail in Chapter 2.

1.3.3 Hypothesis Choices

Another important point to consider is what hypotheses and goals are of interest. Some researchers may only be interested in finding the best fitting model that describes their data. Several studies involving two-part models have not attempted to combine the parts in a meaningful way, but have instead talked about the conclusions found in each part separately.[3, 5, 13] Others are interested in a compound null hypothesis that examines if there is an effect of a covariate on either the continuous or the binary parts of the model. Such a hypothesis does not require the effects in the two parts to be in the same direction with respect to their effect on the overall mean differences. For example, one group could have a higher probability of a non-zero outcome, but also have a lower mean of the outcome given a non-zero outcome. Lachenbruch calls these effects in opposite directions 'dissonant effects'. In contrast, 'consonant effects' are in the same direction relative to the overall mean, i.e. when the group with the highest probability of a non-zero outcome also has the highest mean of the non-zero outcome, both leading to that group having an higher overall mean.[14] Other researchers are interested in making a judgement that combines the two part of the model in a directional manner, i.e. that one group or one treatment is performing better on average than the other, or more specifically that one group has a higher overall mean outcome than the other group.[1, 2, 6, 12]

A general null hypothesis examines whether there are differences in the distributions of group 1 and group 0. This can be written as:

$$\begin{aligned} H_0 &: f_0(Y) = f_1(Y) \\ H_1 &: f_0(Y) \neq f_1(Y). \end{aligned} \tag{1.1}$$

When using a two-part model this can be written in terms of conditional means and probabilities. Lachenbruch examined this question of whether the two groups differ in terms of the probability of zero and the conditional mean.[14] Lachenbruch proposed two-part tests resulting from the following null and compound alternative hypotheses:

$$\begin{aligned} H_0 &: p_0 = p_1 \cap \mu_0 = \mu_1 \\ H_1 &: p_0 \neq p_1 \cup \mu_0 \neq \mu_1 \end{aligned} \tag{1.2}$$

where p_j is the probability of a non-zero outcome given group j and μ_j is the expected mean of the outcome given group j and that the outcome is greater than zero, with groups labeled 0 and 1 where (i.e. $j = 0, 1$). These hypotheses are equivalent to those in Equation 1.1 if a two-part model is assumed.

Tu and Zhou[1], Zhou and Tu[2], and McClerran[12] examine the hypothesis of equality of the overall mean of two groups, $M_j = E(Y|group = j)$ where Y is the semicontinuous outcome. This hypothesis can be written as follows:

$$\begin{aligned} H_0 &: M_0 = M_1 \\ H_1 &: M_0 \neq M_1. \end{aligned} \tag{1.3}$$

McLerran[12] tests this hypothesis using the difference between two means (DM) defined as,

$$DM = M_1 - M_0. \tag{1.4}$$

Zhou and Tu[2] create a confidence interval relating to a ratio of means (RM), with RM defined as,

$$RM = \frac{M_1}{M_0}. \tag{1.5}$$

Tooze et al.[13] also reference the RM in their interpretation of two-part model results. Their application includes adjusting for covariates and they note that the value of RM changes with changing covariate values. In later chapters, this idea will be further developed using a test based on an average ratio of means.

1.4 Description of Remaining Chapters

Chapter 2 introduces two-part models in detail and examines the two-part hypothesis of difference between two groups in terms of the probability of zero and conditional mean (as described above in Equation 1.2). Several tests are compared within this framework for models assuming gamma or log-normal distributions for the continuous part of the model. Chapter 3 examines tests for an equality of means between two groups when no covariates are present using a difference of means test (DM) and a ratio of means test (RM). These tests are then compared under two different distributional models. Both models use logistic regression to model the probability of non-zero vales. Zero-inflated gamma models (ZIG) use gamma regression with a log link to model the non-zero values; zero-inflated log-normal models (ZILN) use a simple linear regression to model the log of the non-zero values. Chapter 4 expands the tests in Chapter 3 to newly proposed tests that will compare the means of two groups adjusting for covariates; first for dichotomous and then for continuous covariates. Under the adjusting framework, two ratio of means statistics are examined: a subject specific ratio of means (RM_{SS}) and a marginal ratio of means (RM_{MAR}). Tests based on DM, RM_{SS} and RM_{MAR} are compared to each other under ZIG and ZILN scenarios. Chapters 5 contains the data example and discussions about model fitting. Chapter 6 presents the final conclusions of the dissertation including a schematic showing which methods were proposed by others and which are original to this dissertation.

CHAPTER 2
TWO-PART TESTS AND MODELS FOR ZERO INFLATED GAMMA AND ZERO INFLATED LOG NORMAL DISTRIBUTIONS

2.1 Introduction

As introduced in Section 1.3.2, semicontinuous data can be modeled in two parts: one consisting of the probability of a non-zero value and the other consisting of the distribution of the continuous non-zero values. These models assume that the zero value are true observed zeros. For the lane departure scenario, the two-part model would model the probability that a subject drives out of their lane and the extent to which they drive outside of their lane.

An important question is how information will be combined across the two parts of the model to address specific scientific hypotheses. This chapter focuses on two-group comparisons based on two-part hypothesis tests. In this type of test, information from the two parts is pooled into a global two degree of freedom (df) test which lacks any constraints regarding the direction of the effects in the two parts. First, specific details concerning the parameterization and likelihood functions of the models are discussed. Then we show how to use these models for estimation and unadjusted two-group hypothesis testing. Finally, we perform simulation studies to examine the Type 1 error rates and power for these methods.

2.2 Two-part Model

The two-part model framework provides an appropriate structure for modeling semicontinuous data.[3, 5] Let Y_i be the semicontinuous outcome for subject i and X_i be a vector of covariates for subject i; let β be the parameters used in modeling the probability of positive responses, and let θ represents mean and dispersion parameters of the conditional distribution of the positive responses. The two parts of the model could contain different sets of covariates, but we will assume that they are the same and use X_i for both parts of the model. Then the probability of a positive response can be denoted as $p_\beta = P(Y_i > 0|X_i)$ and the conditional distribution of the positive responses can be represented as $g_\theta(Y_i|Y_i > 0)$. Also, let the indicator function, $\mathbf{I}(Y_i > 0)$, be defined such that if $Y_i > 0$ then $\mathbf{I}(Y_i > 0) = 1$, else $\mathbf{I}(y_i > 0) = 0$. This framework then results in the following mixture pdf and likelihood:

$$f(y) = (1 - p_\beta)\mathbf{I}(y = 0) + p_\beta g_\theta(y|y > 0) \tag{2.1}$$

$$L(\beta, \theta) = \prod_{Y_i=0} (1 - p_\beta) \prod_{Y_i>0} p_\beta g_\theta(Y_i|Y_i > 0) \tag{2.2}$$

Equation 2.2 can be factored into two parts: one related only to the β parameters involved in estimating p_β and the other including only the parameters involved in estimating the θ parameters. Duan et al. illustrated the factorability of the likelihood in his use of a two-part zero-inflated log-normal regression model.[3] While they used probit regression for the zero values, we will use logistic regression. A generalization of Duan's result is shown in Equation 2.3. Factorization is possible when the Y_i are independent of each other for any model choices for p_β and $g_\theta(Y_i|Y_i > 0)$ which do not induce dependence between β and θ given the data.

$$L(\beta, \theta) = L_1(\beta)L_2(\theta) = \left[\prod_{Y_i=0} (1 - p_\beta) \prod_{Y_i>0} p_\beta\right] \left[\prod_{Y_i>0} g_\theta(Y_i|Y_i > 0)\right] \tag{2.3}$$

This factorization allows for $L_1(\beta)$ and $L_2(\theta)$ to be maximized separately. This result holds whether or not the covariates corresponding to β and θ are the same. However, if regression coefficients from the two parts are constrained to be related, this separate maximization is no longer applicable.[15] This occurs with 'shared parameters' used by Moulton et al.[15] where they assumed that the parameters in the two parts of the model are proportional to each other. The separability also disappears when working with longitudinal outcomes or any other scenario where Y_i are not independent.[16] In this dissertation, all methods developed are for situations with cross-sectional, uncorrelated observations. As the regression parameter is not constrained, the likelihood factorization outlined in Equation 2.3 applies. Under this regression framework, p_β is modeled via β. We will let θ consist of the parameters used to model the positive Y_i using τ and a generic dispersion parameter δ. The corresponding two-part likelihood is then expressed as:

$$\begin{aligned} L(\beta, \tau, \delta) &= L_1(\beta) L_2(\tau, \delta) \\ &= \left[\prod_{Y_i=0} (1 - p_\beta) \prod_{Y_i>0} p_\beta \right] \left[\prod_{Y_i>0} g_{\tau,\delta}(Y_i | Y_i > 0) \right] \end{aligned} \tag{2.4}$$

In modeling p_β we have used logistic regression with $\text{logit}(p_\beta) = X_i\beta$. One way of modeling the positive Y_i is through gamma regression with a log link where $Y_i \sim \Gamma(exp(X_i\tau), \nu)$. For this model, the generic dispersion parameter δ is replaced by ν. Note that $\nu^{-1} = CoV^2$ where CoV is the coefficient of variation. Combining these two pieces of the model as outlined in Equation 2.4 leads to a zero-inflated gamma distributions (ZIG).

Another method of modeling the positive Y_i is through log-normal regression with $\log(Y_i) = X_i\tau + \epsilon_i$ where $\epsilon_i \sim Normal(0, \sigma^2)$. Here δ is replaced by σ^2 and σ^2 is related to the CoV by $\sigma^2 = ln(CoV^2 + 1)$. The overall model combining the logistic regression and the log-normal regression is called zero-inflated log-normal (ZILN). The maximum likelihood estimates $\hat{\beta}$ and $\hat{\tau}$ can be found via a Newton-Raphson algorithm for ZIG; for ZILN, $\hat{\beta}$ can be found via Newton-Raphson and $\hat{\tau}$ can be found via direct maximization. See Section 2.4 for further details.

2.3 Two-part Hypothesis

As described in the Chapter 1 (Equation 1.2), if we are interested in modeling or looking for differences in either or both of the two parts of the models due to a covariate that is used in both parts of the model it may be of interest to look at a compound null hypothesis. If, for example, a covariate (e.g., X_j) is in both parts of the model and we are interested in relationships between the outcome and the covariate, a hypothesis set may look like this:

$$\begin{aligned} H_0 &: \beta_j = 0 \cap \tau_j = 0 \\ H_1 &: \beta_j \neq 0 \cup \tau_j \neq 0 \end{aligned} \tag{2.5}$$

where β_j is the parameter associated with X_j from the logistic portion of the model and τ_j is the parameter associated with X_j in the continuous part of the model.

This hypothesis allows for finding varied patterns of effects. It can find effects where the β_j and τ_j have the same sign, 'consonant effects'. It can also find effects where β_j and τ_j have different signs, 'dissonant effects'. When dissonant effects occur, there can sometimes be a canceling out or near canceling out of effects when looked at on an overall mean scale. For example, in a two-group comparison setting a dissonant effect could occur where two groups have the same overall mean, with one group having a higher proportion of zero values and the other group having a higher conditional mean. In such a setting, there are differences between the groups in terms of both the proportion of nonzero values and the mean of the nonzero values, however, the differences are not in the same direction. The alternative hypothesis is that there are differences in the two groups including either consonant effects or dissonant effects regardless of whether or not the dissonant effects cancel out on a mean scale. This type of hypothesis is useful in addressing differences in probability of non-zero values or the value of the non-zero values between two groups or in relation to a covariate. This is distinct from a hypothesis of difference of the overall mean of the two groups. If one is only interested in consonant effects, or in the comparisons of the overall means of the groups, the mean-based tests which will be described in Chapter 3 could be used.

2.4 Likelihood and Estimation

2.4.1 Zero-Inflated Gamma

The ZIG likelihood follows the format of Equation 2.3 where p_β is modeled as $\text{logit}(p_\beta) = X\beta$ and $g(y|y>0; X, \tau, \delta) = g(y|y>0; X,\tau, \nu^{-1})$ such that

$$g(y|y>0; X, \tau, \nu^{-1}) = \frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu_i}\right)^{\nu} y^{\nu-1}\exp\left(-\frac{\nu y}{\mu_i}\right) \tag{2.6}$$

where μ modeled as $\log(\mu) = X\tau$.[17] This leads a mixture pdf of:

$$f(y) = (1-p_\beta)\mathbf{I}(y=0) + p_\beta\frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu}\right)^{\nu} y^{\nu-1}\exp\left(-\frac{\nu y}{\mu}\right) \tag{2.7}$$

and an overall likelihood of:

$$\begin{aligned} L(\beta,\tau,\nu^{-1}) &= L_1(\beta)L_2(\tau,\nu^{-1}) \\ &= \left[\prod_{Y_i=0}(1-p_\beta)\prod_{Y_i>0}p_\beta\right]\left[\prod_{Y_i>0} g_{\tau,\nu^{-1}}(Y_i|Y_i>0)\right] \\ &= \left[\prod_{Y_i=0}\frac{1}{1+e^{X_i\beta}}\prod_{Y_i>0}\frac{e^{X_i\beta}}{1+e^{X_i\beta}}\right]\left[\prod_{Y_i>0}\frac{1}{\Gamma(\nu)}\left(\frac{\nu}{e^{X_i\tau}}\right)^{\nu} Y_i^{\nu-1}\exp\left(-\frac{\nu Y_i}{e^{X_i\tau}}\right)\right] \end{aligned} \tag{2.8}$$

As mentioned in the general example in Section 2.3, the ZIG likelihood is factorable into one part where the β parameters are the only parameters and another part where the τ parameters and ν^{-1} are the only parameters. Maximizing $L_1(\beta)$ and $L_2(\tau,\nu^{-1})$ separately will also maximize the overall likelihood. This maximization can be performed via a Newton-Raphson algorithm for each part.

The following definitions set up the algorithm for the Newton-Raphson algorithm for logistic regression. Let X be an $n\times q$ matrix of covariates used in both parts of the model where $q-1$ is the number of covariates included in the model. When using two-part models, the design (X) matrices for the two parts could be different. For example, different design matrices could be used to adjust for something in one part of the model, but not the other. However, in this dissertation the focus is on comparing one variable across both parts of the model and we will using using the same design matrix for our methods. The indicator function, $\mathbf{I}(Y_i > 0)$, is defined such that if $Y_i > 0$ then $\mathbf{I}(Y_i > 0) = 1$, else $\mathbf{I}(y_i > 0) = 0$. Let Y_b be an $n \times 1$ vector of the outcome Y_{bi}, where $Y_{bi} = \mathbf{I}(y_i > 0)$ and n is the overall sample size. Also define $\hat{\beta}^{(k)}$ to be a $q \times 1$ vector of the estimates of the β parameters at the k^{th} iteration and $\hat{P}^{(k)}$ to be an $n \times 1$ vector of estimated probabilities that $Y > 0$ with $\text{logit}(\hat{P}^{(k)}) = X\hat{\beta}^{(k)}$.

Then, the score function for the binomial part of the model is a $q \times 1$ vector of first derivatives of the log-likelihood in terms of the β and is $U_b(\beta) = X'(Y_b - P)$ and the Fisher Information, a $q \times q$ matrix of second derivatives, is $I_b(\beta) = \frac{1}{n}X'P(1-P)X$. Finally, the update step of the Newton-Raphson algorithm is $\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + I_b(\hat{\beta}^{(k)})^{-1}U_b(\hat{\beta}^{(k)})$. For each iteration $(k+1)$ of the algorithm, the $\hat{\beta}^{(k)}$ estimates from the previous iteration were used via the update step to find the new $\hat{\beta}^{(k+1)}$ until the $k+1^{th}$ log-likelihood value is within 0.0000001 of the k^{th} log-likelihood value. A step-halving process was also used when the new estimate $\hat{\beta}^{(k+1)}$ lead to a a lower log-likelihood than the previous $\hat{\beta}^{(k)}$ did.

For the gamma regression part of the ZIG model we maximized the likelihood using only the data for which $Y_i > 0$. Y_c is defined as an $n_c \times 1$ vector containing the Y_i for which $Y_i > 0$ where n_c is the number of subjects with non-zero outcomes. Let X_c be a $q \times n_c$ matrix that includes the covariates used for the gamma regression containing only the lines for the observations where $Y_i > 0$. Let $\hat{\tau}^{(k)}$ be the estimate of τ parameters at the k^{th} iteration, and let $\hat{\mu}^{(k)}$ be the vector of estimated conditional means for Y_i given $Y_i > 0$ with $\log(\hat{\mu}^{(k)}) = X_b\hat{\tau}^{(k)}$. For the gamma distribution, the unscaled score function is $U_c(\tau) = X_c' \frac{Y_c - \mu}{\mu}$, with information resulting from only the data with nonzero outcomes defined as $I_c(\tau) = (X_c' X_c)^{-1}$. The update step of the Newton-Raphson algorithm is then $\hat{\tau}^{(k+1)} = \hat{\tau}^{(k)} + I_c(\hat{\tau}^{(k)})^{-1} U_c(\hat{\tau}^{(k)})$. The algorithm continues in the same pattern as described in the above paragraph. The dispersion parameter, ν^{-1}, may be estimated via the deviance $\hat{\nu}_d^{-1}$.[17]

2.4.2 Zero-Inflated Log-Normal

ZILN regression also follows the format of Equation 2.3. As in the ZIG regression, we model $p(X_i, \beta)$ such that $\text{logit}(p(X_i, \beta)) = X_i\beta$. The difference between the two models comes in the continuous part, $g(Y_i|Y_i > 0; X_i, \tau, \delta)$. For the ZILN regression, $g(Y_i|Y_i > 0; X_i, \tau, \delta)$ is defined as $g(Y_i|Y_i > 0; X_i, \tau, \sigma^2)$ such that

$$g(y|y > 0; X, \tau, \sigma^2) = \frac{1}{y\sqrt{2\pi\sigma^2}} exp\left(-\frac{(\ln(y) - \mu_i)^2}{2\sigma^2}\right) \tag{2.9}$$

with σ^2 equivalent to $ln(\nu^{-1} + 1)$. This leads to mixture pdf of

$$f(y) = (1 - p_\beta)\mathbf{I}(y = 0) + p_\beta \frac{1}{y\sqrt{2\pi\sigma^2}} exp\left(-\frac{(\ln(y) - \mu_i)^2}{2\sigma^2}\right) \tag{2.10}$$

and a full likelihood of

$$\begin{aligned} L(\beta, \tau, \sigma^2) &= L_1(\beta)L_2(\tau, \sigma^2) \\ &= \left[\prod_{Y_i=0} (1 - p(X_i, \beta)) \prod_{Y_i>0} p(X_i, \beta)\right] \left[\prod_{Y_i>0} g_{\tau,\sigma^2}(Y_i|Y_i > 0)\right] \\ &= \left[\prod_{Y_i=0} \frac{1}{1 + e^{X_i\beta}} \prod_{Y_i>0} \frac{e^{X_i\beta}}{1 + e^{X_i\beta}}\right] \left[\prod_{Y_i>0} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(ln(Y_i) - \mu_i)^2}{2\sigma^2}\right) \frac{1}{Y_i}\right]. \end{aligned} \tag{2.11}$$

Since the likelihood is factorable, the estimation of β is the same as in ZIG. The estimation of τ utilizing only the data with nonzero outcomes can be obtained via straightforward maximum likelihood estimation where $\hat{\tau} = (X_c' X_c)^{-1} X_c' \log(Y_c)$. Finally, the estimate of σ^2 is $\hat{\sigma^2} = \frac{\sum_{Y_i>0} \log(Y_i) - X_{ci}\hat{\tau}}{n_c - q}$.

2.5 Two-Part Tests

The score function and expected information matrices derived from the full likelihood of ZIG and ZILN models are outlined in Appendix Section A.1.1 and A.2.1 respectively. For example, the score function for ZIG, as found in Appendix Section A.1.1, Equation A.8 , is:

$$U(\beta,\tau) = \begin{pmatrix} U_b(\beta) \\ U_c(\tau) \end{pmatrix} = \begin{pmatrix} X'(Y_b - P) \\ X' D_{Y_b}\left(\frac{Y-\mu}{\mu}\right) \end{pmatrix} \tag{2.12}$$

where Y_b is an n $\times$ 1 vector defined such that the element $Y_{bi} = \mathbf{I}(y_i > 0)$, D_{Y_b} is an n $\times$ n diagonal matrix with Y_b on the diagonals and 0 elsewhere, P is an n $\times$ 1 vector containing for each row i, $P(Y_i > 0) = \frac{exp^{\beta_0+\beta_1 x_i}}{1+exp^{\beta_0+\beta_1 x_i}}$, and $\frac{Y-\mu}{\mu}$ is an n $\times$ 1 vector with elements $\frac{y_i - e^{\tau_0+\tau_1 x_i}}{e^{\tau_0+\tau_1 x_i}}$.

The expected information matrix for ZIG, $\mathcal{I}(\beta,\tau)$, as found in Appendix Section A.1.1 Equation A.27 is

$$\mathcal{I}(\beta,\tau,\nu) = \begin{pmatrix} \mathcal{I}_b(\beta) & 0 \\ 0 & \mathcal{I}_c(\beta,\tau) \end{pmatrix} = \begin{pmatrix} X'D_{p(1-p)}X & 0 \\ 0 & \nu X' D_p X \end{pmatrix} \tag{2.13}$$

where D_p is an n $\times$ n diagonal matrix with $p_i(1-p_i)$ on the diagonals and 0 elsewhere and D_p is also a diagonal n $\times$ n matrix but with p_i on the diagonals.

Similarly for ZILN, Appendix Section A.2.1 outlines the score function and expected information. The expected information matrices for both ZIG and ZILN analyses are block diagonal and can be used to create two-part tests. The two pieces of the tests are independent and each test defined is distributed as χ^2_1. These individual parts can be added together to have an overall two-part test that is distributed as χ^2_2. [3, 14]

2.5.1 Two-Part Wald Test

Given the estimates of $\hat{\beta}$ and $\hat{\tau}$ and the expected information matrices defined above, Wald tests can be created for each β_j and τ_j which are each approximately distributed as χ^2_1. Let $\mathcal{I}_b^{jj}$ be the j^{th} diagonal element of $\mathcal{I}_b^{-1}$ and $\mathcal{I}_c^{jj}$ be the j^{th} diagonal element of $\mathcal{I}_c^{-1}$ then $Var(\beta_j) = \frac{1}{n}\mathcal{I}_b^{-jj}(\beta)$ and $Var(\tau_j) = \frac{1}{n}\mathcal{I}_c^{jj}(\tau)$ are the variance estimates for $\hat{\beta}_j$ and $\hat{\tau}_j$ respectively.

Then, assuming the null hypothesis outlined in Equation 2.5, a two-part Wald test can be defined as follows:

$$\begin{aligned} W(\hat{\beta}_j, \hat{\tau}_j) &= W(\hat{\beta}_j) + W(\hat{\tau}_j) \\ &= \frac{\hat{\beta}_j^2}{Var(\hat{\beta}_j)} + \frac{\hat{\tau}_j^2}{Var(\hat{\tau}_j)} \end{aligned} \tag{2.14}$$

with $W(\hat{\beta}_j, \hat{\tau}_j)$ asymptotically distributed as a χ^2_2. This test can be used under both ZIG and ZILN models using the estimates provided by each model.

2.5.2 Two-Part Score Test

For the two-part score test, in addition to the full model in which $\hat{\beta}$ and $\hat{\tau}$ are estimated, a reduced model needs to be run which sets the parameters related to the variable of interest (e.g. $\hat{\beta}_j$ and $\hat{\tau}_j$) to the hypothesized value of zero. Then $\hat{\beta}_{red}$ and $\hat{\tau}_{red}$ can be found for the ZIG model using the Newton-Raphson algorithm, or the ZILN model using the Newton-Raphson algorithm and by maximum likelihood estimation methods described in Section 2.4. Using the score function from Equation 2.12 and the expected Fisher information as in equation 2.13, the score test can then be found as $U(\hat{\beta}_{red}, \hat{\tau}_{red})\mathcal{I}^{-1}(\hat{\beta}_{red}, \hat{\tau}_{red}, \hat{\nu}_{red})U(\hat{\beta}_{red}, \hat{\tau}_{red})$.[18] This is equivalent to a two-part score test defined as:

$$S(\hat{\beta}_j, \hat{\tau}_j) = U_{bj}(\hat{\beta}_{red})\mathcal{I}_b^{jj}(\hat{\beta}_{red})U_{bj}(\hat{\beta}_{red}) + U_{cj}(\hat{\tau}_{red})\mathcal{I}_c^{jj}(\hat{\beta}_{red}, \hat{\tau}_{red}, \hat{\nu}_{red})U_{cj}(\hat{\tau}_{red}). \tag{2.15}$$

The terms used in the first part of $S(\hat{\beta}_j, \hat{\tau}_j)$ are defined as follows; First, $U_{bj}(\hat{\beta}_{red})$ is the j^{th} element of $U_b(\hat{\beta}_{red}) = X_b'(Y_b - \hat{P}_{red})$ with $\hat{P}_{red} = \frac{exp(X_c\hat{\beta}_{red})}{1+exp(X_c\hat{\beta}_{red})}$ being the estimated probability of a non-zero value for Y under the null. Next, $\mathcal{I}_b(\hat{\beta}_{red}) = \frac{1}{n}X_b'\hat{P}_{red}(1-\hat{P}_{red})X_b$, leading to $\mathcal{I}_b^{jj}(\hat{\beta}_{red})$ as the element in $\mathcal{I}_b(\hat{\beta}_{red})^{-1}$ that corresponds to the parameter of interest (e.g. $\hat{\beta}_j$).

For the terms in the second part of $S(\hat{\beta}_j, \hat{\tau}_j)$, $U_{cj}(\hat{\tau}_{red})$ is the j^{th} element of $U_c(\hat{\tau}_{red}) = X_c'(Y_c - \hat{\mu}_{red})/\hat{\mu}_{red}$ where $\hat{\mu}_{red} = exp(X_c\hat{\tau}_{red})$ is the estimated mean of a non-zero value for Y calculated from the reduced model. Also, $\mathcal{I}_c^{jj}(\hat{\tau}_{red})$ is the j^{th} diagonal element of the inverse Expected Fisher information. As outlined in Equation 2.13, the expected Information for ZIG is $\mathcal{I}_c(\hat{\beta}_{red}, \hat{\tau}_{red}) = \nu X'D_pX$, where D_p is an n $\times$ n diagonal matrix with the elements of $\hat{P}_{red}$ on the diagonals and 0 elsewhere.

2.5.3 Likelihood Ratio Test

As was the case for the two-part score test, to calculate a likelihood ratio test (LRT) the same full and reduced models as described in Section 2.5.1 must be estimated. Use of the likelihood equations in Section 2.4 calculation of a 2 df LRT is straightforward. Using Equation 2.8 for ZIG models and 2.11 for the ZILN models the LRT is

$$LRT = 2(L(\hat{\beta}, \hat{\tau}, \hat{\nu}^{-1}) - L(\hat{\beta}_{red}, \hat{\tau}_{red}, \hat{\nu}_{red}^{-1})) \tag{2.16}$$

where $L\hat{R}T$ is distributed as χ_2^2.

2.5.4 Wald-Wilcoxon Test

The two-part Wald-Wilcoxon test was proposed by Lachenbruch.[14] The probability of a a non-zero outcome is modeled via logistic regression and is, therefore, identical to the first part of the Wald test from Equation 2.14. A test for the difference between the two groups with regards to the non-zero values is then preformed using a Wilcoxon rank sum test. The Wald-Wilcoxon method can only work as is when comparing two groups and not accounting for other covariates. The Wilcoxon rank sum test is defined in Equation 2.17 with the following notation. Let $r_i = rank(Y_i)$, n_{0c} equal the number of non-zero values in group 0, n_{1c} equal the number of non-zero values in group 1, $R_0 = \frac{\sum r_i|group=0}{n_{0c}}$, and $R_1 = \frac{\sum r_i|group=1}{n_{1c}}$ then the Wald-Wilcoxon test will be:

$$\begin{aligned} WWx &= W_b(\hat{\beta}_1) + Wx \\ &= \frac{\hat{\beta}_1^2}{\hat{Var}(\hat{\beta}_1)} + \frac{(R_0 - R_1)^2}{\text{var}(r)(1/n_{0c} + 1/n_{1c})} \end{aligned} \tag{2.17}$$

where under under H_0 as described in Equation 2.5, $Wx \sim \chi_1^2$ and $W_b(\beta_1) \sim \chi_1^2$. This implies that $WWx \sim \chi_2^2$ for the global two-part test.

2.6 Simulation Methods for Two-Part Tests

We performed a simulation study to compare Type 1 error rates and power for the 2 df tests described in the previous sections. The Type 1 error and power results from these tests for ZIG and ZILN distributions under both correct specification and misspecification scenarios were compared. That is, both ZIG and ZILN data sets were simulated and for each data set both ZIG and ZILN models were run and the 2 df tests were created.

In this chapter, Type 1 error and power results are given for simulations of two-group comparisons with no additional covariates. Therefore, β and τ each consist of an intercept parameter (β_0 and τ_0) and a group effect parameter (β_1 and τ_1). We define X_i as a group indicator where $X_i = 0$ for group 0 and $X_i = 1$ for group 1. Data were simulated as both ZIG and ZILN with various sample sizes and parameter values. The simulation of ZIG and ZILN data requires a two-step process. First, for subject i simulate a Y_i^* from a binomial distribution where $\text{logit}(P(Y_i^* = 1|X_{1i} = x_{1i})) = \beta_0 + \beta_1 x_{1i}$. Then, if $Y_i^* = 0$ set the outcome Y_i to equal 0; if $Y_i^* = 1$ then simulate Y_i from either a gamma with $Y_i \sim Gamma(e^{\tau_0+\tau_1 x_{1i}}, \nu)$ or from a log-normal distributions with $\log(Y_i) \sim N(\tau_0 + \tau_1 x_{1i}, \sigma^2)$.

The parameters for the two-part model were estimated using the Newton-Raphson technique outlined in Section 2.4. These methods as well as other methods outlined in this dissertation were programmed using R. For some settings, some of the simulated data sets had either no zeros or no non-zero values resulting in the parameters of the two-part model not being estimable. For these and other settings, the Newton-Raphson estimation methods did not converge. The total Type 1 error and power results removed these data sets from the numerator and the denominator of the percentage calculations. Settings where more than 1% of the data sets had inestimable parameters are excluded from our reported results.

2.6.1 Simulation Settings

Three different sets of simulations were performed. The first set explored Type 1 errors where both groups were drawn from the same distribution, satisfying the null hypotheses of no difference between the two groups. These are true Type 1 errors for tests examined in this chapter and in Chapter 3. For the second set of simulations, there were dissonant effects that canceled each other out, such that for the 1 df mean-based tests these setting are considered to be at a null hypothesis. However, for the 2 df tests, these are settings not under the null hypothesis expressed in Equation 2.5 and so we will present the power results. In the third set of simulations, power results were explored where the effects are either consonant or only existing in one part of the model.

All settings for ZIG and ZILN were matched by means and coefficients of variation. Doing so required creating settings such that $\sigma^2 = ln(\nu^{-1} + 1)$ and intercept terms for the continuous part of the model such that e^{τ_0} for ZIG was equal to the $e^{\tau_0+\sigma^2}$ for the related ZILN analyses. The details of these settings are laid out in

Tables 2.1, 2.2, and 2.3. These tables introduce the labels which will be used in later plots. They also outline the conditional means of the two groups (μ_0 and μ_1), the probability of a non-zero outcome for each group (p_0 and p_1), the overall mean for each group ($\mu_0 p_0 = M_0$ and $\mu_1 p_1 = M_1$), and the β_1 and τ_1 effects needed to achieve those means and probabilities. For each setting of β, τ, and ν^{-1}, sample sizes of n=50, 100, and 200 were used with equal sample sizes in each group and 10,000 data sets in each simulation.

Label	μ_0	p_0	M_0	μ_1	p_1	M_1	e^{β_1}	e^{τ_1}
A	1	0.75	0.75	1	0.75	0.75	1	1
B	1.5	0.5	0.75	1.5	0.5	0.75	1	1
C	3	0.25	0.75	3	0.25	0.75	1	1

Table 2.1: Null settings for two group comparisons 'without covariate adjustment'.

The three settings used in Table 2.1 explore Type 1 error rates. Within each setting the data for both groups are simulated from identical distributions. Each setting has an overall mean of 0.75 but different probabilities and conditional means. For each of the conditional mean and probability combinations presented in the table, three settings were used for the dispersion parameter $\nu^{-1} = 0.5, 1, 2$ and three samples sizes were employed with n per group equaling 25, 50, and 100. Figure 2.1 shows the distributions from which the non-zero data were drawn.

The group of setting in Table 2.2 were used to examine dissonant effects which cancel out when observing the overall mean; that is, situations where the overall group means of the two groups are equal, but are arrived at through different probabilities and conditional means. These settings are derived from the settings in Table 2.1 with one group having the conditional means and probabilities from one of the settings in that table, and the other group having the conditional means and probabilities from

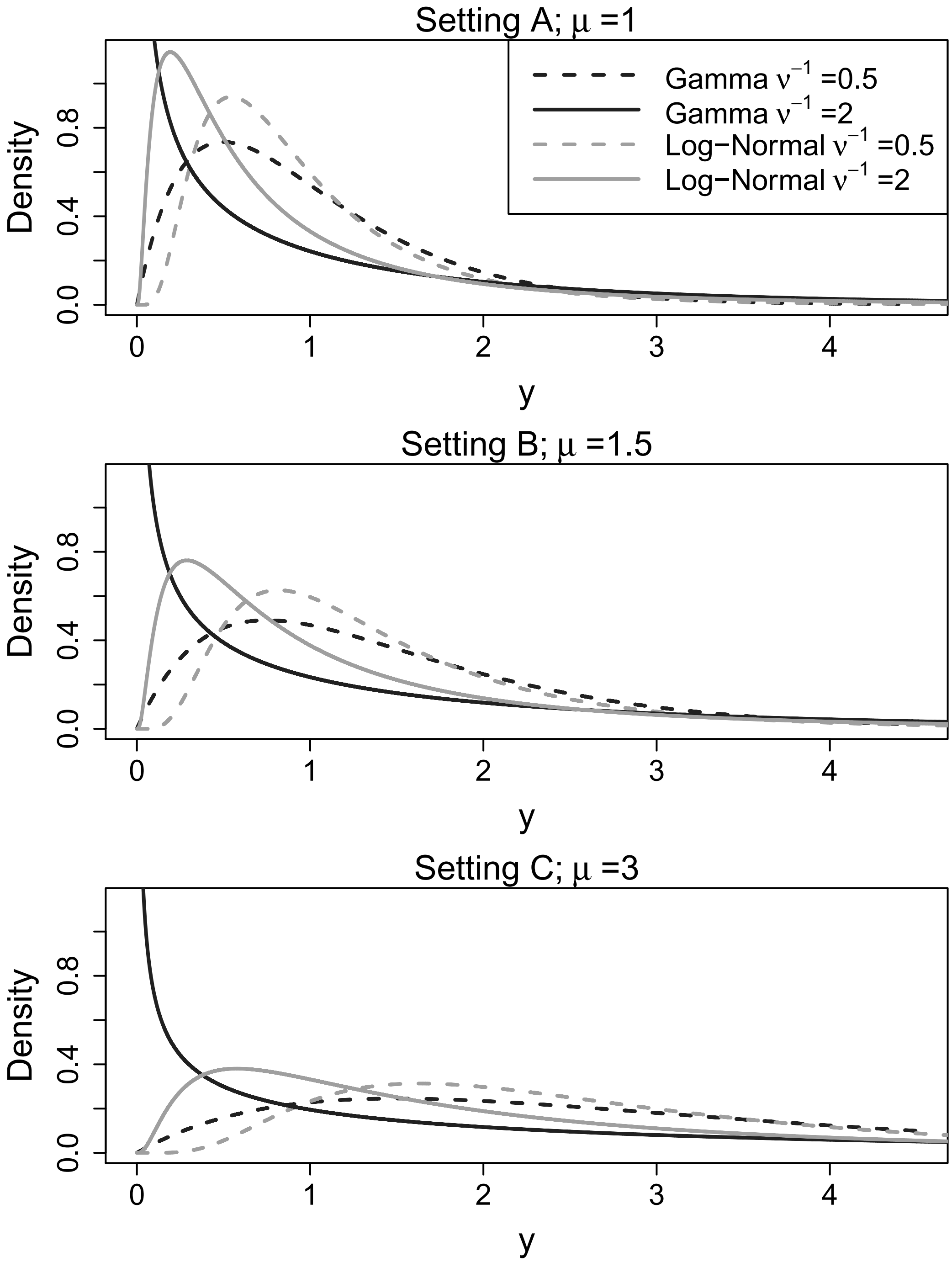

Figure 2.1: Distribution of positive y values from null settings; examples shown at $\nu^{-1} = 0.5$ and $\nu^{-1} = 2$.

another line of that table. For example, for group 0 in setting 'b' μ_0 and p_0 are the same as those for setting 'A' from Table 2.1 and for group 1 in setting 'b' μ_1 and

Comparison[1]	Label	μ_0	p_0	M_0	μ_1	p_1	M_1	e^{β_1}	e^{τ_1}
A vs B	b	1	0.75	0.75	1.5	0.5	0.75	1/3	1.5
A vs C	c	1	0.75	0.75	3	0.25	0.75	1/9	3
B vs C	d	1.5	0.5	0.75	3	0.25	0.75	1/3	2

Table 2.2: Null settings with dissonant effects and equal group means; two group comparisons 'without covariate adjustment'.

[1] Comparison settings come from Table 2.1; e.g., for setting 'b', group 0 has the probabilities and conditional means from setting 'A' in Table 2.1 while group 1 has the probabilities and conditional means from setting 'B'.

p_1 are the same as those for setting 'B' from Table 2.1. The β_1 and τ_1 effects which establish these μ_j and p_j values all differ from zero, but those effects are in opposite directions.

Finally, Table 2.3 outlines simulation settings designed to examine the power to detect group differences at various level of consonant effects coming from one or both parts of the model. For these settings group 0 was given the same distribution as the settings described in Table 2.1. The labels used in Table 2.3 and the resultant plots relate to the settings from which group 0 data were simulated. β_1 and τ_1 were defined such that the overall means for group 1 were $M_1 = 1.08, 1.47$. These group 1 means were achieved in one of three ways, namely varying only β_1, varying only τ_1 or through varying both β_1 and τ_1 to the extent that they each had an equivalent multiplicative effect on the overall mean. Figure 2.2 shows the distributions of positive y values at a few example settings, particularly those with group 0 at setting 'B' and only τ_1 varied. When group 0 came from setting 'A', the desired mean of group 1 could not be attained by changing only β_1 for either level of M_1; also an overall mean of $M_1 = 1.47$ could not be attained by changing β_1 and τ_1 to equivalent extents. All other settings were possible. However, when group 0 was from setting 'B', changing

	Group 0 Setting[1]	M_1	μ_1	p_1	e^{β_1}	e^{τ_1}
β_1 Varied	B	1.08	1.5	0.72	2 4/7	1
	B[2]	1.47	1.5	0.98	49	1
	C	1.08	3	0.36	1 11/16	1
	C	1.47	3	0.49	2 15/17	1
τ_1 Varied	A	1.08	1.44	0.75	1	1.44
	A	1.47	1.96	0.75	1	1.96
	B	1.08	2.16	0.5	1	1.44
	B	1.47	2.94	0.5	1	1.96
	C	1.08	4.32	0.25	1	1.44
	C	1.47	5.88	0.25	1	1.96
β_1 and τ_1 Varied	A	1.08	1.2	0.9	3	1.2
	B	1.47	1.8	0.6	1.5	1.2
	B	1.08	2.1	0.7	2 1/3	1.4
	C	1.08	3.6	0.3	1 2/7	1.2
	C	1.47	4.2	0.35	1 8/13	1.4

Table 2.3: Settings for power analyses; two group comparisons 'without covariate adjustment'.

[1] For group 0 means and probabilities, see corresponding letters in Table 2.1.

[2] Settings for which $< 99\%$ of data sets had inestimable parameters.

only β_1 to obtain a group 1 mean of $M_1 = 1.47$ lead to such a high probability of non-zeros that for more than 1% of the data sets simulated, some or all of the parameters in two-part models could not be estimated. This setting will, therefore, be excluded from plots and discussions.

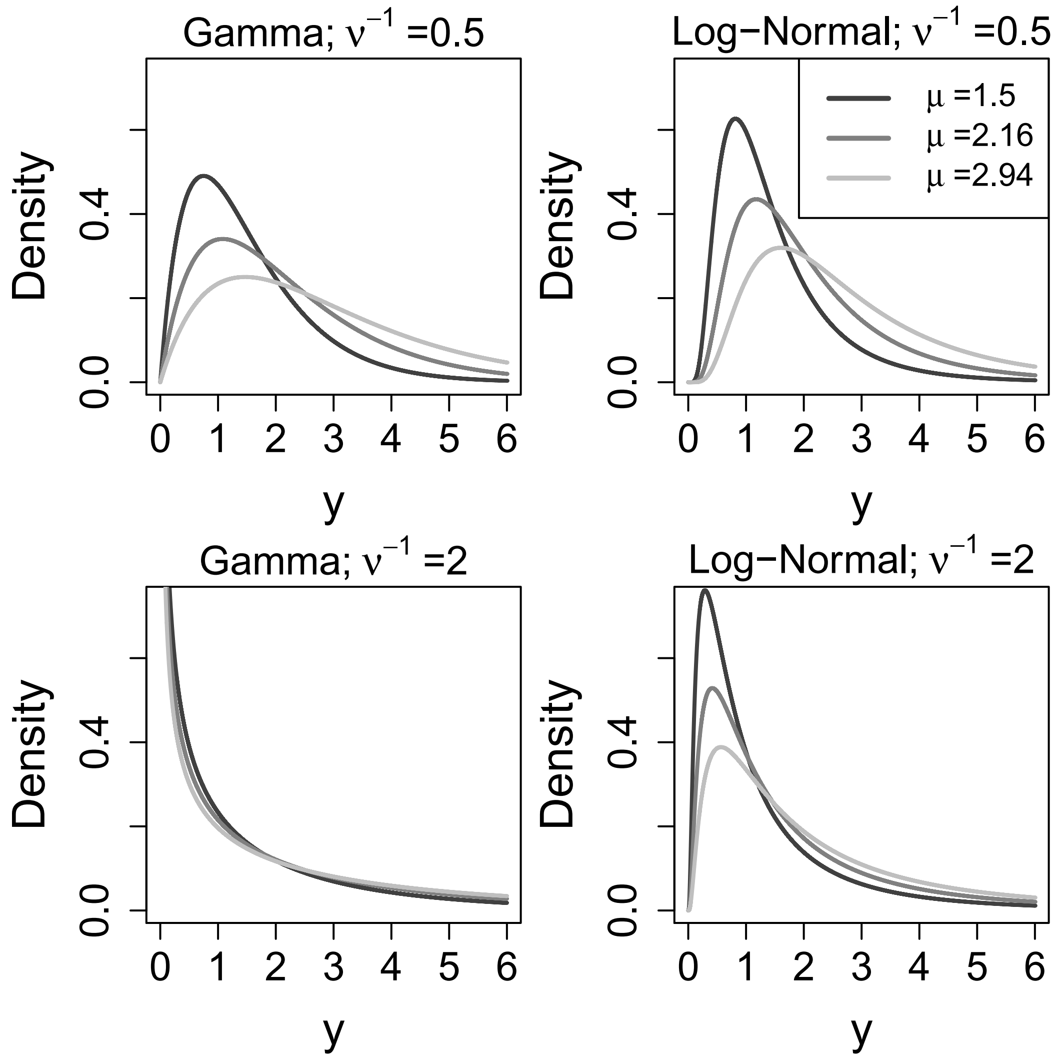

Figure 2.2: Distribution of positive y values from alternative settings with null at setting B and τ_1 changing; examples shown at $\nu^{-1} = 0.5$ and $\nu^{-1} = 2$.

2.7 Simulation Results for Two-Part Tests

2.7.1 Type 1 Error Under General Null Setting

Figures 2.3 - 2.6 illustrate the results of the set of simulations described in Table 2.1. These figures show the observed Type 1 error rates when a nominal Type 1 error rate of 0.05 was assumed. As noted in that table, the setting are labeled 'A' for $P(Y > 0) = .75$ and $E(Y|Y > 0) = 1$, labeled 'B' for $P(Y > 0) = .5$ and $E(Y|Y > 0) = 1.5$, and labeled 'C' for $P(Y > 0) = .25$ and $E(Y|Y > 0) = 3$.

Data simulated from ZIG distributions are presented in the graphs on the left side of the figure and data simulated from ZILN distributions are on the right side of the figure. ZIG analyses are shown with dark grey labels and dotted lines while ZILN analyses are shown with light grey labels and dashed lines. A solid black line shows the nominal level of 0.05 that was used for these analyses with dotted lines at 0.0457 and 0.0543 reflecting $\pm$ 1.96 standard errors of a binomial distributions when p=0.5 and 10,000 different data sets are run. Thus, any observed Type 1 error rates that are outside these bounds are different than would be expected if the Type 1 error were truly equal to the nominal 0.5 level.

Figure 2.3 shows the observed Type 1 error rates when the two-part Wald test described in Section 2.5.1 is used to analyze the outcome. The effects of correct classification and misclassification on Type 1 error differ depending on the distribution from which the data were simulated. The plots on the left show that when the data were simulated as ZIG that for larger sample sizes, most analyses assuming ZIG (dark grey) have slightly conservative Type 1 error rates. The level of conservativeness increases as the coefficient of variation increases. When the sample size was extremely small, n=50 with $P(Y > 0) = 0.25$, the 2 df Wald tests Type 1 error were elevated for ZIG data regardless of distribution assumed. The plots on the right which show results for data simulated as ZILN, demonstrate that when ZIG is falsely assumed to be the distribution Type 1 errors are quite high. This is especially true as ν^{-1} increases, which is likely due to the fact that ZIG and ZILN distributions differ more with large coefficients of variation (CoV) than they do with smaller CoV. When ZILN data are analyzed as ZIG, Type 1 error increases with sample size for large ν^{-1}, but decreases with sample size for smaller values of ν^{-1}. Wald tests for ZILN data correctly modeled as ZILN have slightly high Type 1 error rates when sample size is small, but this improves to near the nominal level as sample size increases.

The Type 1 error results for the 2-part score test defined in Equation 2.15 are

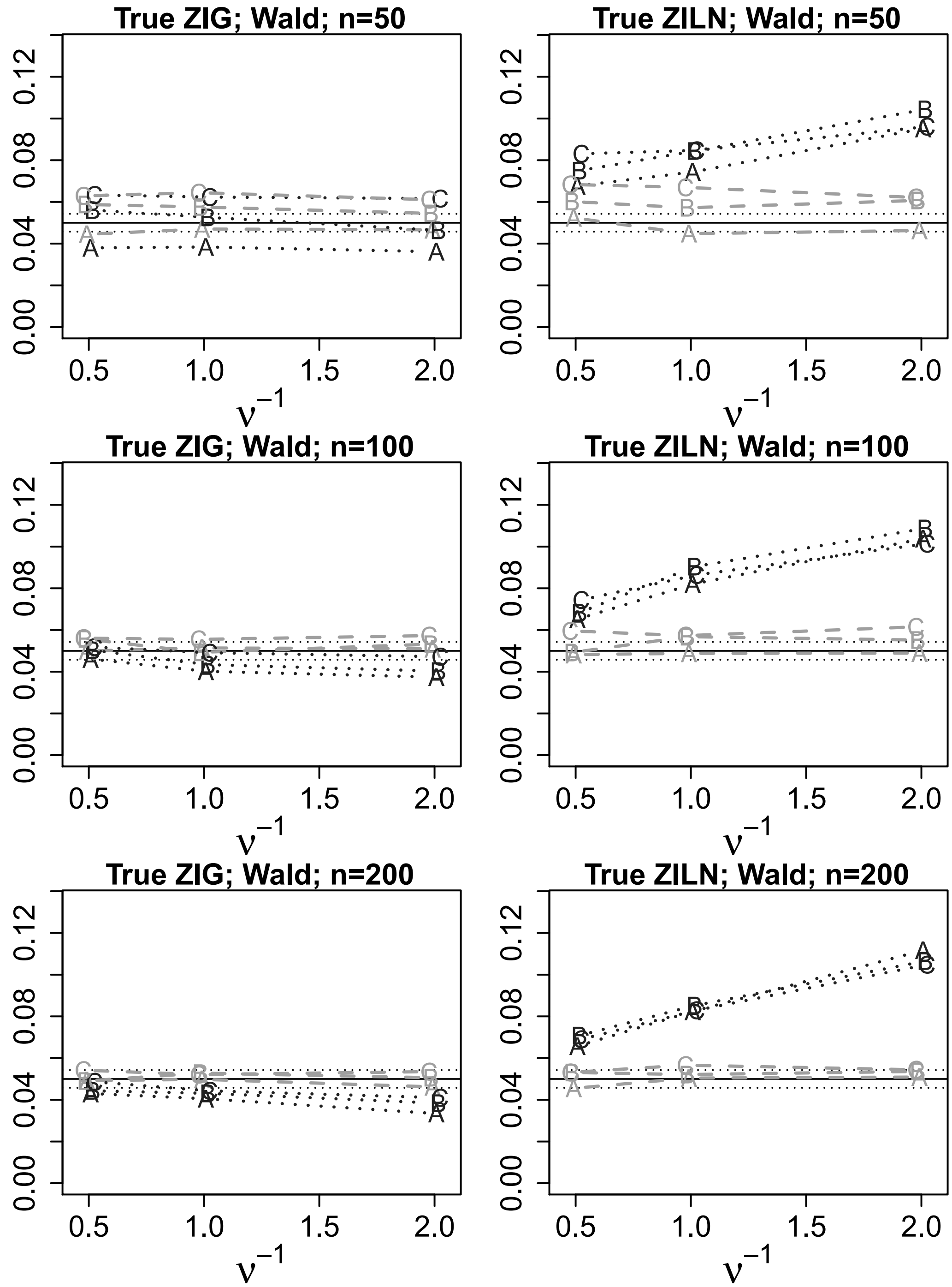

Figure 2.3: Type 1 error for 2 df Wald tests; two groups comparisons 'without covariate adjustment'. Symbols correspond to the setting in Table 2.1 where 'A' is used when $P(Y > 0) = 0.75$, 'B' when $P(Y > 0) = 0.5$, and 'C' when $P(Y > 0) = 0.25$. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

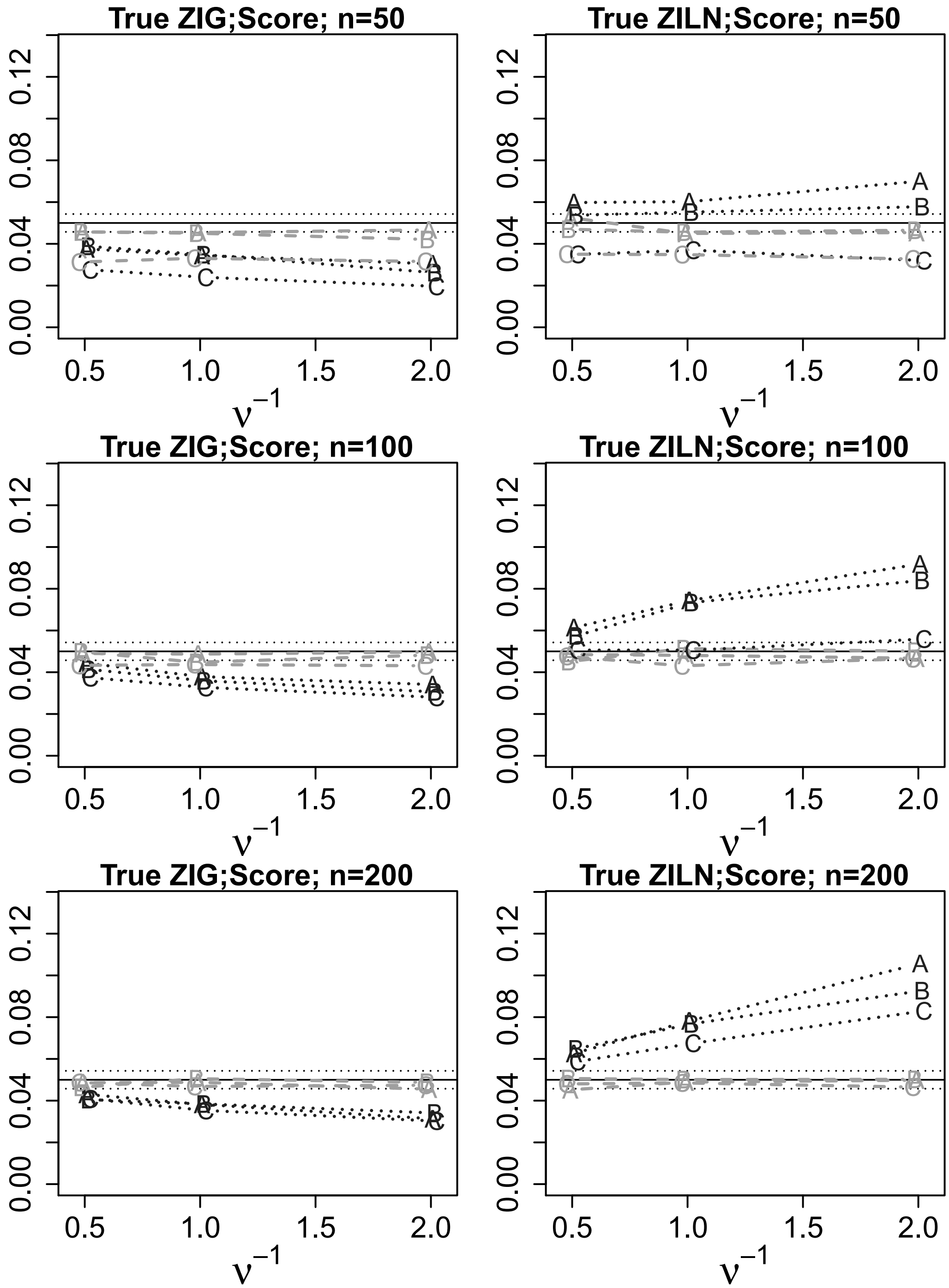

Figure 2.4: Type 1 error for 2 df score tests; two groups comparisons 'without covariate adjustment'. Symbols correspond to the setting in Table 2.1 where 'A' is used when $P(Y > 0) = 0.75$, 'B' when $P(Y > 0) = 0.5$, and 'C' when $P(Y > 0) = 0.25$. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

shown in Figure 2.4. These tests were performed on the same data sets that were used for the other figures in this section. Type 1 error results from data simulated as ZIG are shown in the left hand set of graphs. When the data are analyzed as ZIG, Type 1 errors are conservative and become increasingly so as ν^{-1} increases. Those same data analyzed assuming a ZILN distribution results in slightly conservative Type 1 errors when sample sizes are small, but are at appropriate levels when sample sizes are larger. The right-hand plots show that when ZILN data are properly analyzed assuming ZILN distributions, Type 1 errors are within the appropriate range for sample sizes of n=100 and n=200. However, when n=50, Type 1 errors are slightly conservative for setting 'C' where $P(Y > 0) = 0.25$ can lead to very small samples for the continuous part of the model. On the other hand, when ZILN data are misspecified as ZIG, Type 1 errors are above nominal. Type 1 errors for ZILN data falsely analyzed as ZIG increase with ν^{-1}. They also increase both with overall sample size and with the expected number of non-zeros (e.g. setting 'A' with $P(Y > 0) = .75$ results in larger non-zero samples and higher Type 1 errors than setting 'C' where $P(Y > 0) = .25$). The only exception to the high Type 1 errors for ZILN analyzed as ZIG is when the expected sample size of the continuous part is small (setting 'C' with n=50) where Type 1 errors for ZILN analyzed as ZIG are actually somewhat conservative.

Type 1 error rates using the likelihood ratio tests (LRT) defined in Equation 2.16 for the set of simulations described in Table 2.1 are shown in Figure 2.5. Data simulated from a ZIG distribution (left graphs) lead to Type 1 error rates for LRT that are slightly elevated when sample sizes are small, but are at appropriate levels for larger sample sizes. ZIG and ZILN analyses are very close in Type 1 error, with only a slightly smaller Type 1 error in ZIG analyses when sample size is large, however, both are essentially within the appropriate range so these slight differences may be meaningless. Data simulated from ZILN distributions (right graphs) and correctly analyzed as ZILN lead to slightly elevated Type 1 error rates when sample size is small, but have appropriate Type 1 error rates for larger sample sizes. When that same data are analyzed as ZIG, similar patterns are seen as were seen in Wald and score tests. Namely, Type 1 errors are high, worsen with increase in ν^{-1}, improve slightly with sample size when $\nu^{-1} = 0.5$ and worsen with sample size when $\nu^{-1} = 2$.

Figure 2.6 shows Type 1 error rates for the Wald-Wilcoxon 2-part test described in Equation 2.17. The same simulations were used for these analyses as was used for those described above. Type 1 error rates for the Wald-Wilcoxon test are slightly conservative when sample size is small regardless of the true simulation distribution and the assumed distribution. When sample sizes are larger, the Wald-Wilcoxon tests yield Type 1 errors that are within the expected range.

Figure 2.7 shows the Type 1 error rates for each of the tests in the previous plots for one example mean setting, setting 'B' where $P(Y > 0) = .5$ and $E(Y|Y > 0) = 1.5$. The labels used in this plot refer to the 2 df test used to analyze the data and create the Type 1 error results. Specifically, the label 'W' refers to the Type 1 error for the Wald test, 'S' refers to the score test, 'L' refers to the LRT, and 'X' refers to the Wald-Wilcoxon test. As in the previous plots, the darker grey with dotted lines refers to analyses assuming ZIG distributions and the lighter grey with dashed lines refers to analyses assuming a ZILN distribution. The Wald-Wilcoxon Type 1 error rates are shown in a medium shade of grey connect by a lines with short dashes.

The two main discrepancies in relation to the nominal Type 1 error rates are first, overly conservative observed Type 1 errors when ZIG data are analyzed as ZIG and second, excessively high Type 1 errors when ZILN data are analyzed as ZIG. When the data are ZIG and analyzed as ZIG, the Wald and score tests are the most conservative, with the LRT having Type 1 errors closer to nominal (albeit being slightly high when the sample size is small). For the more extreme situation of high Type 1 errors when ZILN data are analyzed as ZIG, the LRTs are the worst offenders

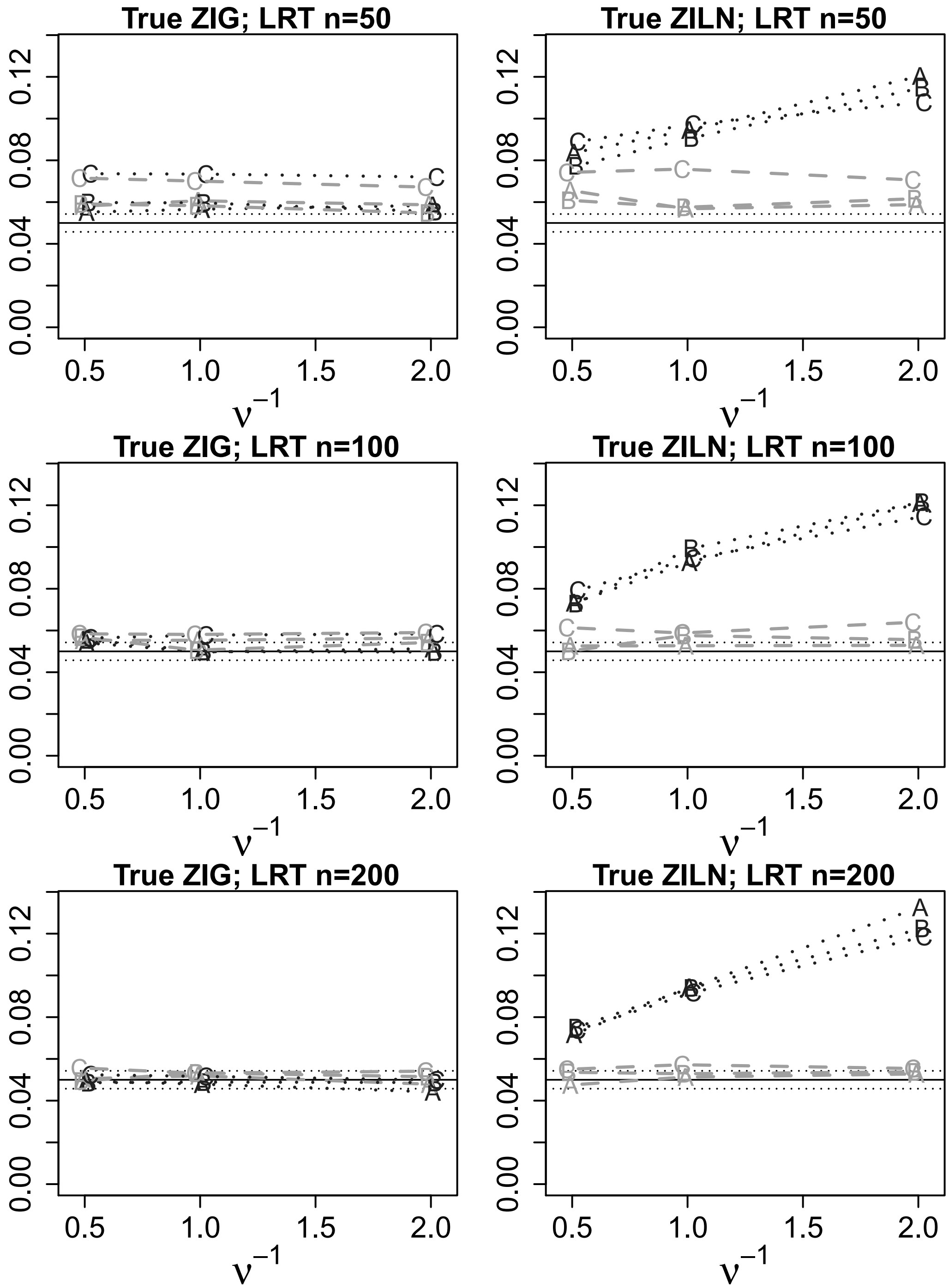

Figure 2.5: Type 1 error for 2 df likelihood ratio tests; two groups comparisons 'without covariate adjustment'. Symbols correspond to the setting in Table 2.1 where 'A' is used when $P(Y > 0) = 0.75$, 'B' when $P(Y > 0) = 0.5$, and 'C' when $P(Y > 0) = 0.25$. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

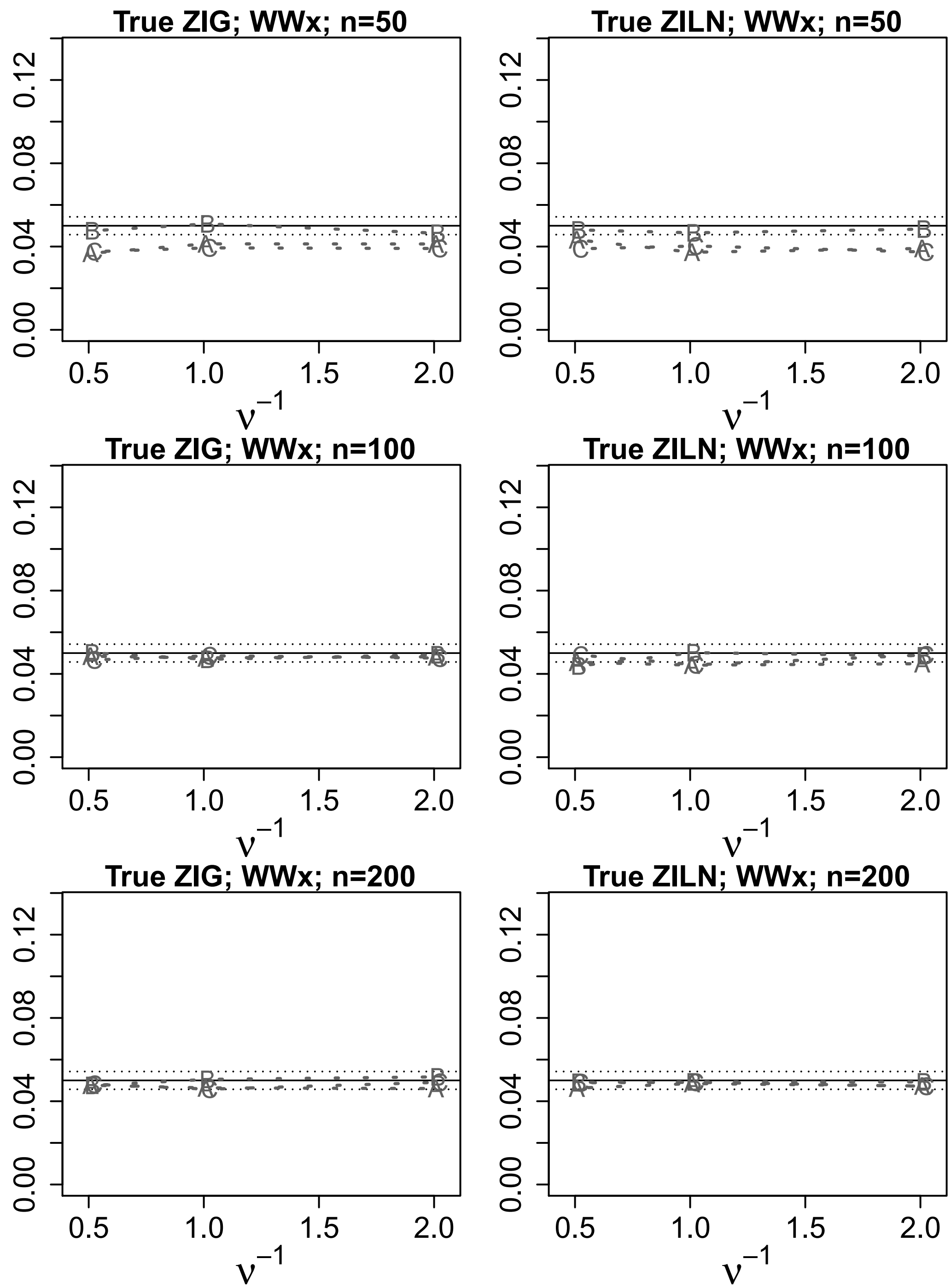

Figure 2.6: Type 1 error for 2 df Wald-Wilcoxon tests. Symbols correspond to the setting in Table 2.1 where 'A' is used when $P(Y > 0) = 0.75$, 'B' when $P(Y > 0) = 0.5$, and 'C' when $P(Y > 0) = 0.25$.

having the highest Type 1 errors for all sample sizes, the Wald tests had the next highest Type 1 error rates, followed by the score tests. The score and Wald tests assuming ZILN, and the Wald-Wilcoxon tests all had appropriate Type 1 error rates.

2.7.2 Power for Dissonant Effects With Equal Group Means

In Table 2.2 three settings were outlined where both groups had the same overall mean, but had different conditional means and different probabilities of non-zero outcomes. A researcher more interested in finding differences based on a two-part model (e.g. looking for differences in the probability of non-zero values and/or differences in the conditional means of the two groups) than in a difference based on the overall group means, could use the two-part tests used in this chapter to find such differences. Figures 2.8 - 2.11 report on the power for finding these differences given the settings in Table 2.2. The labels in the graphs correspond to the labels shown in Table 2.2; 'b' refers to setting where one group has $P(Y > 0) = 0.75$ and the other has $P(Y > 0) = 0.5$, setting 'c' has group 0 with $P(Y > 0) = 0.75$ and group 1 with $P(Y > 0) = 0.25$, and 'd' has one group with $P(Y > 0) = 0.5$ and the other has $P(Y > 0) = 0.25$. These have conditional means that differ between groups in the opposite direction as the probabilities of the two groups differ with the settings calculated in such a way that the two groups still have equivalent overall means. See Table 2.2 for more details. As in previous plots, power for data analyzed as ZIG are shown by dark grey with dotted lines and power for data analyzed as ZILN as shown by light grey with dashed lines.

Figure 2.8 shows the power for finding dissonant effects for different levels of ν^{-1}. From these it is clear that power decreases as ν^{-1} increases. Setting 'c' has the highest powers, having power>0.9 even for n=50; this makes sense as setting 'c' has the most extreme values for both β_1 and τ_1. Settings 'd' and 'b' have equal β_1 effects but differ in terms of τ_1; setting 'd' has higher τ_1 and consequently has a slightly

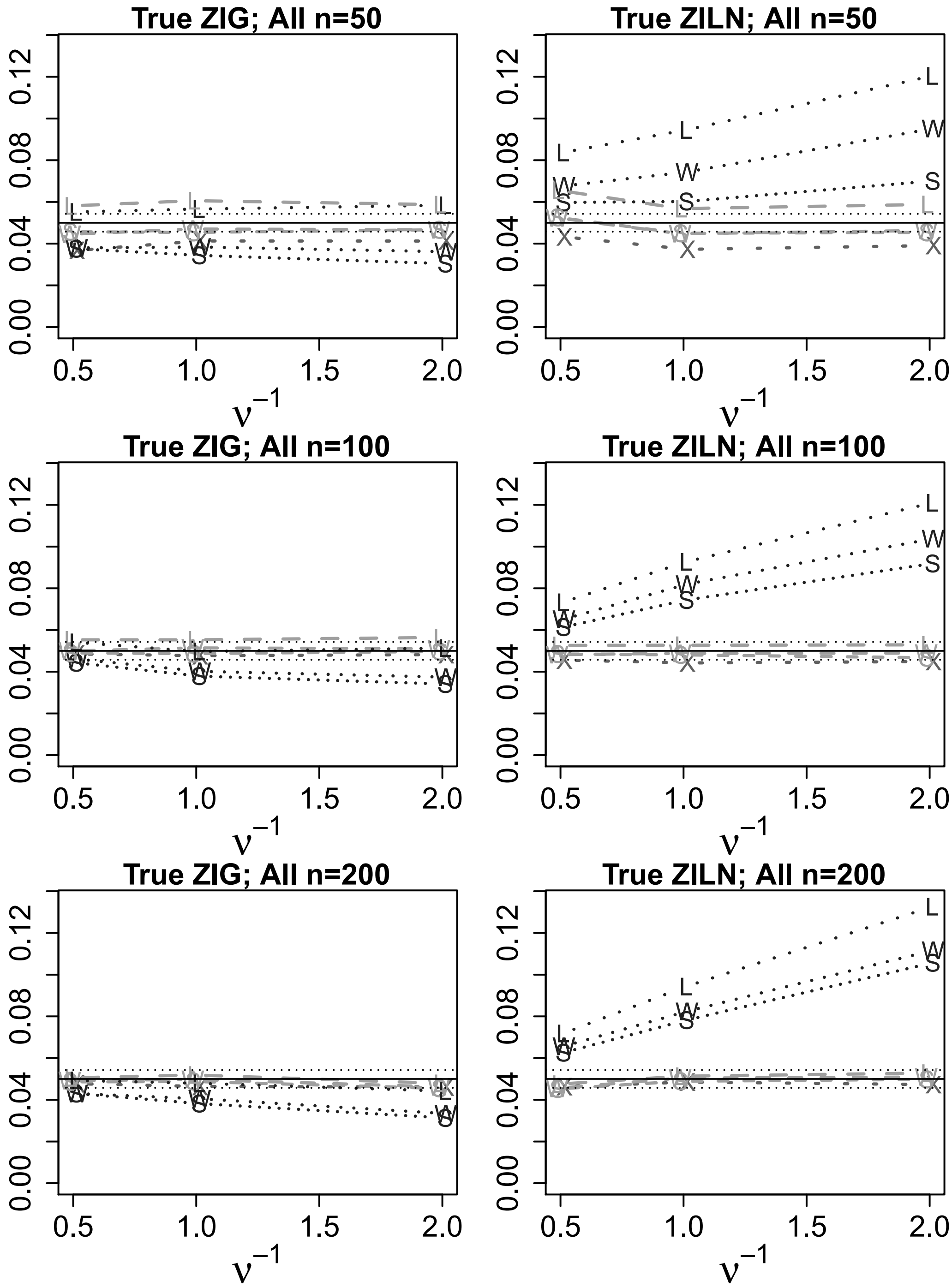

Figure 2.7: Type 1 error for all 2 df tests; no covariate adjusting. Labels represent the 2 df tests used; 'W' is for the Wald tests, 'S' refers to the score tests, 'L' refers to the LRTs, and 'X' refers to the Wald-Wilcoxon tests. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

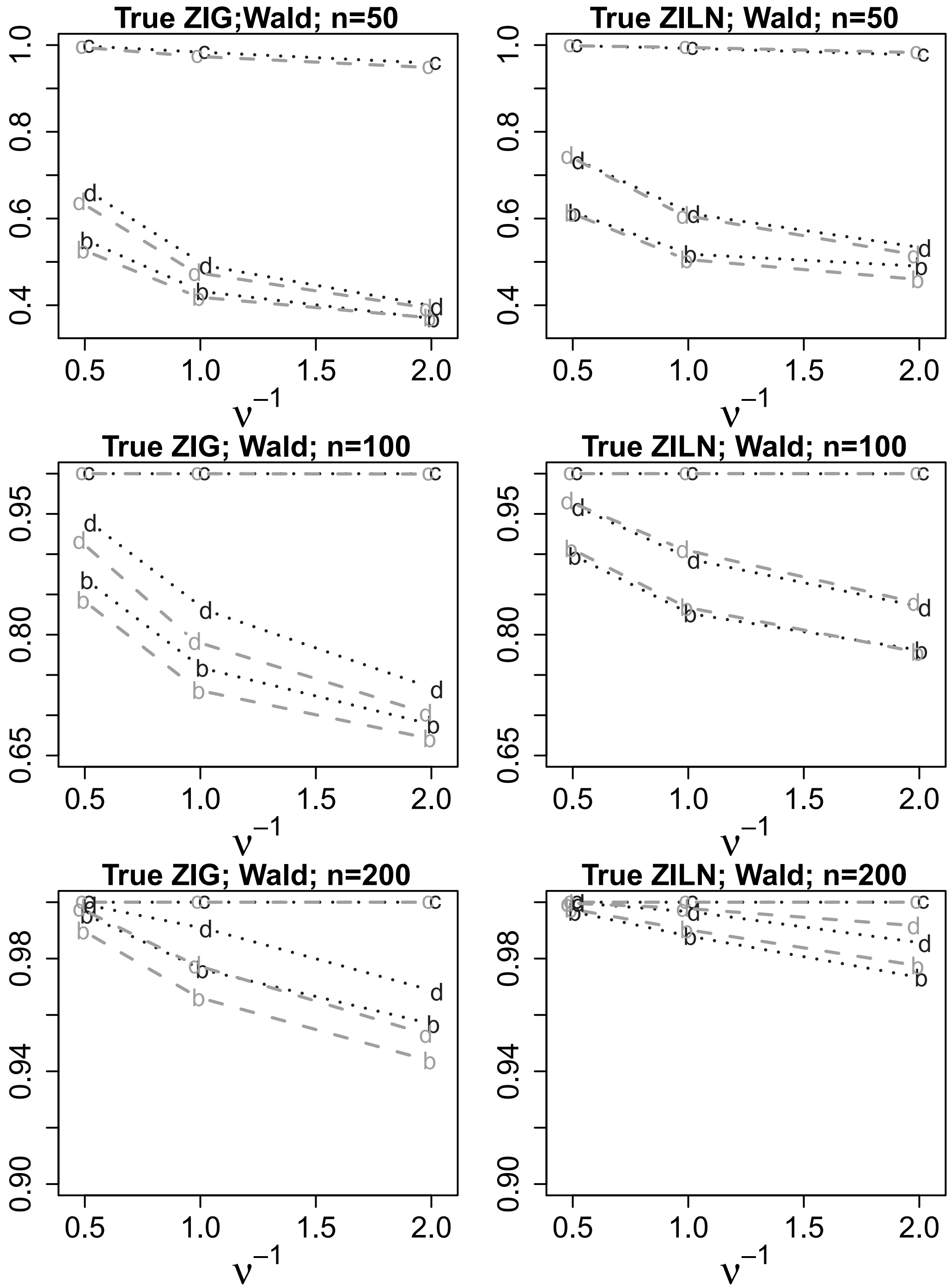

Figure 2.8: Power for dissonant effects and equal group means using a 2 df Wald test; no covariate adjusting. Labels refer to the settings outlined in Table 2.2; for setting 'b' $P(Y > 0|group = 0) = 0.75$ and $P(Y > 0|group = 1) = 0.5$, for setting 'c' $P(Y > 0|group = 0) = 0.75$ and $P(Y > 0|group = 1) = 0.25$, and for setting 'v' $P(Y > 0|group = 0) = 0.5$ and $P(Y > 0|group = 1) = 0.25$. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

higher power even though the $P(Y > 0)$ is smaller for that setting. The power values for the Wald test simulation results shown in this figure show that when the data are truly ZIG, ZIG analyses have higher power than ZILN analyses. When the simulated data are truly ZILN, the powers are much closer, but ZILN slightly outperforms ZIG. It must be noted, however, that when the data are truly ZILN, the ZIG analyses at higher levels of ν^{-1} cannot be trusted because of the high Type 1 errors observed in the previous section.

Figures 2.9 and 2.10 show the power results for the 2 df score and likelihood ratio tests respectively. These show patterns very similar to the Wald test power results. Again, power decrease with ν^{-1}, power increases with the strength of β_1 and τ_1, when the data are truly ZIG, ZIG analyses outperform ZILN analyses. One result which is distinct from those seen for the Wald tests is that for score and LRT power results when n=50 and the data are truly ZILN, ZIG analyses have higher power that ZILN analyses. However, because of high Type 1 errors for misspecified ZIG analyses, ZILN analyses would still be more appropriate.

Figure 2.11 show the power results for this same set of simulations for Wald-Wilcoxon test. With this test, as with those previously examined, power decrease with ν^{-1}, and power increases with the strength of β_1 and τ_1.

Figures 2.12 and 2.13 show a comparison across tests at settings 'b' and 'd' respectively. As in Figure 2.7, 'W' refers to the Type 1 error for the Wald test, 'S' for the score test, 'L' for the LRT, and 'X' for the Wald-Wilcoxon test. Darker grey symbols with dotted lines refer to ZIG-based analyses (that is, analyses assuming ZIG distributions). Lighter grey symbols with dashed lines refer ZILN-based distributions (that is, analyses assuming a ZILN distribution). And, Wald-Wilcoxon results are connected by a medium gray short-dashed line. Among the parametric tests, LRTs have the highest powers when the data are ZIG or the sample size is n=50. For setting 'b' shown in Figure 2.12, Wald and score tests have almost identical power. For setting

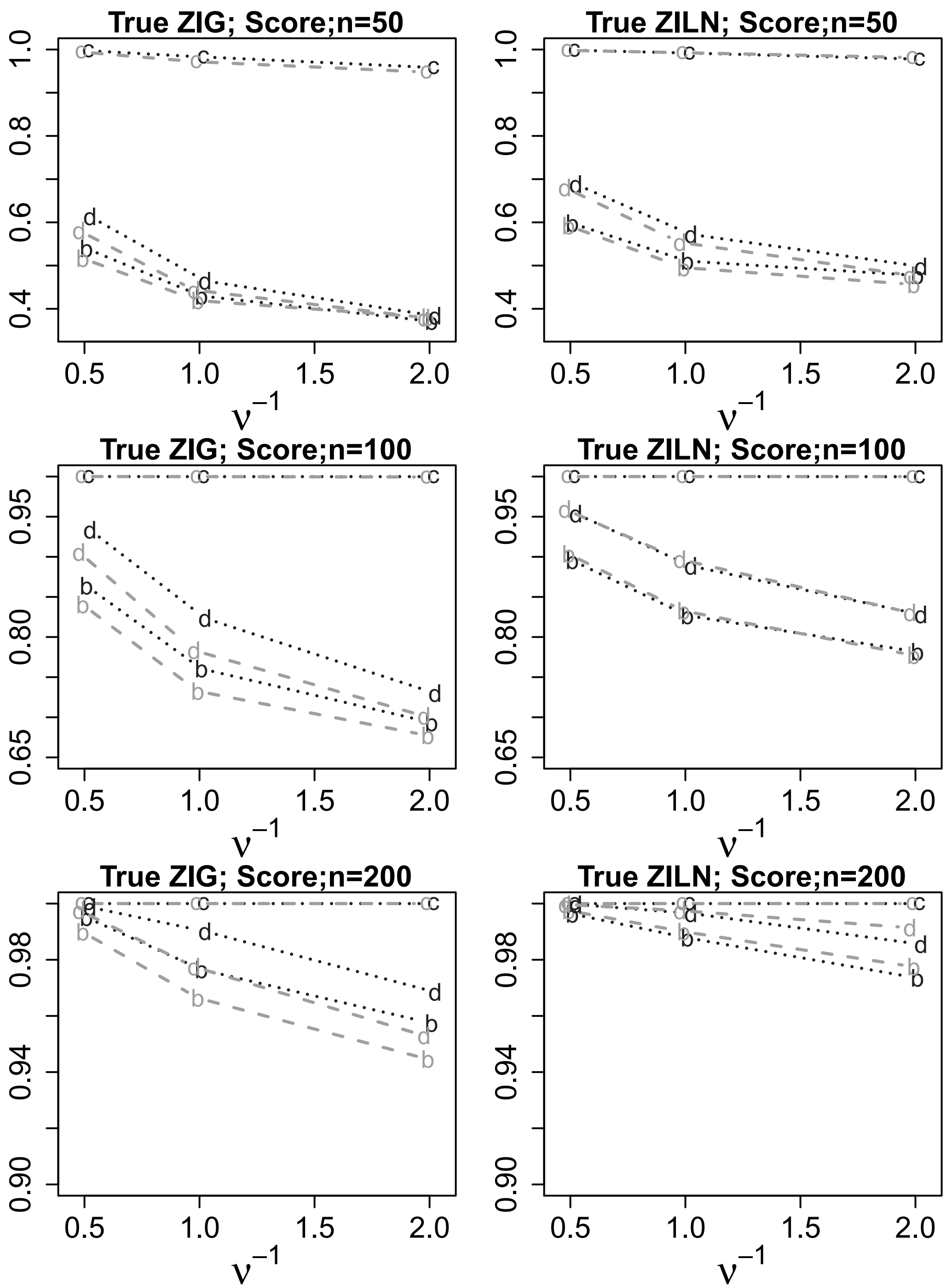

Figure 2.9: Power for dissonant effects and equal group means using a 2 df score test; no covariate adjusting. Labels refer to the settings outlined in Table 2.2; for setting 'b' $P(Y > 0|group = 0) = 0.75$ and $P(Y > 0|group = 1) = 0.5$, for setting 'c' $P(Y > 0|group = 0) = 0.75$ and $P(Y > 0|group = 1) = 0.25$, and for setting 'd' $P(Y > 0|group = 0) = 0.5$ and $P(Y > 0|group = 1) = 0.25$. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

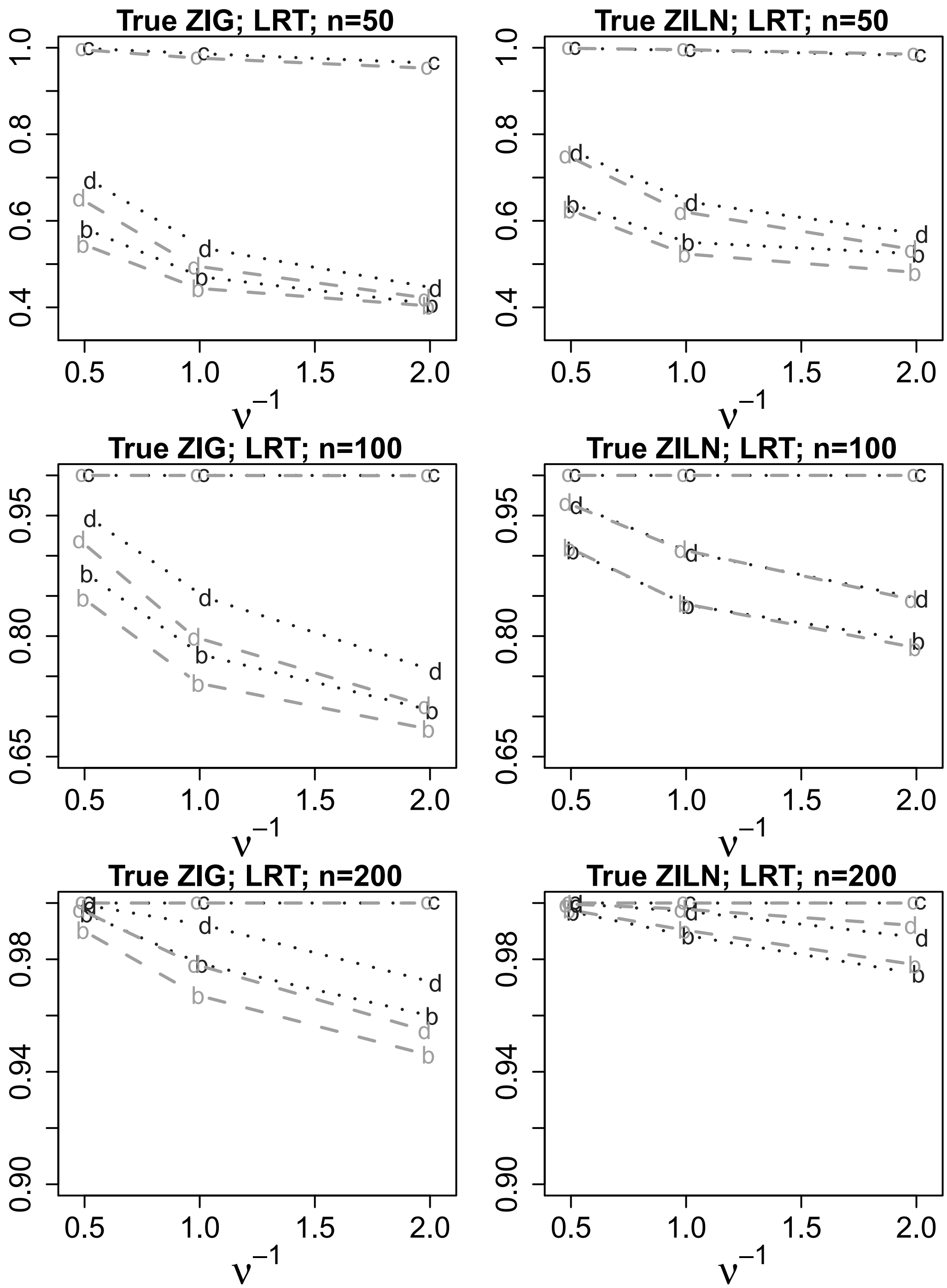

Figure 2.10: Power for dissonant effects and equal group means using a 2 df likelihood ratio test; no covariate adjusting. Labels refer to the settings outlined in Table 2.2; for setting ‘b’ $P(Y > 0|group = 0) = 0.75$ and $P(Y > 0|group = 1) = 0.5$, for setting ‘c’ $P(Y > 0|group = 0) = 0.75$ and $P(Y > 0|group = 1) = 0.25$, and for setting ‘d’ $P(Y > 0|group = 0) = 0.5$ and $P(Y > 0|group = 1) = 0.25$. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

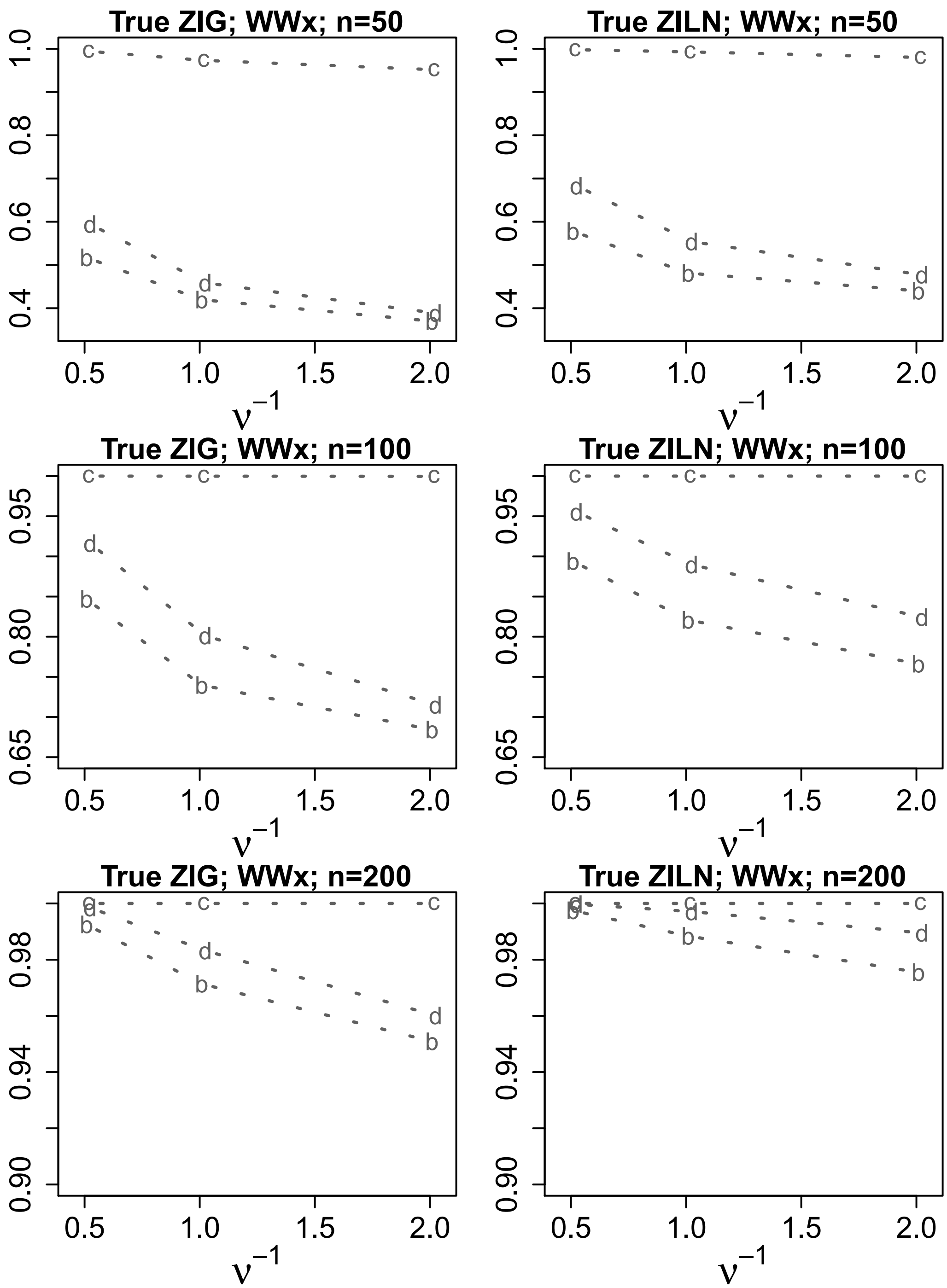

Figure 2.11: Power for dissonant effects and equal group means using a 2 df Wald-Wilcoxon test. Labels refer to the settings outlined in Table 2.2; for setting ‘b’ $P(Y > 0|group = 0) = 0.75$ and $P(Y > 0|group = 1) = 0.5$, for setting ‘c’ $P(Y > 0|group = 0) = 0.75$ and $P(Y > 0|group = 1) = 0.25$, and for setting ‘d’ $P(Y > 0|group = 0) = 0.5$ and $P(Y > 0|group = 1) = 0.25$.

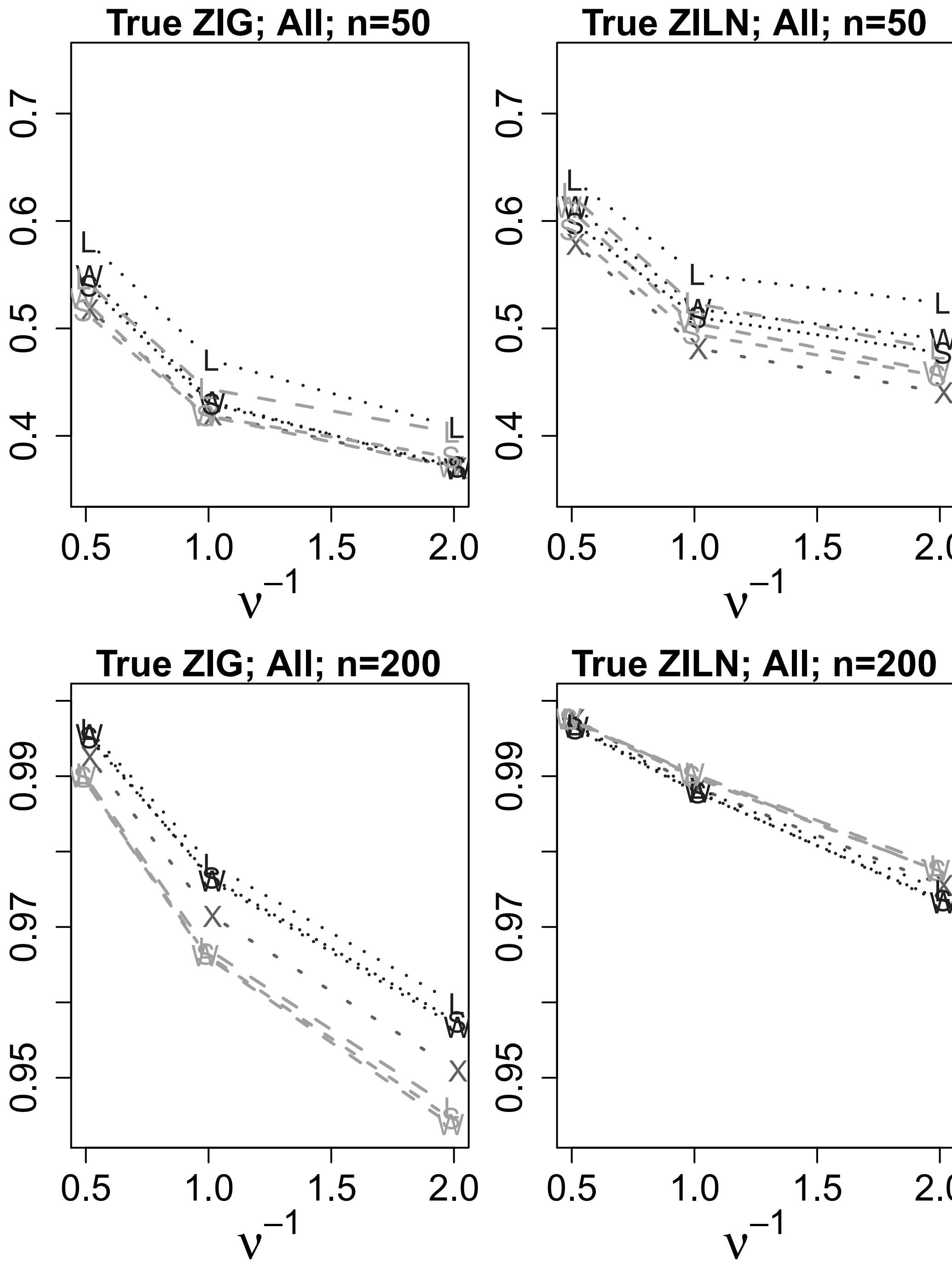

Figure 2.12: Comparing power for dissonant effects with equal group means across all 2 df tests; setting 'b'. Labels represent the 2 df tests used; 'W' is for the Wald tests, 'S' refers to the score tests, 'L' refers to the LRTs, and 'X' refers to the Wald-Wilcoxon tests. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

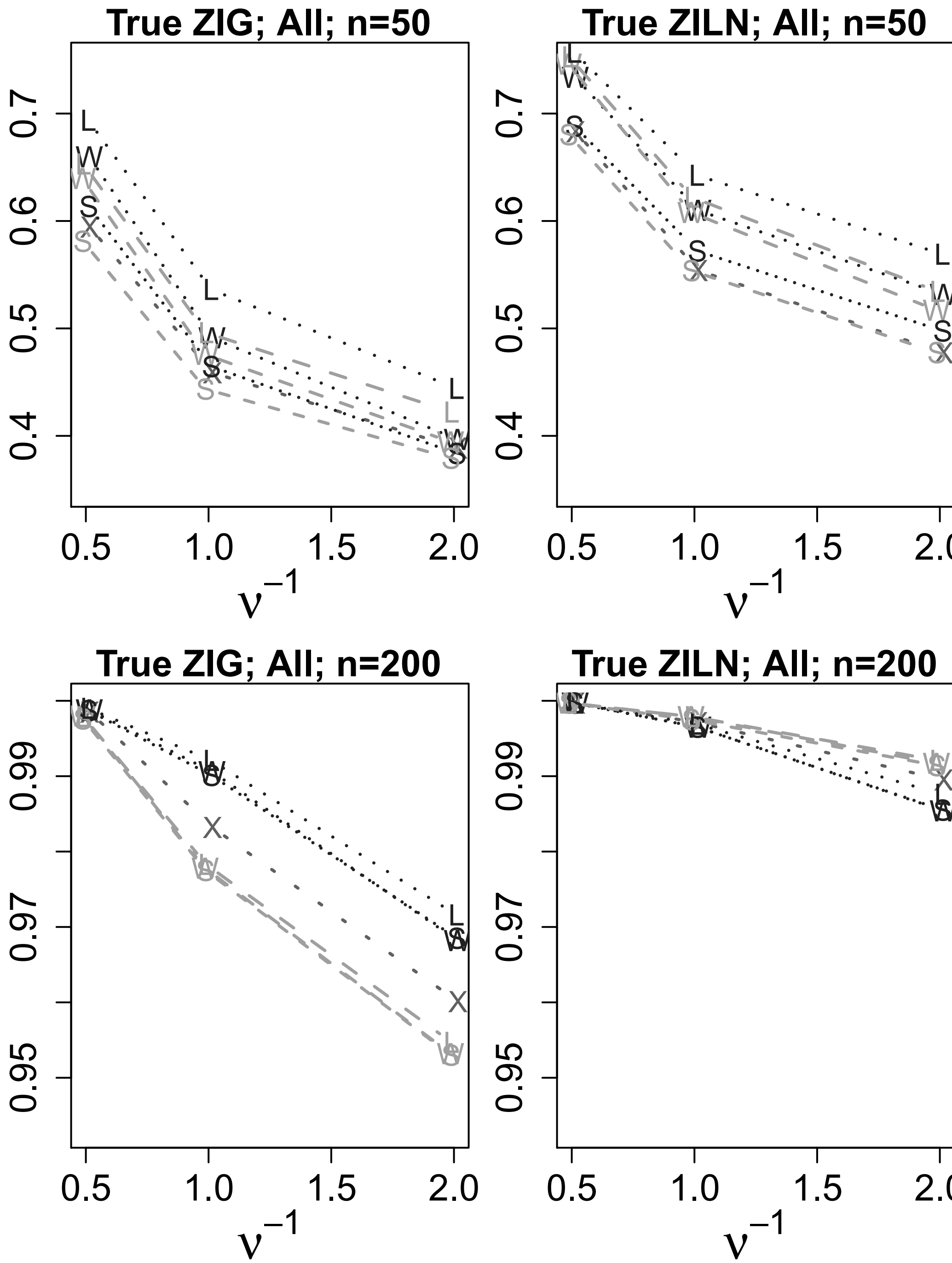

Figure 2.13: Comparing power for dissonant effects with equal group means across all 2 df tests; setting 'd'. Labels represent the 2 df tests used; W is for the Wald tests, 'S' refers to the score tests, 'L' refers to the LRTs, and 'X' refers to the Wald-Wilcoxon tests. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

'd', Figure 2.13 the Wald tests have slightly higher powers than the score tests when n=50. One of the clearest results from these simulations is that when sample size is small, the Wald-Wilcoxon tends have low power relative to both correctly specified and misspecified parametric tests (lower for true ZILN when setting is 'b' and for both data types when setting is 'd'). However, when sample size is larger (n=200) the power for the Wald-Wilcoxon is higher than the power when the distribution is misspecified but lower than when it is correctly specified.

2.7.3 Power for Consonant Effects

Figures 2.14 to 2.19 show the power results for the simulation settings outlined in Table 2.3. Group 0 overall means were set at $M_0 = 0.75$ through the same three combinations of p_0 and μ_0 which were used for the Type 1 error results; also β_1 and τ_1 were varied to lead to group 1 means of $M_1 = 1.08$ and $M_1 = 1.47$. Through these settings, we studied the differences in power for similar effects at different levels of zero-inflation. Figure 2.14 and Figure 2.15 show the power results for two-part Wald tests for $\nu^{-1} = 0.5$ and $\nu^{-1} = 2$ respectively. From these figures we can compare the power results across the settings described in Table 2.3. The across setting trends observed for the Wald tests are roughly similar those found in the score tests and LRTs. Because of this, we will not present figures detailing these results separately. Instead, Figure 2.16 - Figure 2.19 show examples of the differences in power between Wald, score, and likelihood ratio tests on group 0 setting for various group 1 settings.

In Figures 2.14 and 2.15, the symbols 'A', 'B', and 'C' denote the settings used for group 0; specifically, 'A' indicates that $p_0 = 0.75$, 'B' that $p_0 = 0.5$, and 'C' that $p_0 = 0.25$. Both the x-axis values and the plot in which the results are contained relate to the settings for group 1. The x-axis shows the overall mean of group 1, $M_1 = 1.08$ or $M_1 = 1.47$. For the top plots, these group 1 means were achieved by varying the value for β_1 only. For the middle plots, the group 1 means are attained by varying the value for τ_1 only. For the bottom plots, β_1 and τ_1 are set in ways that yielded the same multiplicative effect on the group 1 mean. As in other figures, the plots on the right are for data simulated as ZIG and the plots on the right are for data simulated as ZILN.

From Figure 2.14 it is clear that the higher the value of p_0 the higher the observed power for the 2 df Wald test. For example, when the group effect is in the continuous part of the model (τ_1 varied) settings with 'A' group 0 values ($p_0 = 0.75$) had higher powers than settings with 'B' group 0 values ($p_0 = 0.5$) which had higher powers than settings 'C' ($p_0 = 0.25$). From Table 2.3 it can be noted that with a group 1 mean ($M_1 = 1.08$ or $M_1 = 1.47$) the same values were for the effect τ_1 were utilized, for these settings then it is clear that the extent of zero inflation (driven by p_0 and p_1 which are equal in this case) is the reason for the power differences across lettered settings. For the settings where the group 1 means were attained by varying only β_1 the difference in observed power between setting 'B' and setting 'C' when $M_1 = 1.08$ was considerably larger than those observed when τ_1 alone had been varied. The larger difference in power is likely caused by the following factors. First, since the increase in the group 1 mean (relative to the group 0 mean) was attained by increasing the probability of non-zero outcomes, there are larger difference in zero inflation between settings 'B' and 'C' for the β_1 changing scenarios than there were for the τ_1 changing scenarios. Secondly, as p_0 increases, larger effect sizes (larger β_1 values) are needed to obtain the desired the group 1 means by changing only p_1.

When the group 1 overall means are obtained through a combination of β_1 and τ_1 effects, power was lower than when the effects were only in one part of the model for group 0 settings 'B' and 'C', but higher for setting 'A' when the data are truly ZIG. Setting 'A' in these bottom graphs, has the lowest zero-inflation of all of the settings ($p_0 = 0.75$, $p_1 = 0.9$) pictured and also the highest β_1 value; so for setting 'A' the β_1 effect is strong and likely driving the power results. For the other settings,

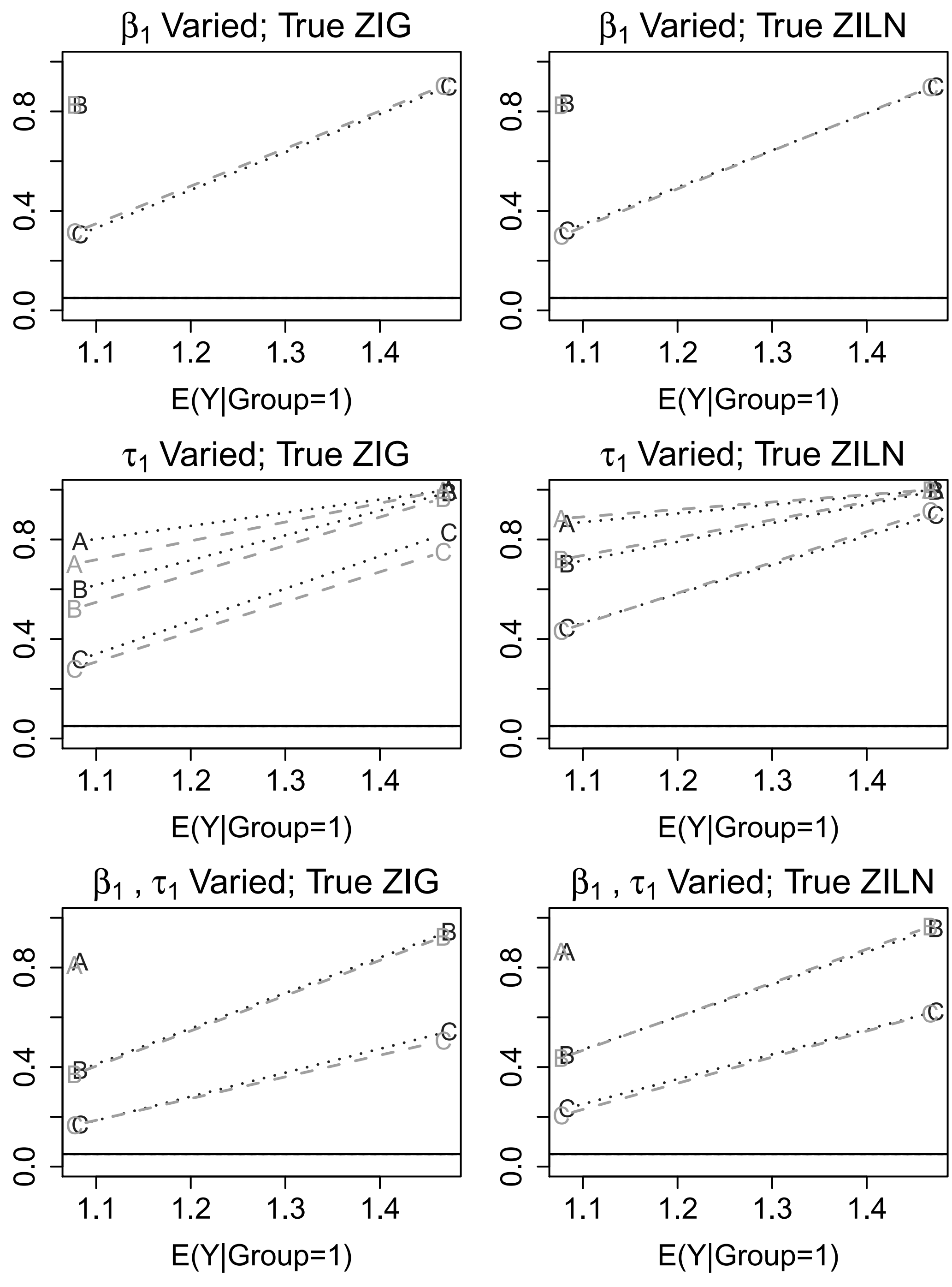

Figure 2.14: Power for consonant effects; two-part Wald tests where $\nu^{-1} = 0.5$. Symbols correspond to the group 0 settings with 'A' when $P(Y > 0|group = 0) = 0.75$, 'B' when $P(Y > 0|group = 0) = 0.5$, and 'C' when $P(Y > 0|group = 0) = 0.25$. The overall mean of the second group is on the x-axis. See Table 2.3 for more setting details. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

power is lower when the effects are split between the two parts of the model. This seems to be primarily due to the lower effect sizes required from each part when the overall mean difference is obtained by a combination of the two parts than when it was obtained from only one part alone. In Figure 2.14, the coefficient of variation is small, specifically $\nu^{-1} = 0.5$. At low levels of ν^{-1} the gamma and log-normal distributions only differ slightly. Similarly in this figure, it can be seen that differences in power between the two settings differ only slightly.

When the non-zero group effect is in the binomial part of the model, there is essentially no difference in power between ZIG and ZILN models under both true ZIG and true ZILN simulated data sets. Power differences between ZIG and ZILN analyses are also slight when non-zero group effects are in both parts of the model. When the effect is in the continuous part of the model (τ_1 varied) there is some distinction between correctly specified and misspecified models with correctly specified models having slightly higher power levels than the incorrectly specified models. The differences between correctly specified and misspecified analyses are the greatest when the data are truly ZIG. In that case, ZIG models have higher power than the misspecified ZILN models.

In Figure 2.15 the power results for 2 df Wald tests using the settings outlined in Table 2.3 and $\nu^{-1} = 2$ are shown. Similar to the results seen for $\nu^{-1} = 0.5$, when $\nu^{-1} = 2$ as the level of zero inflation decreases the power increases. When $\nu^{-1} = 2$ and the non-zero group effect was in the binomial part of the model the power levels remained similar to those seen when $\nu^{-1} = 0.5$. For the settings where there is a non-zero group effect in the continuous part of the model, the power levels for $\nu^{-1} = 2$ are lower than those seen when $\nu^{-1} = 0.5$. Figure 2.15 also shows that for $\nu^{-1} = 2$, when non-zero group effects are exclusively in the continuous part of the model power is lower when the true distribution is ZIG than when the true distribution is ZILN. Also, misspecification has a greater effect when $\nu^{-1} = 2$ and the data are ZIG than

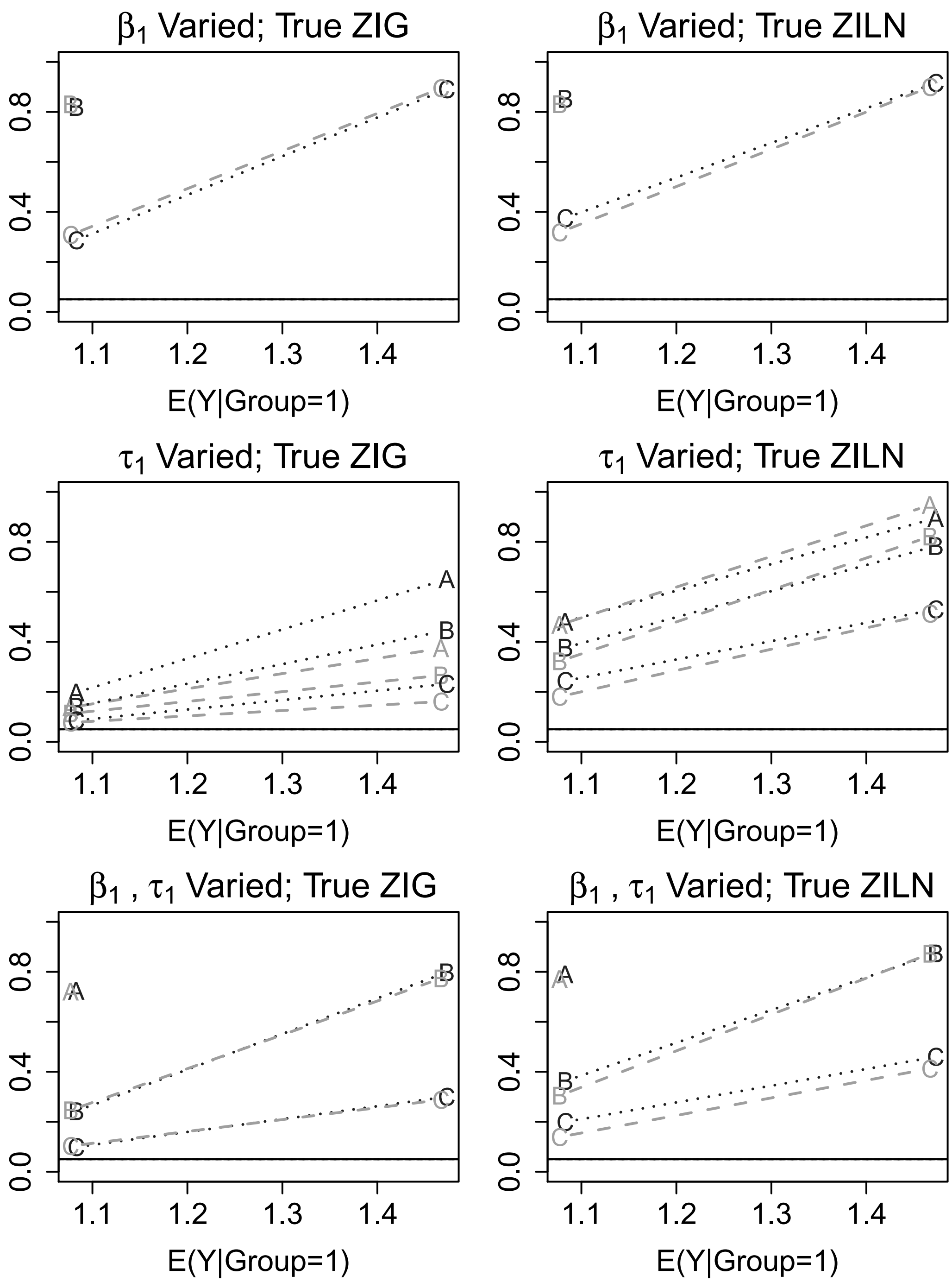

Figure 2.15: Power for consonant effects; two-part Wald tests where $\nu^{-1} = 2$. Symbols correspond to the group 0 settings with 'A' when $P(Y > 0|group = 0) = 0.75$, 'B' when $P(Y > 0|group = 0) = 0.5$, and 'C' when $P(Y > 0|group = 0) = 0.25$. The overall mean of the second group is on the x-axis see Table 2.3 for more setting details. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

it did when $\nu^{-1} = 0.5$. Specifically, ZIG analyses have substantially higher power than ZILN analyses when the data are truly ZIG. When that data are truly ZILN, on the other hand, ZIG analyses have only a slightly higher power than ZILN analyses in spite of the fact than with this high coefficient of variation Type 1 error had also been high.

The general power trends found in score and likelihood ratio tests are similar to those just described for Wald tests. There are, however, some differences in power across Wald, score, and likelihood ratio tests. These are summarized for group 0 setting 'B' in Figures 2.16 through 2.19. In these figures, the symbols used represent the various two-part tests; 'W' is used for the power results of the Wald tests, 'S' for those of the score tests, 'L' for those of the LRTs, and 'X' for the power results of the Wald-Wilcoxon tests. Figure 2.16 shows the power results for all of the settings in Table 2.3 where group 0 is from setting 'B' and $\nu^{-1} = 0.5$. Figure 2.17 is a zoomed in version of Figure 2.16 so that smaller differences in power can be seen. Figure 2.18 shows the power results when group 0 is from setting 'B' and $\nu^{-1} = 2$. Figure 2.19 is a zoomed in version of Figure 2.18.

Figures 2.16 and 2.17 show that when $\nu^{-1} = 0.5$, there are only slight differences in power when comparing various two-part tests. The largest differences between the power levels observed for the various tests are seen when the data are ZIG and τ_1 is varied to create an overall group 1 mean of 1.08. In that case, the ZIG LRT had the highest power, followed by the ZIG Wald, and the ZIG Score tests. The Wald-Wilcoxon test had power that was between the ZIG and ZILN tests, and the ZILN power results had a smaller range but with Wald tests having the highest power, followed by LRTs, and finally score tests. When the data are ZILN and τ_1 is varied to create an overall group 1 mean of 1.08, the differences among tests are much smaller with ZILN tests having higher powers that ZIG tests and Wald tests having slightly higher powers than score tests. At all other settings, the differences between

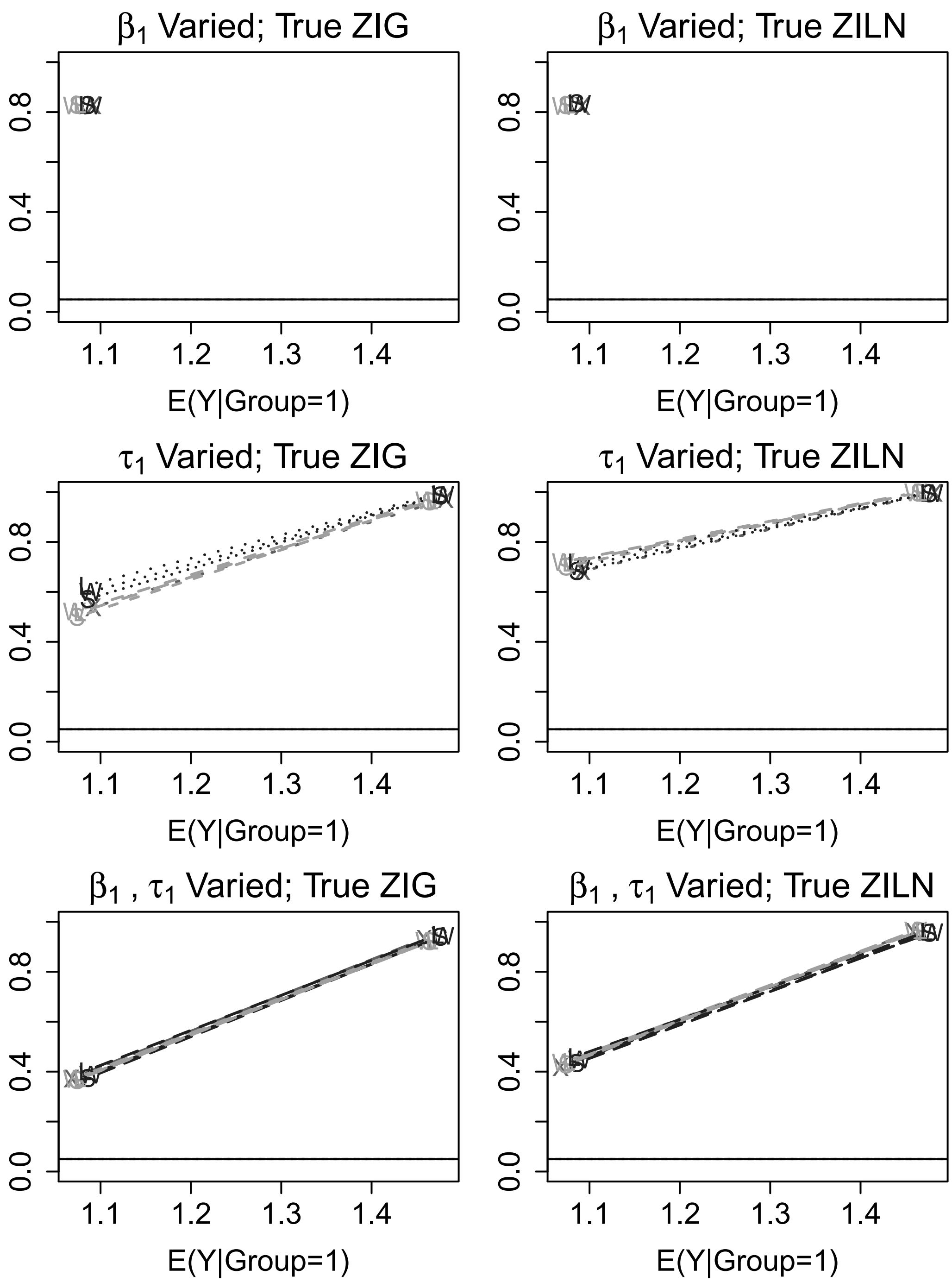

Figure 2.16: Comparing power across all 2 df tests. Setting 'B' with $\nu^{-1} = 0.5$ and no covariate adjusting. Labels represent the 2 df tests used; 'W' is for the Wald tests, 'S' refers to the score tests, 'L' refers to the LRTs, and 'X' refers to the Wald-Wilcoxon tests. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

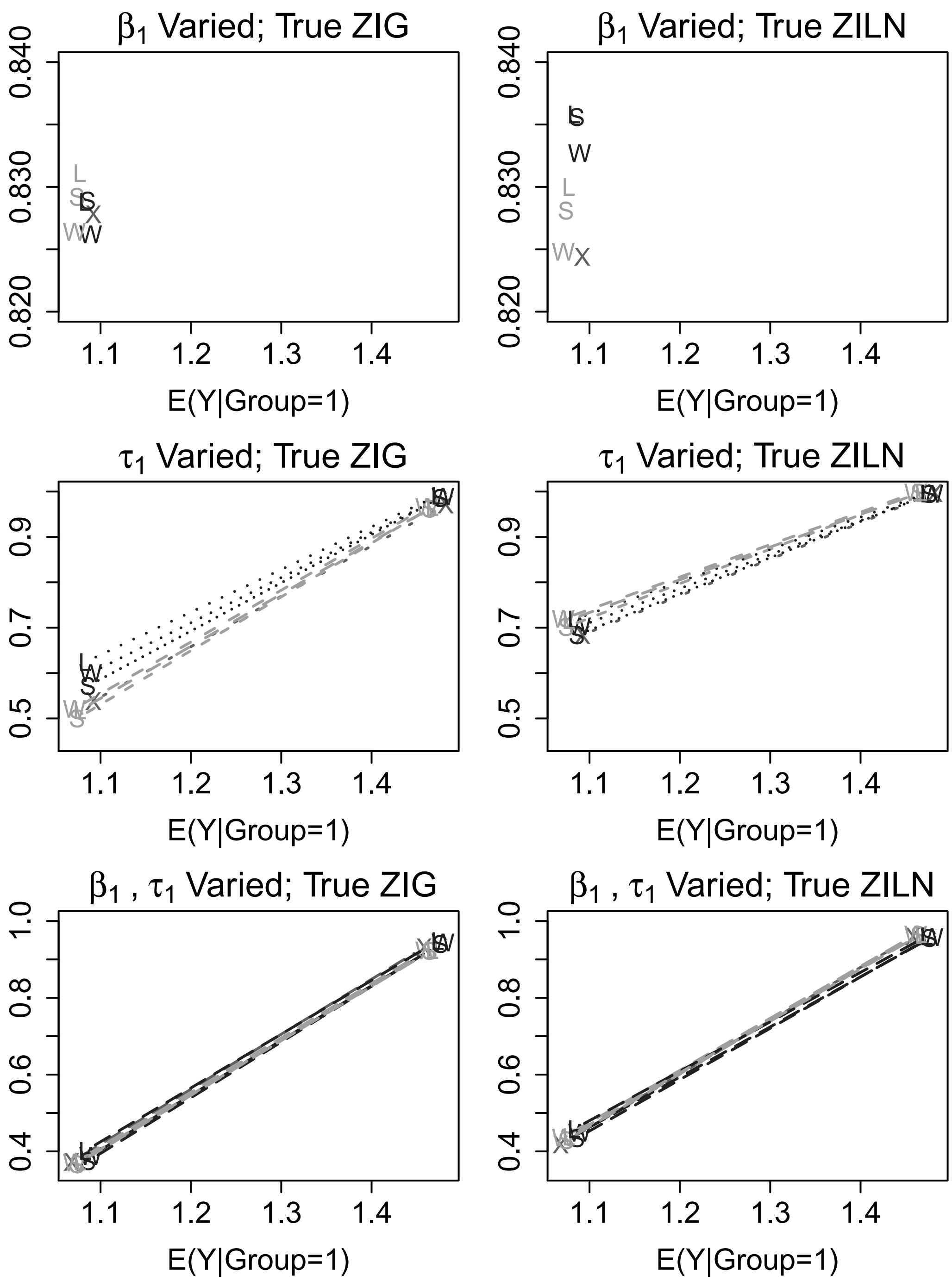

Figure 2.17: Zoomed in version of Figure 2.16. Comparing power across all 2 df tests. Setting 'B' with $\nu^{-1} = 0.5$ and no covariate adjusting. Labels represent the 2 df tests used; 'W' is for the Wald tests, 'S' refers to the score tests, 'L' refers to the LRTs, and 'X' refers to the Wald-Wilcoxon tests. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

groups are even slighter.

Figures 2.18 and 2.19 demonstrate that when $\nu^{-1} = 2$ there are greater differences between the power levels of the various two-part tests. However, these differences remain primarily when the non-zero group effect is entirely in the continuous part of the ZIG and ZILN models. When the non-zero group effects driving the power results are in the binomial part of the model, the distribution of the model both in terms of ZIG vs. ZILN and in terms of the values of ν^{-1} make little impact on the power of the two-part tests. This is because the first part of the two-part tests is identical across ZIG and ZILN models and is not impacted by ν^{-1}. On the other hand, when a strong non-zero group effect is in the continuous part of the model the distribution from which the data come, the model choice, and the value of ν^{-1} impact the power results. When the data are truly ZIG and the effect is in the continuous part of the model, power is highest for ZIG LRTs, followed by ZIG Wald tests, followed by ZIG score tests and the Wald-Wilcoxon tests (order of these two switch with M_1 value). Finally the ZILN tests have the lowest powers but the powers of the various ZILN two-part tests are approximately equivalent. When the data are truly ZILN, the powers for the ZILN analyses are also very similar to each other with the Wald-Wilcoxon having a slightly lower power that ZILN analyses.

2.8 Summary

In this chapter, two-part models (ZIG and ZILN) used for analysing semicontinuous data were introduced. Then the Newton-Raphson maximization techniques used to find the parameter estimates for the ZIG and ZILN models were outlined. Two-part tests developed from these models were proposed, including Wald, score, LRT, and Wald-Wilcoxon tests. Finally, results of a simulation study were presented, including both correctly specified and misspecified ZIG and ZILN analyses. The simulation study examined Type 1 error rates, and power levels for both dissonant and

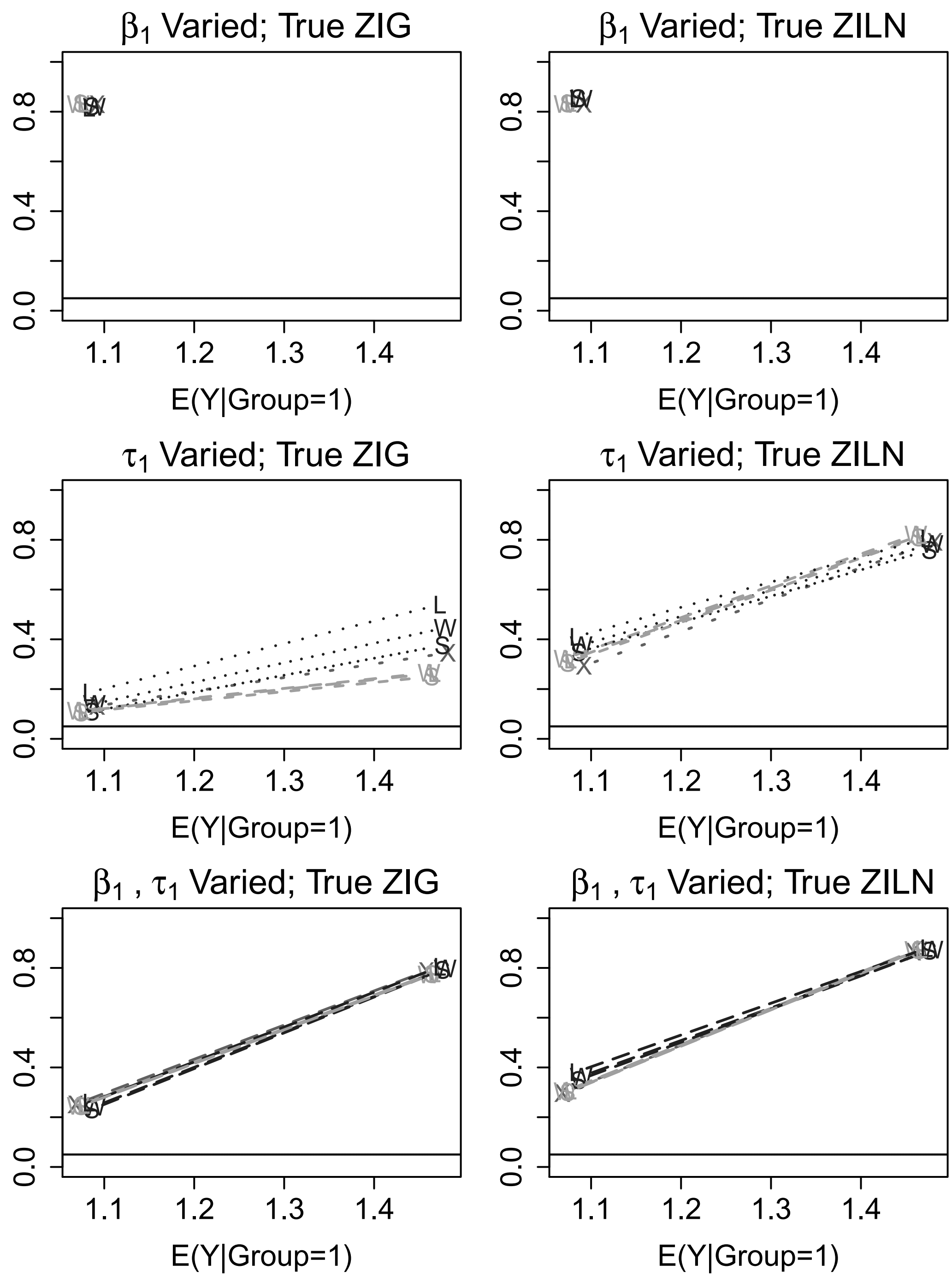

Figure 2.18: Comparing power across all 2 df tests. Setting 'B' with $\nu^{-1} = 2$ and no covariate adjusting. Labels represent the 2 df tests used; 'W' is for the Wald tests, 'S' refers to the score tests, 'L' refers to the LRTs, and 'X' refers to the Wald-Wilcoxon tests. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

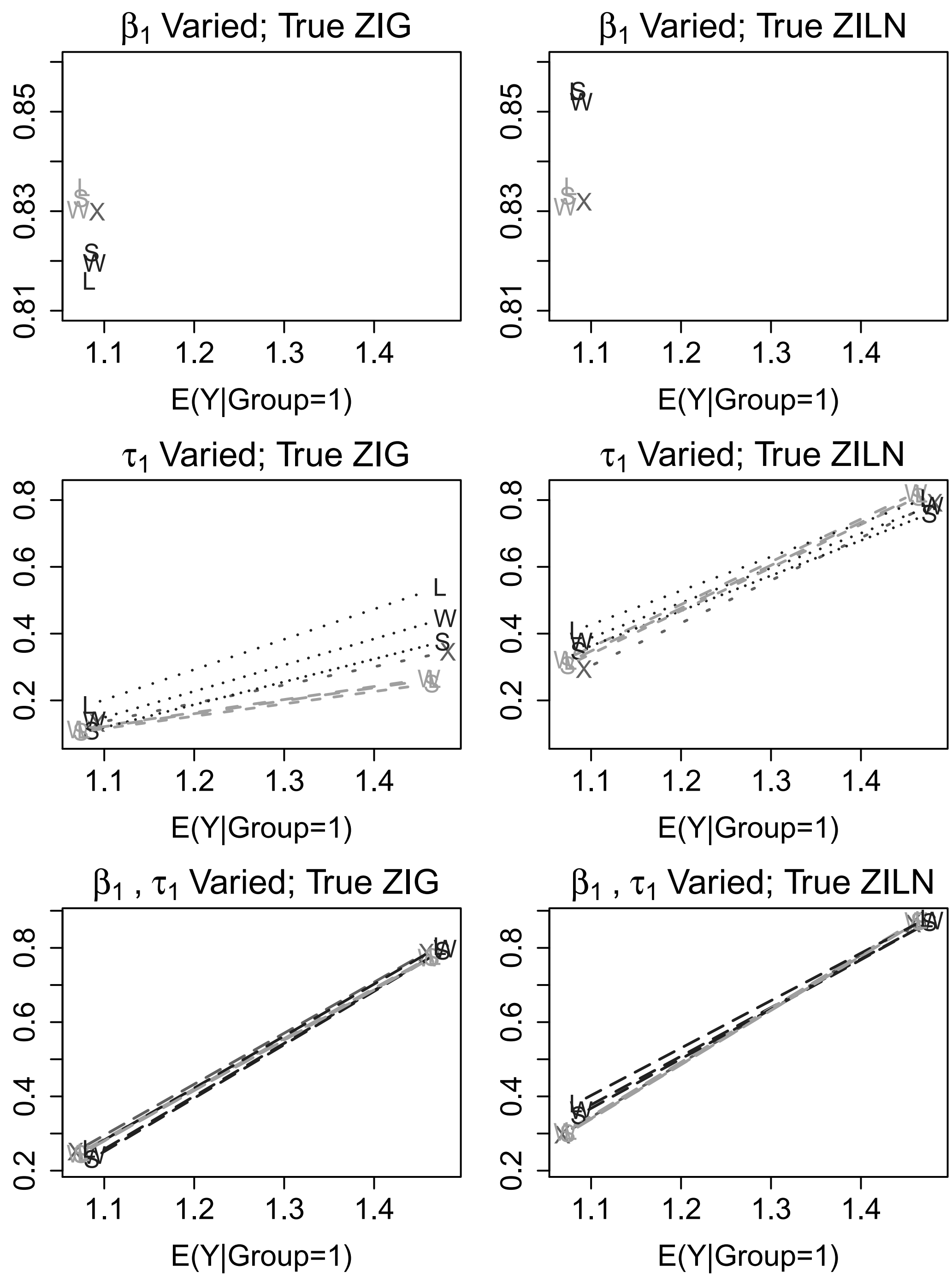

Figure 2.19: Zoomed in version of Figure 2.18. Comparing power across all 2 df tests. Setting 'B' with $\nu^{-1} = 2$ and no covariate adjusting. Labels represent the 2 df tests used; 'W' is for the Wald tests, 'S' refers to the score tests, 'L' refers to the LRTs, and 'X' refers to the Wald-Wilcoxon tests. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

consonant effects. The settings used in the simulation study also included various sample sizes, levels of zero-inflation, and values of ν^{-1}.

For sample sizes of n=100 and n=200, ZILN data correctly analysed assuming ZILN had appropriate Type 1 error rates for Wald tests, score tests and LRTs. When ZIG data was correctly analysed assuming ZIG, Type 1 error rates for n=100 and n=200 were slightly conservative for Wald and score tests, becoming more conservative as ν^{-1} increased. For ZIG data assuming ZIG, Type 1 error rates were within the appropriate range for LRTs. For data simulated as ZIG and misspecified as ZILN, Wald and score tests had appropriate Type 1 error rates and LRTs had slightly elevated Type 1 error rates which came closer to nominal as sample size increased. Conversely, prohibitively high Type 1 error rates were observed when data simulated as ZILN were misspecified as ZIG. Such misspecified LRTs had the highest Type 1 error rates, followed by Wald tests, and then score tests. Also, the Type 1 error rates became more elevated as ν^{-1} increased. The Wald-Wilcoxon test had conservative to appropriate Type 1 error rates.

Among all tests, the power to find significant group differences in the presence of dissonant effects increased with the strength of the effect of group (β_1 and τ_1). Among the settings utilized, this was true even when the the simulation with the larger τ_1 also had a greater level of zero-inflation. When the data were simulated as ZIG, analyses assuming ZIG had higher power levels than analyses assuming ZILN. For ZILN data, the power levels for analyses assuming ZIG and those assuming ZILN were similar. For all tests and parameters, power decreased as ν^{-1} increased. Wald-Wilcoxon tests had power levels that were among the lowest when the sample size was small, and at levels between the correctly specified and the misspecified analyses when the sample size was larger.

Power to find consonant effects using the two-part tests outlined in this chapter increases as the level of zero-inflation increases. Power also increases with the strength of the group effect, and decreases as ν^{-1} increases. The largest differences between analyses assuming ZIG and analyses assuming ZILN occur when the data are simulated from ZIG distributions and the non-zero group effect is in the continuous part of the model ($\tau_1 > 0$). When this is the case, analyses assuming ZIG distributions have higher power than those assuming ZILN distributions. This difference can be seen when $\nu^{-1} = 0.5$, but is greatest when $\nu^{-1} = 2$. For these settings, Wald-Wilcoxon tests have lower power than analyses assuming ZIG, but higher than analyses assuming ZILN. For the other settings, differences in power among analyses assuming ZIG, analyses assuming ZILN, and analyses using the Wald-Wilcoxon test are small, and the relative value of the power levels differs by setting.

CHAPTER 3
MEAN-BASED TWO GROUP COMPARISON TESTS WITHOUT COVARIATE ADJUSTMENT UNDER ZERO INFLATED GAMMA AND ZERO INFLATED LOG NORMAL DISTRIBUTIONS

3.1 Mean-Based Hypothesis

In Chapter 2, tests which examined differences between two groups due to either or both parts of the two-part model were the focus; that is both probability differences and conditional mean differences were of interest regardless of the overall difference in means. A researcher may instead be interested in making a qualitative judgment as to which group has a higher mean. Policy decisions may need to be made in applications of semicontinuous data which require a mean-based hypothesis. For example, when comparing health expenditures under two health care plans the overall mean spending under each plan is the important feature as it determines the overall costs associated with each plan. When interested in the effect of group on the overall mean, the hypothesis defined in Equation 1.3 would be the starting point for tests of this nature. This null hypothesis posits that the overall means of the two groups are equal, compared to an alternative hypothesis that the overall group means are not equal. The tests developed in this chapter are for such a hypothesis set. In the lane departure example this would equate to testing that the Parkinson's disease subjects and the elderly control subjects have the same average lane departure severity score.

3.2 Mean of ZIG and ZILN Distributions

This chapter focuses on comparing two groups via mean-based tests with no additional covariates. We continue the use of the two-part models defined in Section 2.2 where $P(Y > 0|X_1)$ is modeled via logistic regression and the conditional mean $E(Y|Y > 0, X_1)$ is modeled by gamma regression using a log-link (for ZIG models) or log-normal regression (for ZILN models). We define X_j to be an indicator of group membership with $X_1 = j$ for group j where $j = 0, 1$. The resultant overall mean is of the form:

$$M_j = E(Y|X_1 = j) = P(Y > 0|X_1 = j)E(Y|Y > 0, X_1 = j) = p_j\mu_j. \qquad (3.1)$$

For the ZIG regression as defined in Section 2.4.1, we define τ_0 as an intercept term and τ_1 as a group effect for the continuous part of the model term. We define β_0 as an intercept term and β_1 as a group effect term for the logistic regression portion of the model. Then, the conditional mean for group j is:

$$\mu_j = E(Y|Y > 0, X_1 = x_1) = e^{\tau_0+\tau_1 x_1}, \qquad (3.2)$$

and the overall mean for group j is:

$$M_j = E(Y|X_1 = x_1) = e^{\tau_0+\tau_1 x_1}\frac{e^{\beta_0+\beta_1 x_1}}{1 + e^{\beta_0+\beta_1 x_1}} \qquad (3.3)$$

where $x_1 = j$ for $j = 0, 1$.

ZILN regression is defined in Section 2.4.2 and defines τ_0, τ_1, β_0, and β_1 similarly to their definition described above for ZIG. The conditional part of the model utilizes log-normal regression. The mean of a log-normal distribution, where $\log(Z) \sim N(\mu, \sigma^2)$, is [19]

$$E(Z) = e^{\mu+\frac{\sigma^2}{2}}. \qquad (3.4)$$

Using two-part model notation, the conditional mean given group membership and $Y > 0$ is

$$\mu_j = E(Y|Y > 0, X_1 = x_1) = e^{\tau_0+\tau_1 x_1+\sigma^2/2} \qquad (3.5)$$

and the overall mean given group membership is:

$$M_1 = E(Y|X_1 = x_1) = e^{\tau_0+\tau_1 x_1+\sigma^2/2}\frac{e^{\beta_0+\beta_1 x_1}}{1 + e^{\beta_0+\beta_1 x_1}}. \qquad (3.6)$$

3.3 Mean Difference

One way to test a mean-based hypothesis is through a difference between the two means. The difference of the means (DM) is defined as:

$$DM = M_1 - M_0 = E(Y|X_1 = 1) - E(Y|X_1 = 0) \tag{3.7}$$

McLerran proposed a contrast statement in SAS **Proc Nlmixed** to test the 'difference across treatments' using a DM based test and ZIG distributional assumptions.[12]

3.3.1 Mean Difference Assuming a Zero Inflated Gamma Distribution

The difference of means (DM) follows easily from the definition of the mean of the ZIG distribution (see Equation 3.3). Under the ZIG framework, the estimator for DM given the parameters β_0, β_1, τ_0, and τ_1 is as follows:

$$\begin{aligned} DM(\beta, \tau) &= E(Y|X_1 = 1) - E(Y|X_1 = 0) \\ &= \frac{e^{\beta_0+\beta_1}}{1 + e^{\beta_0+\beta_1}} e^{\tau_0+\tau_1} - \frac{e^{\beta_0}}{1 + e^{\beta_0}} e^{\tau_0}. \end{aligned} \tag{3.8}$$

It can be shown that

$$\sqrt{n}\left(DM(\hat{\beta}, \hat{\tau}) - DM(\beta, \tau)\right) \xrightarrow{d} N(0, \Psi_{DM}) \tag{3.9}$$

where

$$\Psi_{DM} = \frac{\mu_0^2 p_0(1 - p_0)}{w_0} + \frac{\mu_1^2 p_1(1 - p_1)}{w_1} + \nu^{-1}\frac{\mu_0^2 p_0}{w_0} + \nu^{-1}\frac{\mu_1^2 p_1}{w_1}, \tag{3.10}$$

and w_j is the weight for group j. From this result, a variance of $DM(\hat{\beta}, \hat{\tau})$ can be defined for finite sample sizes. Specifically, since $w_0 = \frac{n_0}{n}$ and $w_1 = \frac{n_1}{n}$,

$$\begin{aligned} \hat{Var}(DM(\hat{\beta}, \hat{\tau})) =& \hat{\mu}_0^2 \frac{\hat{p}_0(1 - \hat{p}_0)}{n_0} + \hat{\mu}_1^2 \frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} \\ &+ \hat{\mu}_0^2 \frac{\hat{p}_0 \hat{\nu}^{-1}}{n_0} + \hat{\mu}_1^2 \frac{\hat{p}_1 \hat{\nu}^{-1}}{n_1} \end{aligned} \tag{3.11}$$

and $\hat{p}_0 = P(Y > 0|X_1 = 0) = \frac{e^{\hat{\beta}_0}}{1+e^{\hat{\beta}_0}}$, $\hat{p}_1 = P(Y > 0|X_1 = 1) = \frac{e^{\hat{\beta}_0+\hat{\beta}_1}}{1+e^{\hat{\beta}_0+\hat{\beta}_1}}$, $\hat{\mu}_0 = e^{\hat{\tau}_0}$, $\hat{\mu}_1 = e^{\hat{\tau}_0+\hat{\tau}_1}$, and n_j is the number of subjects in group j. See Appendix Sections A.1.1 and A.1.2 for the proof of these results culminating in Equation A.38.

Given Equations 3.8, 3.9, and 3.11 Wald-type test statistics and confidence intervals can be created. For example, the DM test statistic (used for the Type 1 error and power results presented in this chapter) is

$$D = \frac{DM(\hat{\beta}, \hat{\tau})}{\sqrt{\hat{Var}(DM(\hat{\beta}, \hat{\tau}))}} \tag{3.12}$$

which is asymptotically distributed as a standard normal under the null hypothesis of $DM = 0$.

3.3.2 Mean Difference Assuming a Zero Inflated Log-Normal Distribution

Following from the definition of the mean of the ZILN distribution in Equation 3.3; the estimator for DM (assuming homogeneous σ^2) is as follows:

$$\begin{aligned} DM(\beta, \tau, \sigma^2) &= M_1 - M_0 = E(Y|X_1 = 1) - E(Y|X_1 = 0) \\ &= \frac{e^{\beta_0+\beta_1}}{1 + e^{\beta_0+\beta_1}} e^{\tau_0+\tau_1+\sigma^2/2} - \frac{e^{\beta_0}}{1 + e^{\beta_0}} e^{\tau_0+\sigma^2/2}. \end{aligned} \tag{3.13}$$

It can be shown that

$$\sqrt{n}\left(DM(\hat{\beta}, \hat{\tau}, \hat{\sigma^2}) - DM(\beta, \tau, \sigma^2)\right) \xrightarrow{d} N(0, \Psi_{DM}) \tag{3.14}$$

where

$$\Psi_{DM} = \frac{\mu_0^2 p_0(1 - p_0)}{w_0} + \frac{\mu_1^2 p_1(1 - p_1)}{w_1} + \frac{\mu_0^2 p_0 \sigma^2}{w_0} + \frac{\mu_1^2 p_1 \sigma^2}{w_1} + \frac{\sigma^4}{2(w_0 p_0 + w_1 p_1)}(p_1\mu_1 - p_0\mu_0)^2, \tag{3.15}$$

From this result, a variance of $DM(\hat{\beta}, \hat{\tau}, \hat{\sigma}^2)$ can be defined for finite sample sizes.

Specifically, since $w_0 = \frac{n_0}{n}$ and $w_1 = \frac{n_1}{n}$,

$$\begin{aligned}\hat{Var}(DM(\hat{\beta}, \hat{\tau}, \hat{\sigma}^2)) =& \hat{\mu}_0^2 \frac{\hat{p}_0(1-\hat{p}_0)}{n_0} + \hat{\mu}_1^2 \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} \\ &+ \hat{\mu}_0^2 \frac{\hat{p}_0 \hat{\sigma}^2}{n_0} + \hat{\mu}_1^2 \frac{\hat{p}_1 \hat{\sigma}^2}{n_1} \\ &+ \frac{\hat{\sigma}^4}{2(n_0\hat{p}_0 + n_1\hat{p}_1)} (\hat{p}_1^2 \hat{\mu}_1 + \hat{p}_0^2 \hat{\mu}_0)\end{aligned} \tag{3.16}$$

and $\hat{p}_0 = P(Y > 0|X_1 = 0) = \frac{e^{\hat{\beta}_0}}{1+e^{\hat{\beta}_0}}$, $\hat{p}_1 = P(Y > 0|X_1 = 1) = \frac{e^{\hat{\beta}_0+\hat{\beta}_1}}{1+e^{\hat{\beta}_0+\hat{\beta}_1}}$, $\hat{\mu}_0 = e^{\hat{\tau}_0 + \frac{\hat{\sigma}^2}{2}}$, $\hat{\mu}_1 = e^{\hat{\tau}_0+\hat{\tau}_1+\frac{\hat{\sigma}^2}{2}}$, and n_j is the number of subjects in group j. See Appendix A.2.1 and A.2.2 for a further outline of these delta method results, with the proof culminating in Equation A.86.

The variance of DM for a ZILN distribution in Equation 3.16 is similar in form to the variance of DM for a ZIG distribution found Equation 3.8 except that $\hat{Var}(DM(\hat{\beta}, \hat{\tau}, \hat{\sigma}^2))$ includes an additional term because of the inclusion of the parameter $\hat{\sigma}^2$.

Finally, a DM test statistic can be calculated as:

$$D = \frac{DM(\hat{\beta}, \hat{\tau}, \hat{\sigma^2})}{\sqrt{\hat{Var}(DM(\hat{\beta}, \hat{\tau}, \hat{\sigma^2}))}} \tag{3.17}$$

which is asymptotically distributed as a standard normal under the null hypothesis of $DM = 0$.

3.4 Ratio of Means

Zhou and Tu [11] have proposed a confidence interval estimation technique for the ratio of two group means (RM). They allow the two groups to have different σ^2 values. For this dissertation, we have simplified this assumption and require a homogeneous σ^2 value. We create a hypothesis test using RM both for ZILN as they did, but also for ZIG which, to our knowledge, has not been considered before. A ratio of means between two groups is defined as:

$$RM = \frac{M_1}{M_0} = \frac{E(Y|X_1 = 1)}{E(Y|X_1 = 0)}. \tag{3.18}$$

3.4.1 Ratio of Means Assuming a Zero Inflated Gamma Distribution

RM for a ZIG distribution follows from the definition of RM in Equation 3.18 and the estimator for the mean of a ZIG distribution. Equation 3.19 shows RM in terms of the model parameters.

$$\begin{aligned} RM(\beta, \tau) &= \frac{M_1}{M_0} = \frac{E(Y|X_1 = 1)}{E(Y|X_1 = 0)} \\ &= \frac{\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} e^{\tau_0+\tau_1}}{\frac{e^{\beta_0}}{1+e^{\beta_0}} e^{\tau_0}} \\ &= e^{\beta_1+\tau_1} \frac{1+e^{\beta_0}}{1+e^{\beta_0+\beta_1}} \end{aligned} \tag{3.19}$$

For constructing hypothesis tests and confidence intervals, $\log(RM(\beta, \tau))$ will be used. The null hypothesis of RM=1 translates to a null hypothesis on the log scale of $\log(RM) = 0$, where $\log(RM)$ can be expressed as:

$$\begin{aligned} \log(RM(\beta, \tau)) &= \beta_1 + \tau_1 + \log(1 + e^{\beta_0}) - \log(1 + e^{\beta_0+\beta_1}) \\ &= \beta_1 + \tau_1 - \log(1 - p_0) + \log(1 - p_1). \end{aligned} \tag{3.20}$$

It can be shown that $\log(RM(\beta, \tau))$ has the following asymptotic distribution,

$$\sqrt{n}\left(\log(RM(\hat{\beta}, \hat{\tau})) - \log(RM(\beta, \tau))\right) \xrightarrow{d} N(0, \Psi_{log(RM)}) \tag{3.21}$$

where

$$\Psi_{\log(RM)} = \frac{p_0(1-p_0)}{w_0} + \frac{p_1(1-p_1)}{w_1} + \nu^{-1}\left(\frac{1}{w_0 p_0} + \frac{1}{w_1 p_1}\right) \tag{3.22}$$

From this result, a variance of $\log(RM(\hat{\beta}, \hat{\tau}))$ can be defined for finite sample sizes.

Specifically, since $w_0 = \frac{n_0}{n}$ and $w_1 = \frac{n_1}{n}$,

$$\hat{Var}(\log(RM(\hat{\beta}, \hat{\tau})))) = \nu^{-1}\left(\frac{n_0\hat{p}_0 + n_1\hat{p}_1}{n_0\hat{p}_0 n_1\hat{p}_1}\right) + \frac{1-\hat{p}_0}{n_0\hat{p}_0} + \frac{1-\hat{p}_1}{n_1\hat{p}_1}. \tag{3.23}$$

See Appendix A.1.3 for a further outline of these delta method results, with the proof culminating in Equation A.45.

Because the cancelation of parameters in Equation 3.19, this variance has fewer contributing terms than the variance for DM explored previously. Specifically, $\hat{Var}(\log(RM(\hat{\beta}, \hat{\tau}))))$ depends on the dispersion parameter and $\hat{p}_0$ and $\hat{p}_1$ but does not depend on the mean of the non-zero values M_0 or M_1.

From Equations 3.20, 3.21, and 3.23 Wald-type test statistics and confidence intervals based on $\hat{Var}(\log(RM(\hat{\beta}, \hat{\tau}))))$ can be created. The $\log(RM)$ based test statistic (used for the Type 1 error and power results presented in this chapter) is

$$R = \frac{\log(RM(\hat{\beta}, \hat{\tau}))}{\sqrt{\hat{Var}(\log(RM(\hat{\beta}, \hat{\tau})))}}, \tag{3.24}$$

which is asymptotically distributed as a standard normal. Using these same results, an $(1-\alpha)\%$ C.I. for RM can be calculated. First, a $(1-\alpha)\%$ C.I. for $\log(RM(\beta, \tau))$ can be expressed as L_{lRM}, U_{lRM} where $L_{lRM} = \log(RM(\hat{\beta}, \hat{\tau})) + z_{\alpha/2}Var(\log(RM(\hat{\beta}, \hat{\tau})))$ and $U_{lRM} = \log(RM(\hat{\beta}, \hat{\tau})) + z_{1-\alpha/2}Var(\log(RM(\hat{\beta}, \hat{\tau})))$. Then exponentiating the upper and lower limits of the $(1-\alpha)\%$ C.I. for $\log(RM(\beta, \tau))$ gives a $(1-\alpha)\%$ C.I. for RM, which is $(U_{RM}, L_{RM}) = (e^{U_{lRM}}, e^{L_{lRM}})$.

3.4.2 Ratio of Means Assuming a Zero Inflated Log Normal Distribution

A benefit of using RM for a ZILN mean comparison, is that the estimator of RM does not depend on σ^2 (if a constant σ^2 between the two groups is assumed). The following equation set defines RM under the ZILN framework and demonstrates the canceling out of σ^2:

$$\begin{aligned} RM(\beta,\tau) &= \frac{E(Y|X_1=1)}{E(Y|X_1=0)} \\ &= \frac{\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} e^{\tau_0+\tau_1+\sigma^2/2}}{\frac{e^{\beta_0}}{1+e^{\beta_0}} e^{\tau_0+\sigma^2/2}} \\ &= e^{\beta_1+\tau_1} \frac{1+e^{\beta_0}}{1+e^{\beta_0+\beta_1}}. \end{aligned} \tag{3.25}$$

Because of the canceling out of σ^2, the estimator of RM for ZIG and ZILN distributions are identical. The estimator for the variance of $\log(RM)$ under ZILN will be similar to that found under the ZIG regression scenario outlined above. It can be shown that,

$$\sqrt{n}\left(\log(RM(\hat{\beta},\hat{\tau})) - \log(RM(\beta,\tau))\right) \xrightarrow{d} N(0, \Psi_{log(RM)}) \tag{3.26}$$

where

$$\Psi_{log(RM)} = \sigma^2\left(\frac{w_0p_0 + w_1p_1}{n_0p_0w_1p_1}\right) + \frac{1-p_0}{w_0p_0} + \frac{1-p_1}{w_1p_1} \tag{3.27}$$

Then for finite sample sizes, the variance of $\log(RM)$ can be estimated as

$$\hat{Var}(\log(RM)) = \sigma^2\left(\frac{n_0\hat{p}_0 + n_1p_1}{n_0\hat{p}_0n_1\hat{p}_1}\right) + \frac{1-\hat{p}_0}{n_0\hat{p}_0} + \frac{1-\hat{p}_1}{n_1\hat{p}_1}. \tag{3.28}$$

This can be proven in using the delta method in a similar fashion to those proofs referenced above.

Using the results found in Equations 3.25 - 3.28 hypothesis test and confidence interval frameworks can be created similar to those found for ZIG regression in the previous section.

3.5 Simulation Methods for Comparing Two Groups Without Covariate Adjustment

The simulations settings presented in this chapter are identical to those in Chapter 2. However, here we are interested in the Type 1 error rates and power for mean-based tests, specifically for the DM and RM tests defined in the previous sections. We compare these tests for ZIG and ZILN distributions under both correct specification and misspecification scenarios. They are also compared to the 'naive' methods of a Student's t-test and a Wilcoxon rank sum test.

These mean-based tests were performed on the data simulated in Section 2.6. The ZIG and ZILN data were simulated via a two-step process first simulating 0's and 1's from a binomial distribution and then simulating from a gamma or a log-normal to replace the 1's with a continuous distribution. Total sample sizes of n=50, 100, and 200 were used with equal sample sizes for each group, with 10,000 data sets created for each setting. The parameters for the two-part model were estimated using the Newton-Raphson technique outlined in Section 2.4. When parameters of the two-part model were not estimable, the data sets were removed, and any settings where more than 1% of the data sets had inestimable parameters were excluded from the simulation study.

3.5.1 Simulation Settings

Tables 3.1, 3.2, and 3.3 show the same settings as were seen in Tables 2.1, 2.2, and 2.3, augmented with the values of DM and RM attained at those settings. Tables 3.1 and 3.2 include settings with equal group means where DM=0 and RM=1. These will be used to explore Type 1 errors. Table 3.1 includes settings where group 0 and group 1 are drawn from identical distributions with $\beta_1 = 0$ and $\tau_1 = 0$. In Table 3.2, the overall group means are equal to each other but are obtained through dissonant effects with β_1 and τ_1 both non-zero. Table 3.3 shows the settings used for the power analyses with two different mean comparison levels. The lower setting has DM=0.33 and RM=1.44 and the higher setting has DM=0.72 and RM=1.96.

Label	μ_0	p_0	M_0	μ_1	p_1	M_1	e^{β_1}	e^{τ_1}	RM	DM
A	1	0.75	0.75	1	0.75	0.75	1	1	1	0
B	1.5	0.5	0.75	1.5	0.5	0.75	1	1	1	0
C	3	0.25	0.75	3	0.25	0.75	1	1	1	0

Table 3.1: Null settings for two group comparisons 'without covariate adjustment', with DM and RM values.

Comparison[1]	Label	μ_0	p_0	M_0	μ_1	p_1	M_1	e^{β_1}	e^{τ_1}	RM	DM
A vs B	b	1	0.75	0.75	1.5	0.5	0.75	1/3	1.5	1	0
A vs C	c	1	0.75	0.75	3	0.25	0.75	1/9	3	1	0
A vs B	d	1.5	0.5	0.75	3	0.25	0.75	1/3	2	1	0

Table 3.2: Null settings with dissonant effects and equal group means; two group comparisons 'without covariate adjustment', with DM and RM.

[1] Comparison settings come from Table 2.1; e.g., for setting 'b' group 0 has the probabilities and conditional means from setting 'A' in Table 2.1 while group 1 has the probability and conditional means from setting 'B'.

3.6 Simulation Results Comparing Two Groups via Mean-Based Tests Without Covariate Adjustment

3.6.1 Type 1 Error Under General Null Setting

Figures 3.1 to 3.3 show the Type 1 error rates coming from the settings explained in Table 3.1. The labels 'A', 'B', and 'C' represent the settings so labeled in Table 3.1 with 'A' having the highest probability of a non-zero outcome and the lowest conditional mean and 'C' having the lowest probability of a non-zero outcome and the highest conditional mean. For all figures, dark gray labels and dotted lines represent analyses assuming ZIG and light gray labels and dashed lines represent

	Group 0 Setting[1]	M_1	μ_1	p_1	e^{β_1}	e^{τ_1}	RM	DM
β_1 Varied	B	1.08	1.5	0.72	2 4/7	1	1.44	0.33
	B[2]	1.47	1.5	0.98	49	1	1.96	0.72
	C	1.08	3	0.36	1 11/16	1	1.44	0.33
	C	1.47	3	0.49	2 15/17	1	1.96	0.72
τ_1 Varied	A	1.08	1.44	0.75	1	1.44	1.44	0.33
	A	1.47	1.96	0.75	1	1.96	1.96	0.72
	B	1.08	2.16	0.5	1	1.44	1.44	0.33
	B	1.47	2.94	0.5	1	1.96	1.96	0.72
	C	1.08	4.32	0.25	1	1.44	1.44	0.33
	C	1.47	5.88	0.25	1	1.96	1.96	0.72
β_1 and τ_1 Varied	A	1.08	1.2	0.9	3	1.2	1.44	0.33
	B	1.47	1.8	0.6	1.5	1.2	1.44	0.33
	B	1.08	2.1	0.7	2 1/3	1.4	1.96	0.72
	C	1.08	3.6	0.3	1 2/7	1.2	1.44	0.33
	C	1.47	4.2	0.35	1 8/13	1.4	1.96	0.72

Table 3.3: Settings for power analyses; two group comparisons 'without covariate adjustment', with DM and RM values.

[1] For group 0 means and probabilities, see corresponding letters in Table 3.1.

[2] Settings for which $< 99\%$ of data sets could be solved.

analyses assuming ZILN. A solid black line shows the nominal level of 0.05 that was used for these analyses with dotted lines at 0.0457 and 0.0543 denote the bounds of acceptable Type 1 error. This result reflects comes from $\pm$ 1.96 standard errors of a binomial distributions when p=0.5 and 10,000 different data sets are run. Any observed Type 1 error rates that are outside these bounds are different than would be expected if the Type 1 error were truly equal to the nominal 0.5 level.

Figure 3.1 shows Type 1 error rates for the DM based test outlined in Section 3.3. The plots on the left represent the findings where data were simulated from ZIG distributions. With data coming from a ZIG distribution, Type 1 errors for the DM based test are excessively conservative. They are increasingly conservative as ν^{-1} increases and more conservative for ZILN analyses than for ZIG analyses. Type 1 errors decrease with the $P(Y > 0)$ leading setting 'C' to have the lowest Type 1 errors. When correctly analyzing data and ZIG, Type 1 error rates become closer to nominal as sample size increases for all settings. This over conservativeness seems to be a matter of slow convergence of the DM test to normality. When ZIG data are incorrectly analyzed as ZILN, for settings 'A' and 'B', Type 1 errors become closer to nominal as sample size increases but with setting 'C' (where $P(Y > 0) = .25$ and $E(Y|Y > 0) = 3$) Type 1 errors become more anti-conservative as sample size increases. The lowest observed Type 1 error occurs when setting is 'C', $\nu^{-1} = 2$, and ZIG data are analyzed as ZILN at a level of 0.0009.

The graphs on the right side of Figure 3.1 include the results for data simulated as ZILN. When the ZILN data are analyzed as ZILN, Type 1 errors are conservative especially as ν^{-1} increases. As sample size increases this tendency is counteracted and Type 1 error rates become closer to nominal. This same ZILN data when analyzed as ZIG typically has an increased Type 1 error. This Type 1 error increases with ν^{-1} and with sample size. There are large differences apparent in the Type 1 errors across settings with a higher $P(Y > 0)$ leading to a higher Type 1 error; for n=50

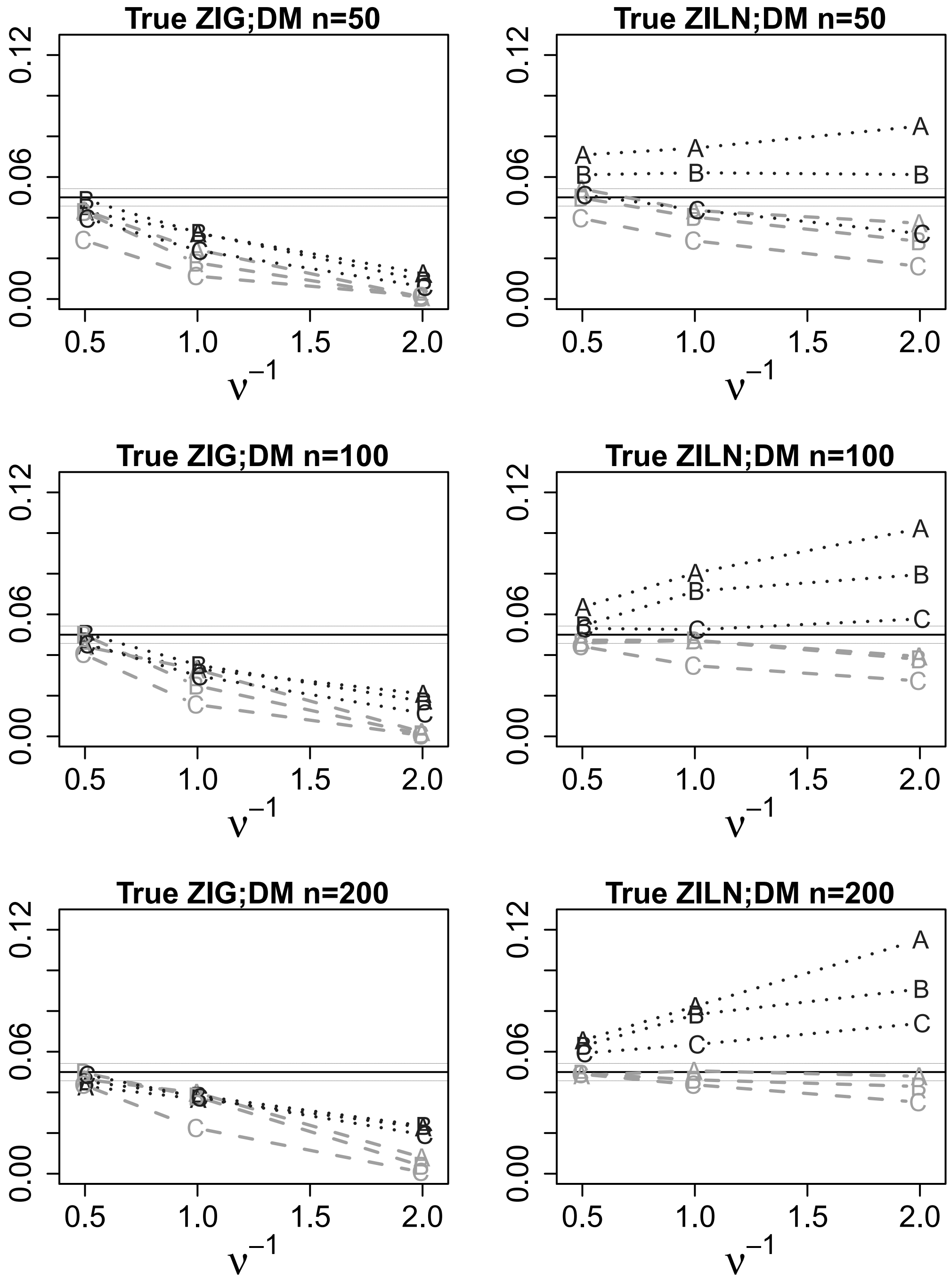

Figure 3.1: Type 1 error for tests based on DM; two group comparison 'without covariate adjustment's. Symbols correspond to the setting in Table 2.1 where 'A' is used when $P(Y > 0) = 0.75$, 'B' when $P(Y > 0) = 0.5$, and 'C' when $P(Y > 0) = 0.25$. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

and setting 'C', Type 1 errors are conservative at higher levels of ν^{-1} whereas for all other settings and sample sizes ZILN data analyzed assuming ZIG lead to increased Type 1 error rates.

Figure 3.2 shows the Type 1 error rates for the test based on $\log(RM)$ described in Section 3.4. The plots on the left contain the Type 1 error rates for the $\log(RM)$ based tests when the data are simulated from ZIG distributions. The Type 1 errors for the $\log(RM)$ based tests with ZIG data are closer to nominal than they were for DM based tests. Also, unlike the results seen for the DM tests, for $\log(RM)$ ZIG data analyzed as ZILN have higher Type 1 error rates than ZIG data analyzed as ZIG. These differences between DM tests and $\log(RM)$ tests are likely due to the lack of need to include σ^2 in the estimation of the $\log(RM)$ which leads to a smaller test variance. Type 1 errors for $\log(RM)$ for ZIG data are within or lower than acceptable Type 1 error limits except when sample size is small (setting 'C' when n=50, where the expected sample size for the continuous part is 6.25 per group, is the only setting with Type 1 error greater than 0.06). Also with ZIG data analyzed as ZIG, Type 1 error becomes more conservative as ν^{-1} increases. However, the conservativeness seen for the $\log(RM)$ isn't as extreme as in DM but is instead similar to the slightly conservative Type 1 error rates that were observed for the two-part Wald and score tests in Section 2.7.1.

The right hand plots in Figure 3.2 show the Type 1 error rates observed when the data were simulated as ZILN. With ZILN data, $\log(RM)$ tests assuming ZILN have appropriate Type 1 error, except where Type 1 error is only slightly elevated when n=50 and $\nu^{-1} = 2$ for settings 'A' and 'B'. When ZILN data are analyzed as ZIG, the Type 1 errors for the $\log(RM)$ tests are inflated. This worsens as sample size increases, both overall sample size and the expected sample size for the continuous part, and as ν^{-1} increases.

Figure 3.3 shows the Type 1 error rates for the mean-based or directional tests

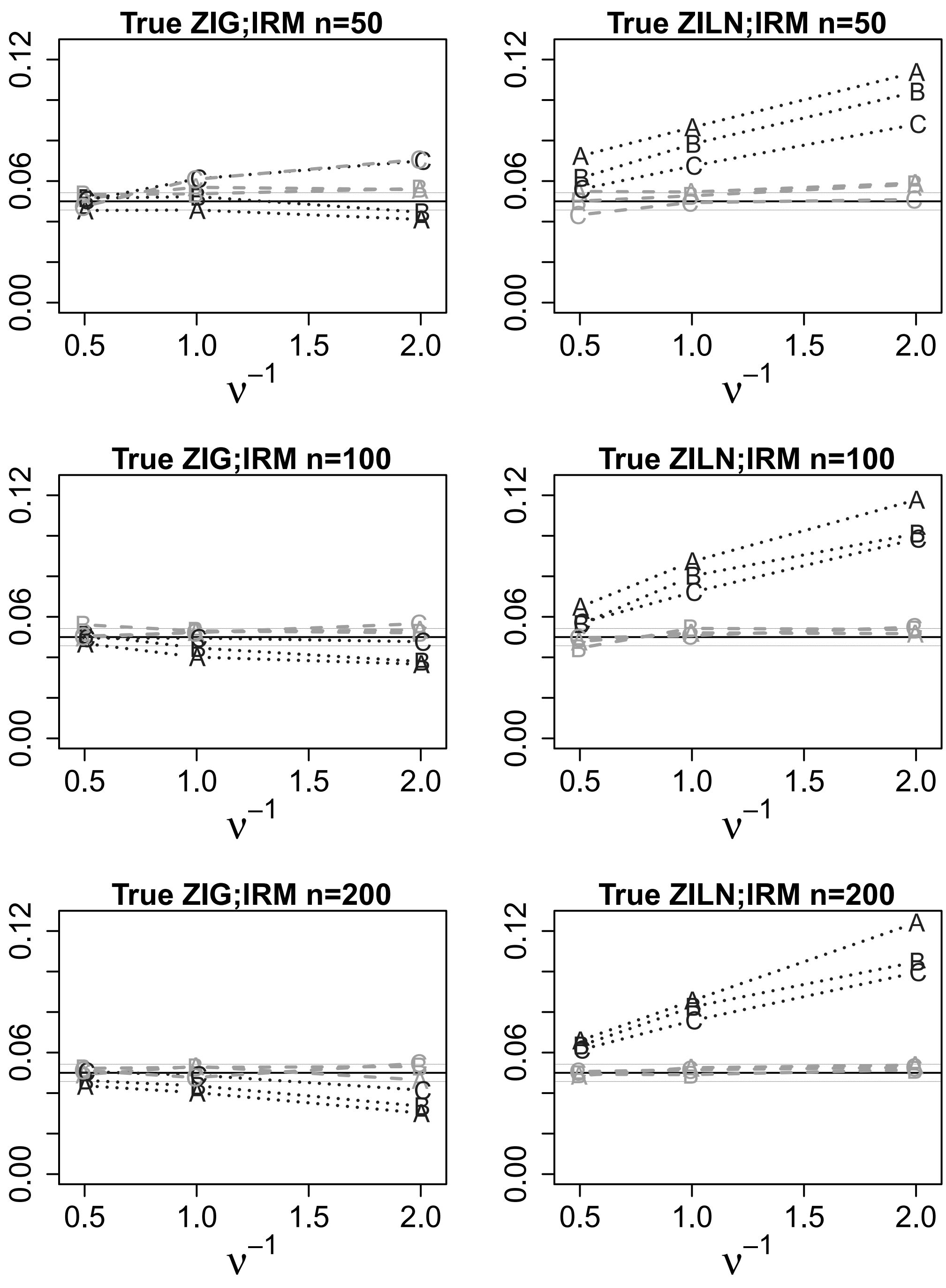

Figure 3.2: Type 1 error for tests based on $\log(RM)$; two group comparison 'without covariate adjustment's. Symbols correspond to the setting in Table 2.1 where 'A' is used when $P(Y > 0) = 0.75$, 'B' when $P(Y > 0) = 0.5$, and 'C' when $P(Y > 0) = 0.25$. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

used above at one example mean setting. The setting used is here is 'B' where $P(Y > 0) = 0.5$ and $E(Y|Y > 0) = 1.5$, but the trends are similar to those seen in the other settings. The labels in this plot identify the various tests used. Specifically, the label 'D' refers to the Type 1 error for the difference of means DM test, 'R' refers to the test based on the $\log(RM)$, 'T' refers to Type 1 error for the t-test, and 'X' refers to the Wilcoxon rank sum test. As in the previous plots, the darker grey with dotted lines refers to analyses assuming ZIG distributions and the lighter grey with dashed lines refers to analyses assuming a ZILN distribution. The Wald-Wilcoxon is shown in a grey that is in between the dark and light shades used for the distributional tests and uses a short dash, and the t-test is shown with a black line and an alternating dash dot pattern.

The Type 1 errors for t-tests are within the appropriate bounds to somewhat conservative. Type 1 error increases with ν^{-1}. As sample size increases the Type 1 errors become closer to nominal. The Type 1 errors for the Wilcoxon rank sum test are all within the appropriate Type 1 error bounds. In general, the Wilcoxon, ZILN $\log(RM)$, and t-tests stay closest to the nominal levels. The plots on the left side demonstrate that when the data are ZIG DM has the most excessively conservative Type 1 errors of all of the tests. The right hand plots show that the $\log(RM)$ tests have the most extremely high Type 1 errors when ZILN data are incorrectly analyzed assuming ZIG.

3.6.2 Type 1 Error Dissonant Effects with Equal Group Means

Table 3.2 shows settings where the groups have the same overall mean but the means were obtained through dissonant effects. That is, within each setting the two groups have the same overall mean but different probabilities of a non-zero outcome and different conditional means given a non-zero outcome. In Chapter 2 these were considered to be powers because they did not coincide with the null hypothesis used

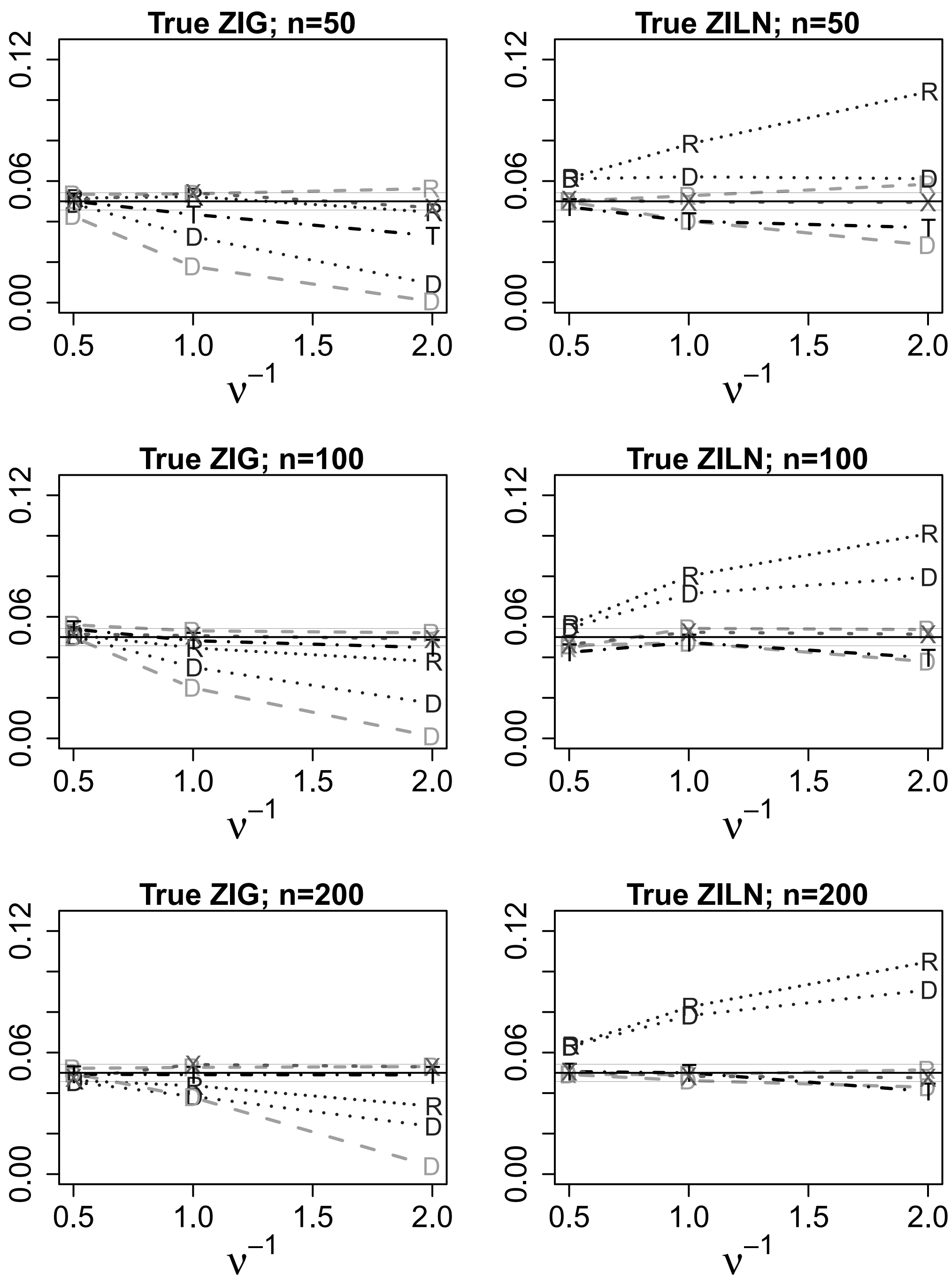

Figure 3.3: Type 1 error for mean-based tests using setting 'B'; two group comparison 'without covariate adjustment's. 'D' refers to DM test, 'R' refers to RM test, 'T' refers to the t-test, and 'X' refers to the Wilcoxon rank sum test. For DM and RM, darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

in that chapter, i.e. no difference between two groups in terms of the probability of a non-zero outcome or in terms of the conditional mean. However, in this chapter the focus is on tests based on the hypothesis in Equation 1.3, where the null hypothesis refers to no difference in the overall mean of the two groups. The settings laid out in Table 3.2 have equivalent means for the two groups and so will rightly be considered Type 1 errors, with DM=0 and RM=1 for these settings.

The Type 1 error results for DM under dissonant effects seen in Figure 3.4 are similar to the DM Type 1 error results previously seen. Namely, the Type 1 error rates decrease as ν^{-1} increases when the data are simulated as ZIG and increases with ν^{-1} when ZILN data are analyzed as ZIG. Also for DM, ZIG data analyzed as ZILN exhibit extremely low Type 1 error rates which are lower than the Type 1 error for rates ZIG data analyzed as ZIG. However, Type 1 error results for DM tests are higher in general among ZIG data for these mean-based null settings than they were for the settings where the two groups had the same probabilities and conditional means. This shift upwards in Type 1 error compared to the general null settings is particularly noticeable for setting 'c'. For this setting, when appropriately analyzed as ZIG, Type 1 error rates are above nominal for sample sizes of 50 or 100 but lessen as sample size increases. For setting 'c', although the overall mean is equal, $P(Y > 0)$ and $E(Y|Y > 0)$ differ to the largest extent between the two groups. The increased Type 1 error may be due to the fact that in some of the random samples the effects may not cancel out perfectly, leading to finding an overall effect when there shouldn't be one. When the sample size is larger this situation becomes less likely.

When the data are truly ZILN, Type 1 error rates for DM tests applied to dissonant effect simulations are higher than for the settings of the previous section. For setting 'c', when appropriately analyzed as ZILN, Type 1 errors are elevated but decrease with sample size. For settings 'b' and 'd', slightly conservative Type 1 errors get closer to nominal levels as sample size increases. When ZILN data are

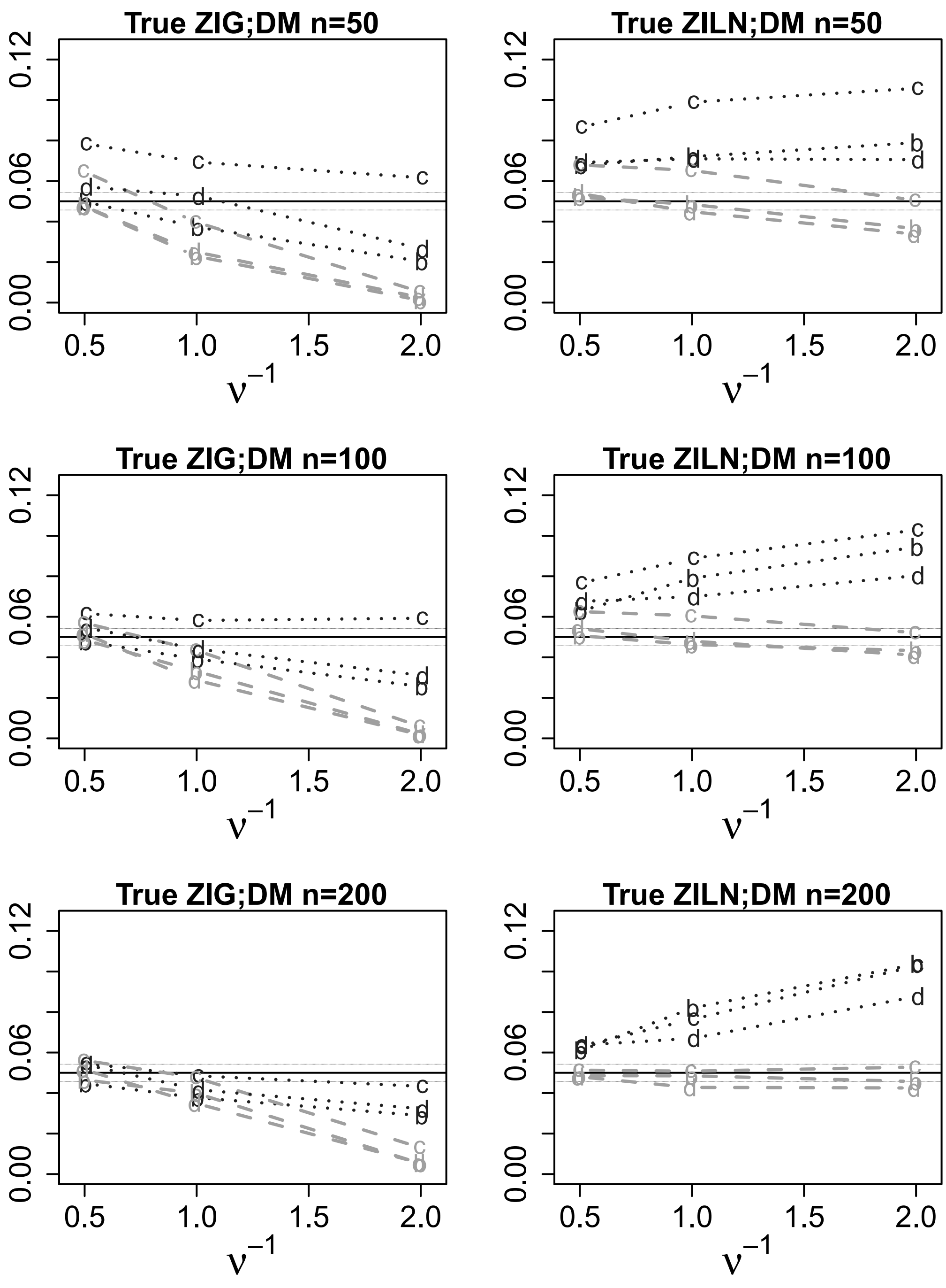

Figure 3.4: Type 1 error for tests based on DM; dissonant effects; two group comparison 'without covariate adjustment's. Symbols correspond to the setting in Table 2.2 where setting 'b' indicates $P(Y > 0|group = 0) = 0.75$ and $P(Y > 0|group = 1) = 0.5$, setting 'c' indicates $P(Y > 0|group = 0) = 0.75$ and $P(Y > 0|group = 1) = 0.25$, and setting 'd' indicates $P(Y > 0|group = 0) = 0.5$ and $P(Y > 0|group = 1) = 0.25$. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

analyzed as ZIG, the relationship between Type 1 error across settings and sample sizes gets more complicated than it was in the no effect cases of the previous section. Setting ‘c’ has a higher Type 1 error than ‘b’ and ‘d’ in the smaller sample sizes. However, as sample size increases, this effect disappears and ‘b’ has the largest Type 1 error rates, likely due to having the smallest level of zero inflation. There appears to be a combination of effects taking place. First, there is the effect of ZILN data misspecified as ZIG which was demonstrated in Section 3.6.1 to lead to increasing Type 1 errors when sample sizes increase. Then there is likely another effect that arises where the dissonant effects may not properly cancel out well enough in smaller sample sizes leading to a high Type 1 error that decreases with sample size. These two contradictory trends meet in the ZILN data and lead to the change in which setting has the highest Type 1 errors.

Figure 3.5 illustrates the Type 1 error results for the $\log(RM)$ based test statistic applied to the same dissonant data settings used for Figure 3.4. The left graphs show that when the data are truly ZIG, Type 1 errors are fairly close to nominal when ZILN is assumed; and when the data are analyzed as ZIG there is some decrease in Type 1 error as ν^{-1} increases. Type 1 error decreases as ν^{-1} increases, but to a much smaller extent than was the case for DM based tests. When the data are distributed as ZILN and a ZILN distribution is assumed, Type 1 errors are within the appropriate range except for slightly elevated Type 1 errors for setting ‘b’ when n=50. When ZILN data are analyzed assuming ZIG, elevated Type 1 errors are again observed along with an increase in Type 1 error as ν^{-1} increases. For $\log(RM)$ tests, settings with higher $P(Y > 0)$ tend to have more elevated Type 1 errors than those with lower $P(Y > 0)$ even though those settings with lower $P(Y > 0)$ have larger dissonant effects. Comparing settings ‘b’ and ‘d’ can be illustrative of the interplay between dissonant effect size and the expected sample size for the continuous part of the model. Setting ‘b’ has higher probabilities of non-zero outcomes than does

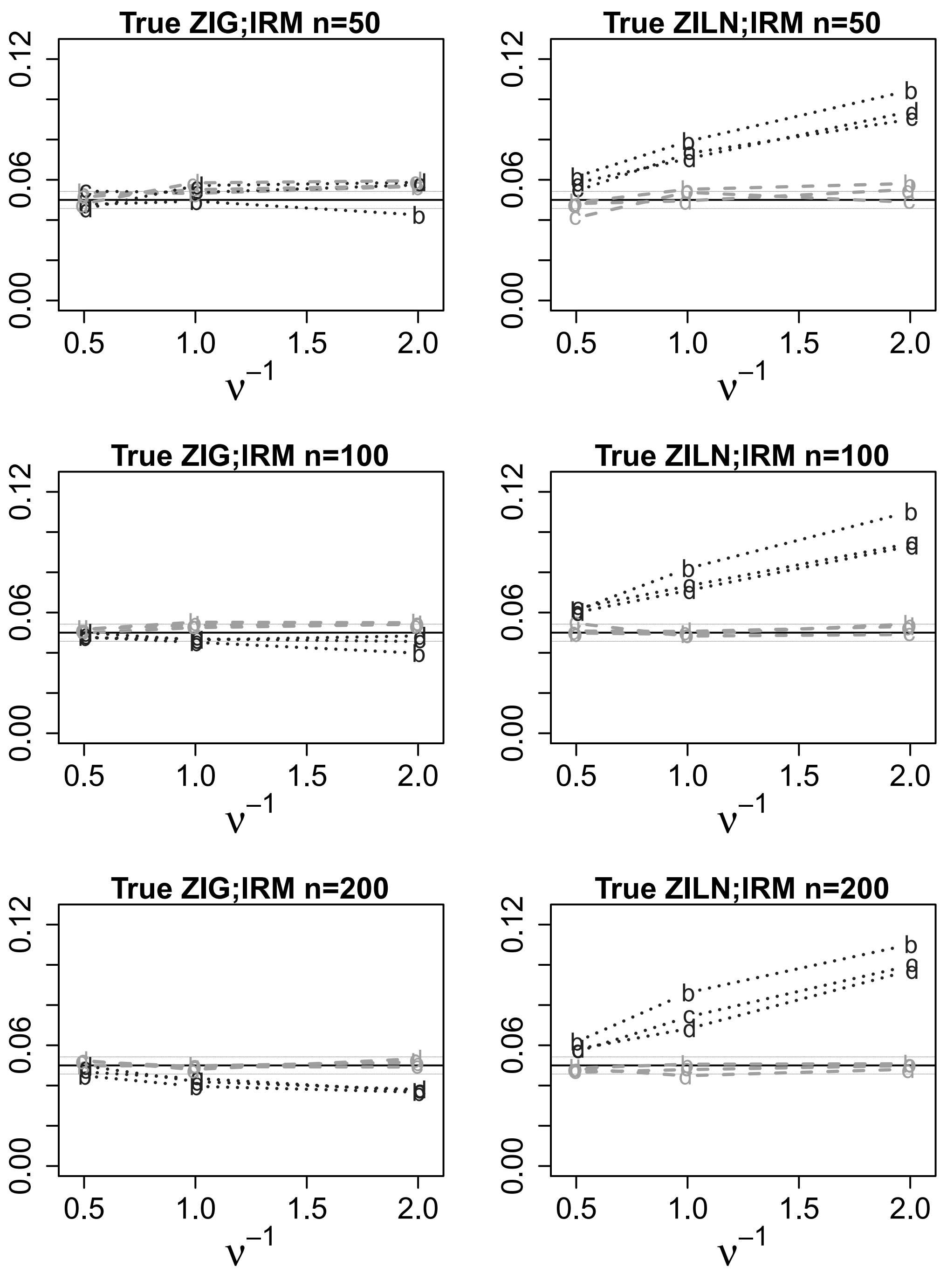

Figure 3.5: Type 1 error for tests based on $\log(RM)$; dissonant effects; two group comparison 'without covariate adjustment's. Symbols correspond to the setting in Table 2.2 where setting 'b' indicates $P(Y > 0|group = 0) = 0.75$ and $P(Y > 0|group = 1) = 0.5$, setting 'c' indicates $P(Y > 0|group = 0) = 0.75$ and $P(Y > 0|group = 1) = 0.25$, and setting 'd' indicates $P(Y > 0|group = 0) = 0.5$ and $P(Y > 0|group = 1) = 0.25$. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

setting ‘d’, but setting ‘d’ has slightly larger dissonant effect sizes.

Figure 3.6 shows the Type 1 error results for the t-test for these same simulations, where the dissonant effects cancel out in the mean. Just as with DM and $\log(RM)$ based tests, the t-test has a null hypothesis of equality for the overall means of the two groups. The t-test assumes a normal distribution; that assumption is violated severely in these settings. Type 1 errors for the t-test are slightly elevated for settings ‘c’ and ‘d’. Type 1 error is higher where effect sizes and the number of zeros expected are greater. On the other hand, for setting ‘b’ which has the fewest zero values and the weakest dissonant effects, the Type 1 errors are at appropriate levels.

The Wilcoxon rank sum test evaluates whether two distributions are equal. The group 0 and group 1 distributions simulated here have equal means, but differ in distribution. Specifically, one group has a higher probability of a non-zero outcome and the other group has a higher conditional mean. Because of this it must be noted that these results are power results in terms of the Wilcoxon test hypotheses. In Figure 3.7, the power to detect the difference between two groups when there are dissonant effects that cancel out in terms of the mean is shown. This test is not a good test when directionality and mean-based comparison are of interest. On the other hand, when the power results for the Wilcoxon test in Figure 3.7 are compared to the two-part tests for the same data shown in Chapter 2 Figures 2.8 - 2.11, it is clear that the power to find the difference between the two groups is lower for the Wilcoxon tests than for the 2 d.f. tests. Therefore, regardless of which scenario is of interest to a researcher, the Wilcoxon rank sum tests performs poorly compared to the other options we’ve observed.

Figures 3.8, 3.9, and 3.10 show a comparison of the mean-based tests with one plot per setting so that the Type 1 errors of the tests can be compared. The labels used for these Type 1 error plots are as follows: ‘D’ represents DM test, ‘R’ the test based on $\log(RM)$, and ‘T’ for the t-test. Dark grey with dotted lines refers

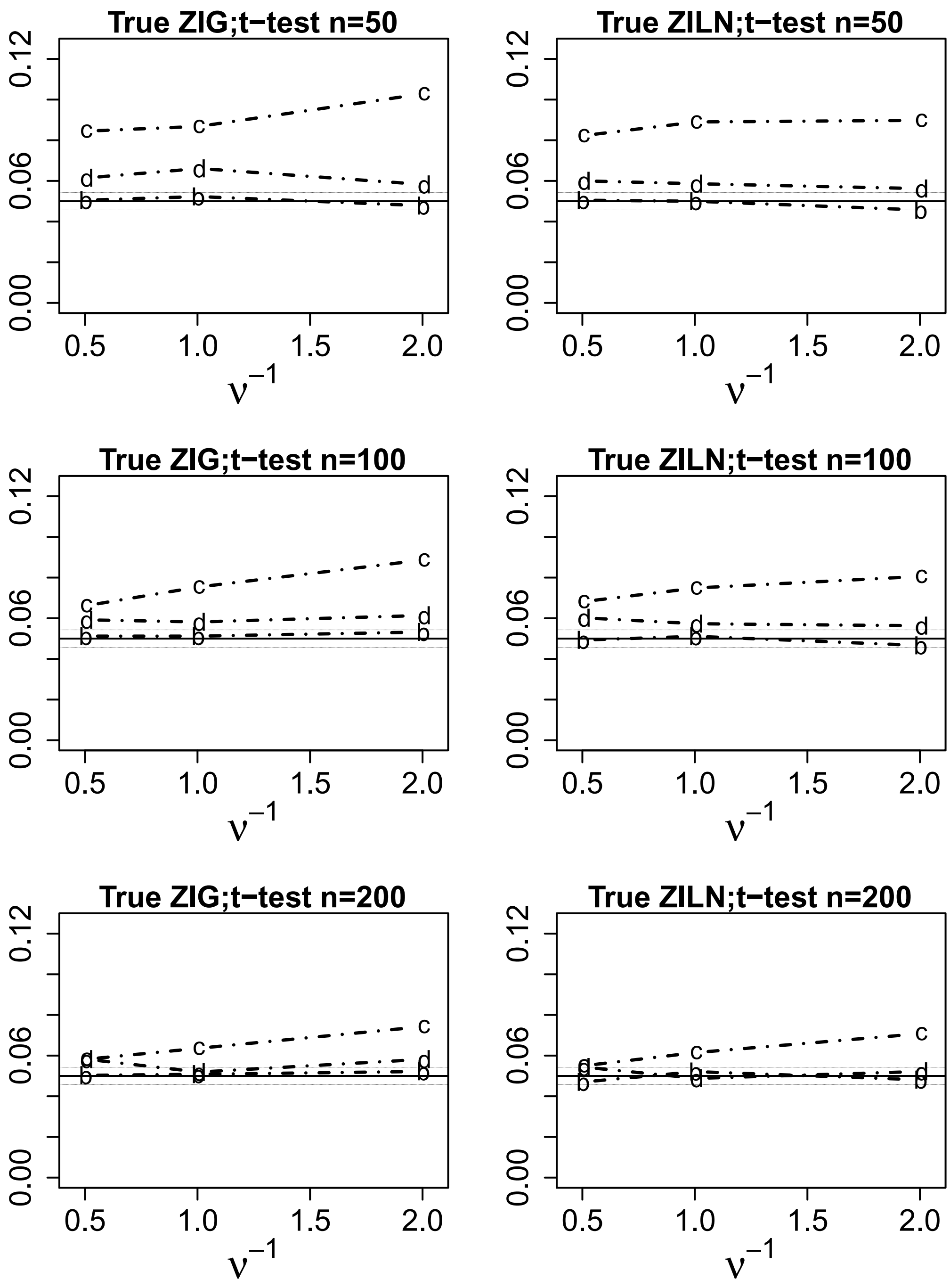

Figure 3.6: Type 1 error for t-tests; dissonant effects; two group comparison 'without covariate adjustment's. Symbols correspond to the setting in Table 2.2 where setting 'b' indicates $P(Y > 0|group = 0) = 0.75$ and $P(Y > 0|group = 1) = 0.5$, setting 'c' indicates $P(Y > 0|group = 0) = 0.75$ and $P(Y > 0|group = 1) = 0.25$, and setting 'd' indicates $P(Y > 0|group = 0) = 0.5$ and $P(Y > 0|group = 1) = 0.25$.

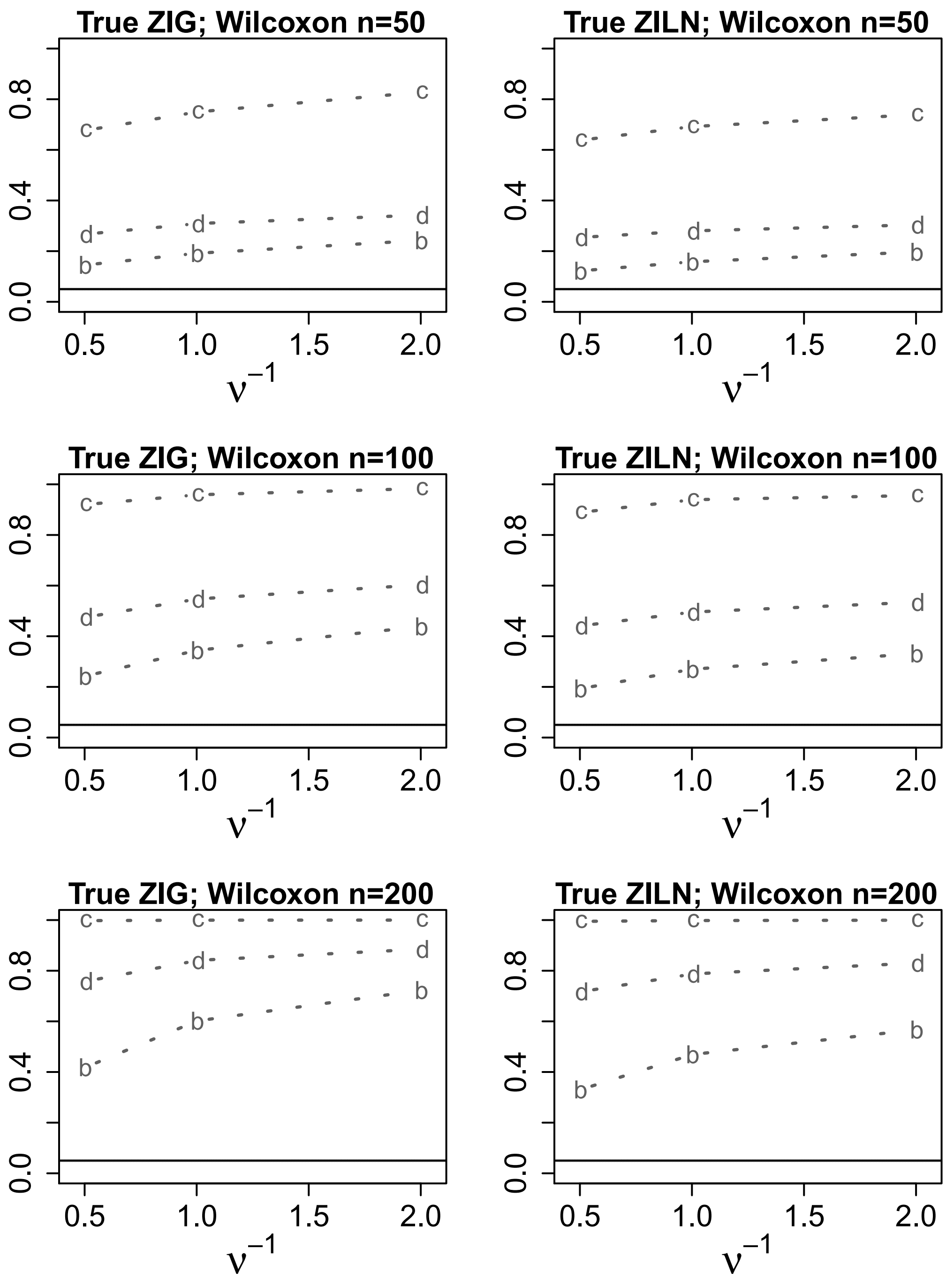

Figure 3.7: Power for Wilcoxon rank sum tests; dissonant effects. Symbols correspond to the setting in Table 2.2 where setting 'b' indicates $P(Y > 0|group = 0) = 0.75$ and $P(Y > 0|group = 1) = 0.5$, setting 'c' indicates $P(Y > 0|group = 0) = 0.75$ and $P(Y > 0|group = 1) = 0.25$, and setting 'd' indicates $P(Y > 0|group = 0) = 0.5$ and $P(Y > 0|group = 1) = 0.25$.

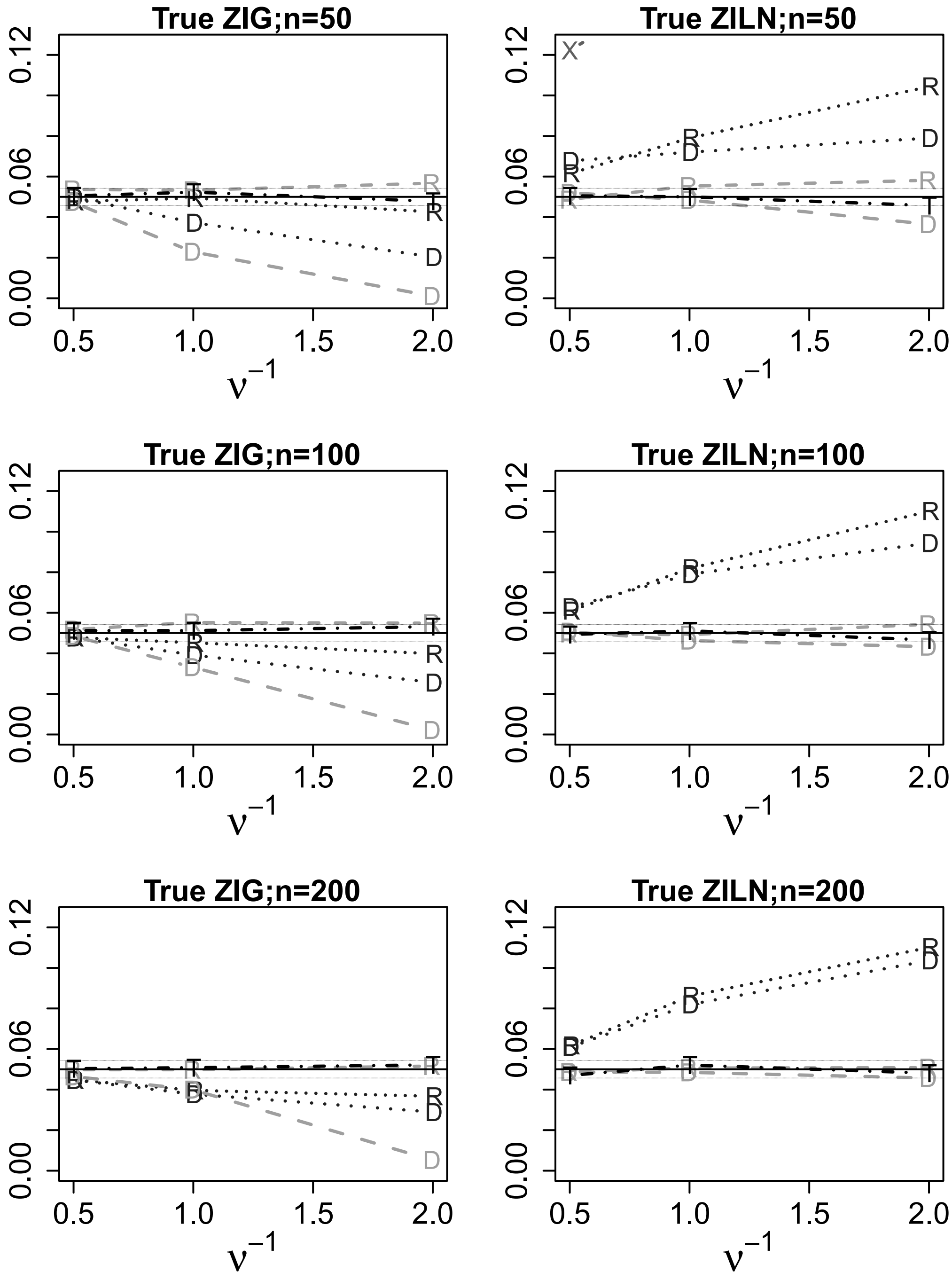

Figure 3.8: Type 1 error for mean-based tests with dissonant effects; setting 'b'; two group comparison 'without covariate adjustment's. 'D' refers to DM test, 'R' refers to RM test, 'T' refers to the t-test, and 'X' refers to the Wilcoxon rank sum test. For DM and RM, darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

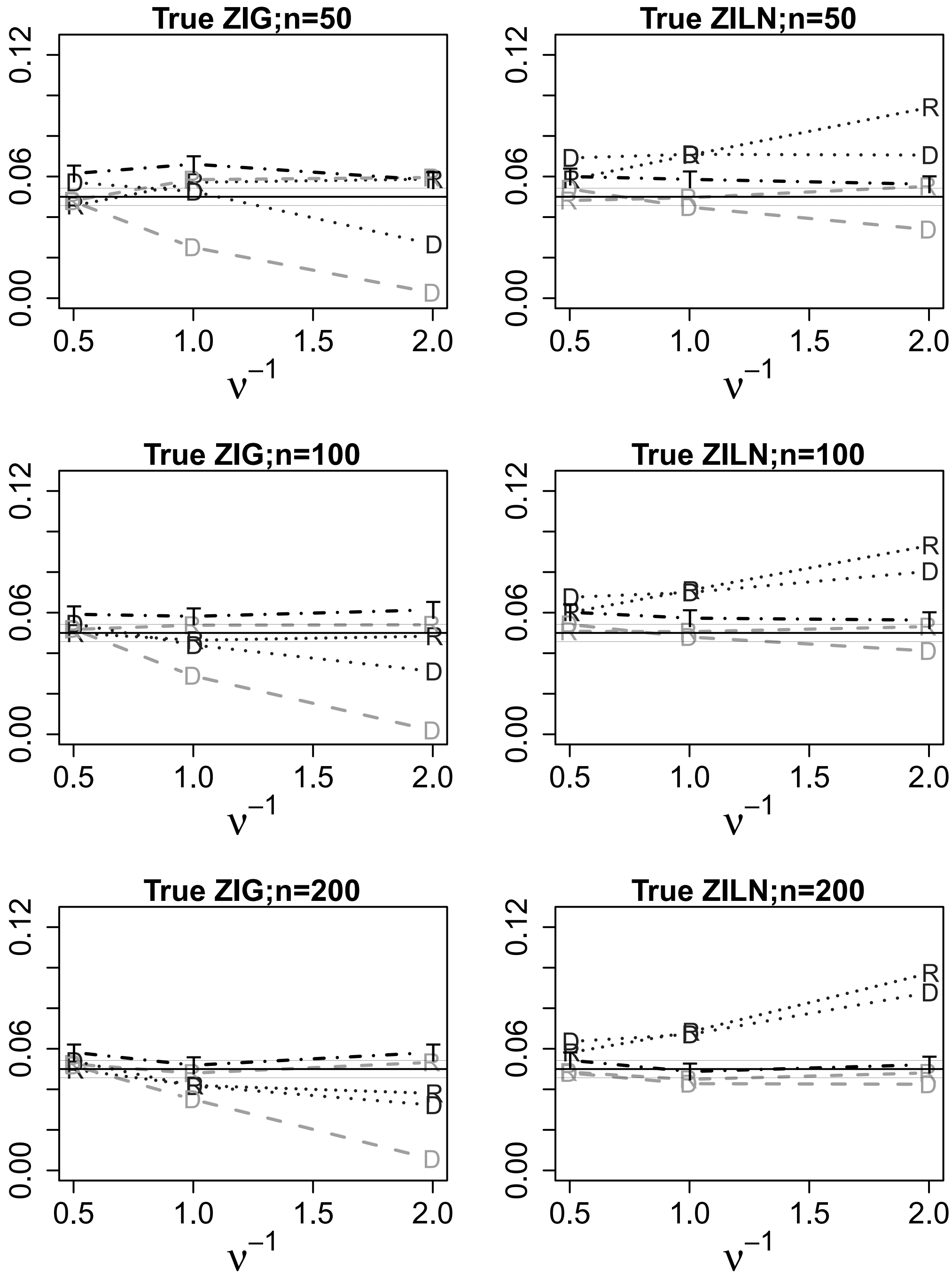

Figure 3.9: Type 1 error for mean-based tests with dissonant effects; setting 'd'; two group comparison 'without covariate adjustment's. 'D' refers to DM test, 'R' refers to RM test, 'T' refers to the t-test, and 'X' refers to the Wilcoxon rank sum test. For DM and RM, darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

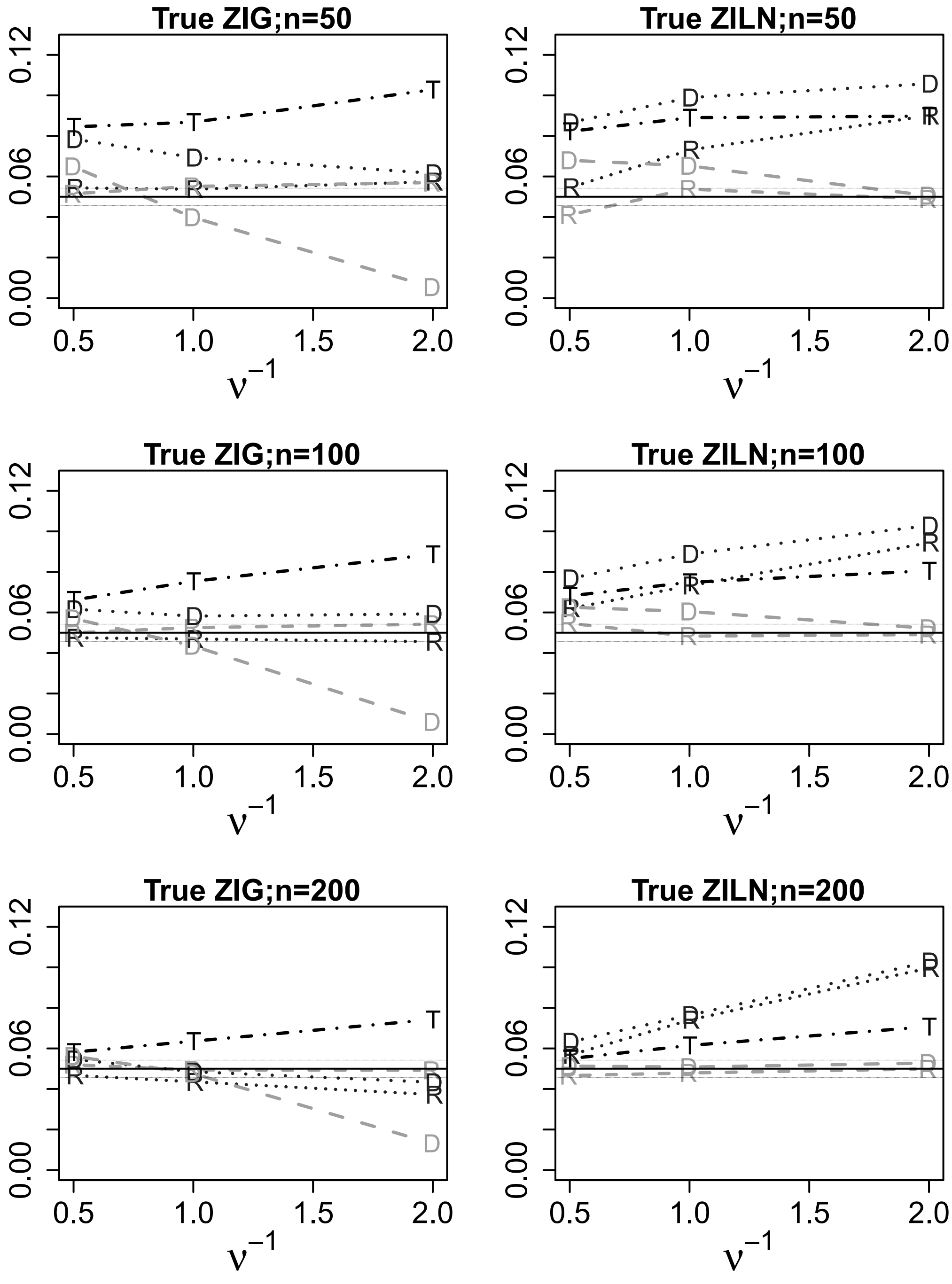

Figure 3.10: Type 1 error for mean-based tests with dissonant effects; setting 'c'; two group comparison 'without covariate adjustment's. 'D' refers to DM test, 'R' refers to RM test, 'T' refers to the t-test, and 'X' refers to the Wilcoxon rank sum test. For DM and RM, darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

to assuming a ZIG distribution, the light grey with dashed lines refers to assuming a ZILN distribution, and the t-test is shown with a black line and an alternating dash dot pattern. For all settings, when the data are simulated as ZIG the DM test assuming ZILN has the most extremely low Type 1 errors. When the data are from a ZILN distribution, the ZILN analyses for both DM and $\log(RM)$ tend to be closer to nominal than those assuming ZIG for the same metric. Under ZIG data and setting 'c', the setting with the largest dissonant effects and greatest number of zeros (see Figure 3.10), the t-test and DM assuming ZIG have high Type 1 errors, with the t-test having the more extremely high Type 1 error rates. Setting 'd' shown in Figure 3.9 has slightly elevated Type 1 errors for the t-test as well. When the data are ZILN and the setting is 'b' (Figure 3.8), when the data are analyzed as ZIG $\log(RM)$ yields higher Type 1 errors than DM; for setting 'd', Figure 3.9, when the data are analyzed as ZIG $\log(RM)$ has higher Type 1 errors than DM for $\nu^{-1} = 2$, but DM has higher Type 1 errors than RM when ν^{-1}=0.5. For setting 'c', Figure 3.10, when ZILN data were analyzed as ZIG, DM had the highest Type 1 errors at all levels of ν^{-1}; the t-test also has elevated Type 1 errors for this setting when the data are simulated from ZILN.

3.6.3 Power for Consonant Effects

Power results for DM, $\log(RM)$, t-tests, and Wilcoxon rank sum tests are shown in Figures 3.11 - 3.15. Specifically, Figure 3.11 and Figure 3.12 include powers for the DM based test statistic when ν^{-1} equals 0.5 and 2 respectively, Figure 3.13 includes powers for the $\log(RM)$ based test statistic when $\nu^{-1} = 2$, and Figure 3.14 and Figure 3.15 include power results for all four tests for group 0 setting 'B' when ν^{-1} equals 0.5 and 2 respectively. In Figures 3.11 - 3.13 the plotting symbols represent the group 0 mean settings, as in previous sections, with 'A' when $p_0 = 0.75$, 'B' when $p_0 = 0.5$, and 'C' when $p_0 = 0.25$. In the figures which compare power across tests, Figures 3.14 and 3.15, the plotting symbols represent the test statistic used with 'D' for the DM test statistics, 'R' for the $\log(RM)$ test statistics, 'T' for the t-test, and 'X' for the Wilcoxon rank sum test. In all of the figures, plots on the left contain data simulated as ZIG and plots on the right include data simulated as ZILN. Darker grey with dotted lines refers to analyses which assume ZIG distributions and the lighter grey with dashed lines refers to analyses which assume ZILN distributions.

Power for DM based tests when $\nu^{-1} = 0.5$ are presented in Figure 3.11. Regardless of the effect varied to reach the group 1 means, power increases as level of zero-inflation decreases; setting 'A' has higher power levels, than setting 'B', followed by setting 'C'. Within each type of setting, power increases as M_1 value (and hence the DM value) increases. For all levels of zero-inflation, when the group 1 mean is increased relative to the group 0 mean through effects in both the binomial and continuous parts of the model, power is higher than when the increased mean is entirely in the continuous part of the model (entirely due to an increased τ_1 value). This is different than what was observed for the two-part models where the settings with the non-zero effects split across both parts of the model had lower power levels for 'B' and 'C' settings than those based solely on the continuous part. In these DM power results, the highest powers were seen when β_1 was varied, but these powers are not as high as those seen for two-part tests in Chapter 2. In this figure, where $\nu^{-1} = 0.5$ misspecification has only a small impact on power with correctly specified analyses having slightly higher power levels than misspecified analyses. When $\nu^{-1} = 0.5$, results for the power of the test statistic based on $\log(RM)$ were very similar to those seen for DM. As such a figure for $\log(RM)$ power results when $\nu^{-1} = 0.5$ is not included, but $\log(RM)$ results are included in Figures 3.14 and 3.15 which compare power across test statistics.

Figure 3.12 illustrates the power results for DM based tests when $\nu^{-1} = 2$. In comparing Figure 3.12 to Figure 3.11 it is clear that power decreases as ν^{-1} increases.

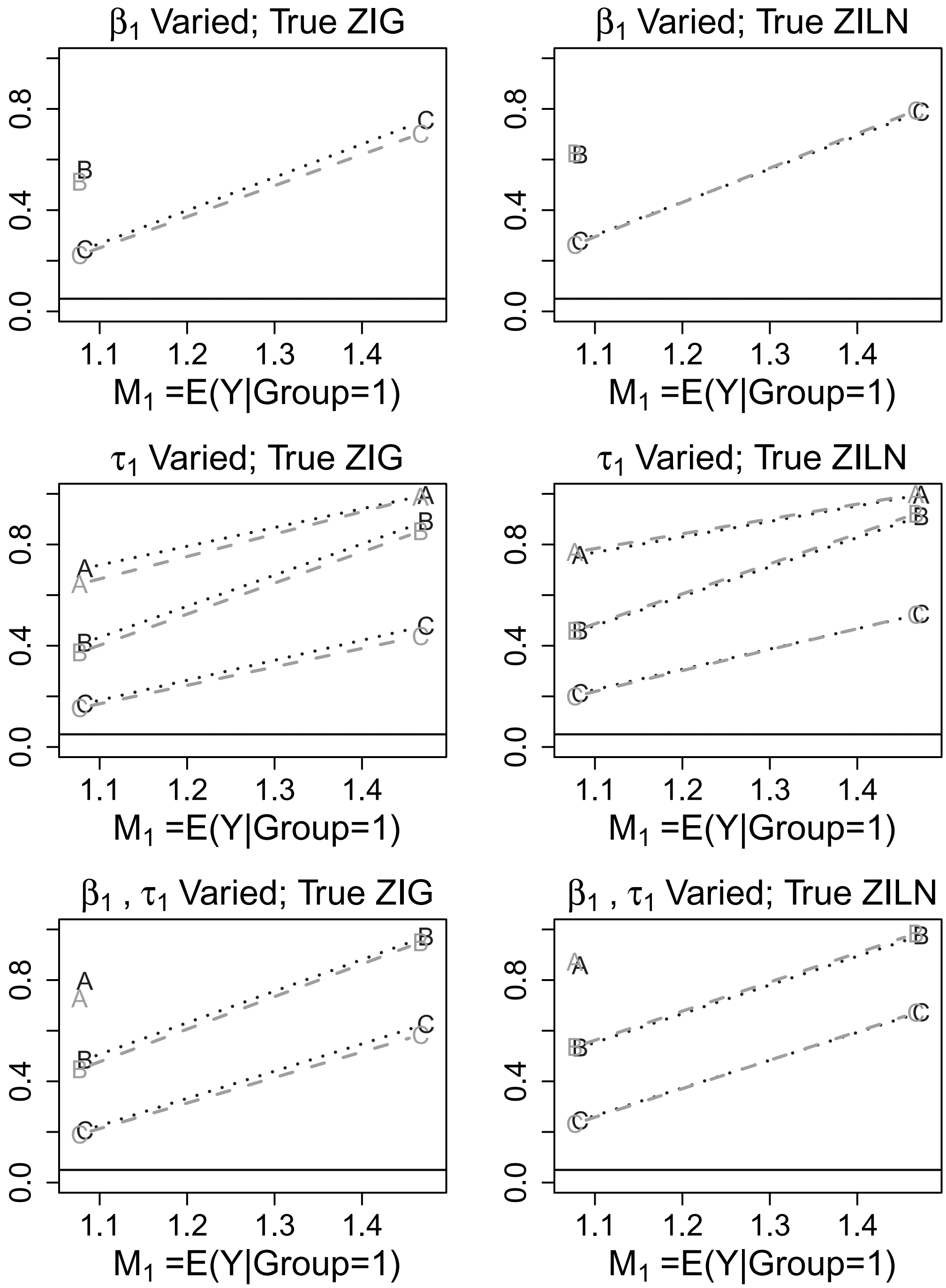

Figure 3.11: Power for DM based tests when $\nu^{-1} = 0.5$ and n=200; two group comparison 'without covariate adjustment'. Symbols correspond to the group 0 settings with 'A' when $P(Y > 0|group = 0) = 0.75$, 'B' when $P(Y > 0|group = 0) = 0.5$, and 'C' when $P(Y > 0|group = 0) = 0.25$. The overall mean of the second group is on the x-axis see Table 3.3 for more setting details. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

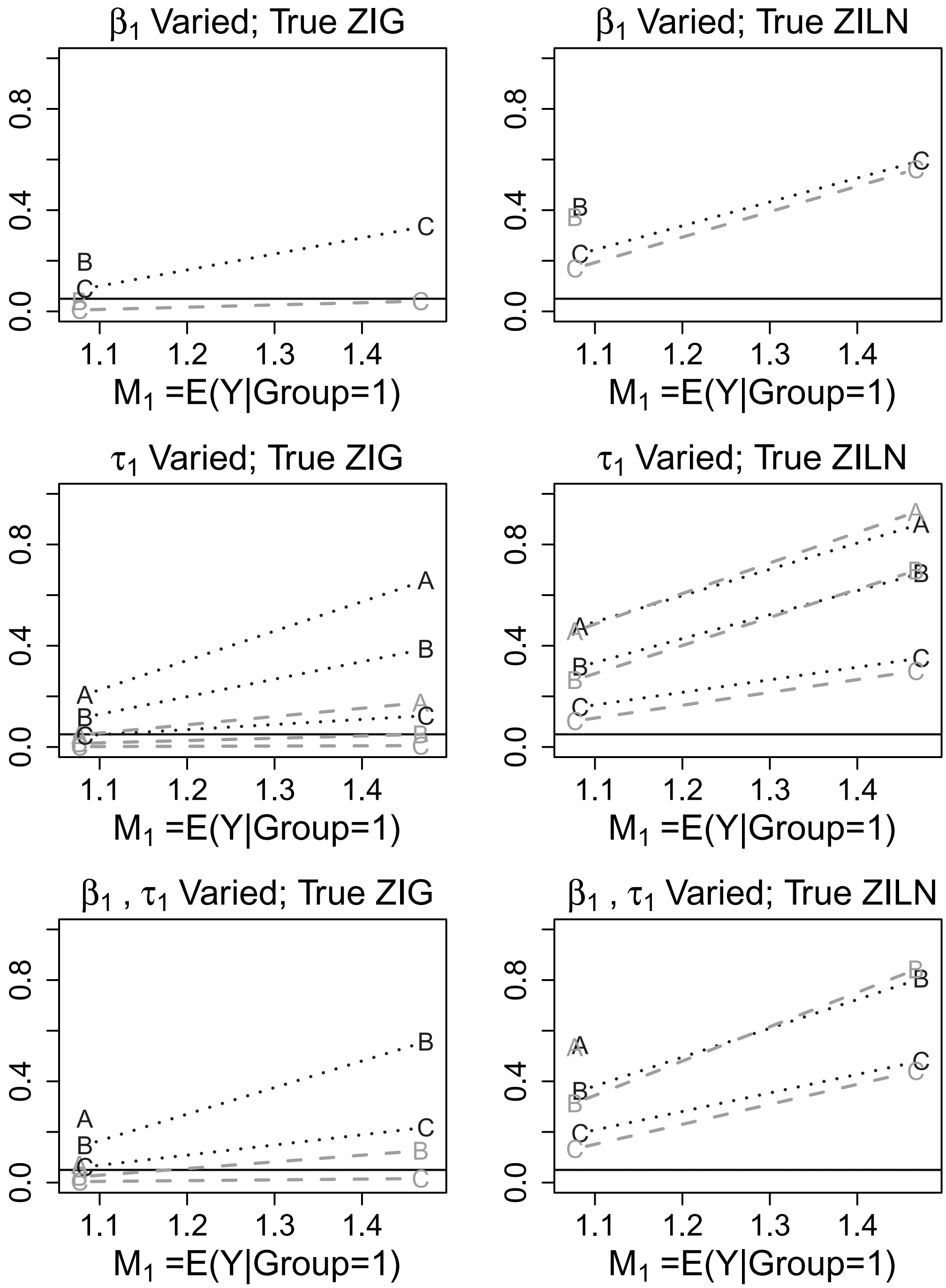

Figure 3.12: Power for DM based tests for $\nu^{-1} = 2$ and n=200; two group comparison 'without covariate adjustment'. Symbols correspond to the group 0 settings with 'A' indicates $P(Y > 0|group = 0) = 0.75$, 'B' indicates $P(Y > 0|group = 0) = 0.5$, and 'C' indicates $P(Y > 0|group = 0) = 0.25$. The overall mean of the second group is on the x-axis. See Table 3.3 for more setting details. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

As in previous results it can be seen that power decreases as the level of zero inflation increases. When the data are simulated from a ZIG distribution, power levels for analyses assuming ZIG are considerably higher than power for analyses assuming ZILN. Power levels for ZILN analyses of ZIG data are extremely low; for example, for setting 'C' power is below 0.05 even for the largest DM values which correspond to the highest values of M_1. When the data are ZILN, power is sometimes higher when a ZIG distribution is assumed and sometimes higher when a ZILN distribution is assumed. However, at this high level of ν^{-1} ZIG analyses are not acceptable for ZILN data due to high Type 1 error rates(see Section 3.6.1).

Figure 3.13 shows the power results for the $\log(RM)$ based test statistic when $\nu^{-1} = 2$. When the data are simulated from a ZILN distribution, the power results are very similar to those seen for DM, with only small differences in power between ZIG and ZILN analyses. The power results for the data simulated as ZIG show smaller differences between the powers for ZIG and ZILN analyses than were seen with DM when $\nu^{-1} = 2$. Correctly specified ZIG analyses using $\log(RM)$ test statistics do have higher power than the misspecified ZILN analyses, but the difference is not as extreme as it was for DM. Some of the reduced power is due to the misspecified model and might show up regardless of the metric used with that model. One the other hand, for DM the power was even lower, which is likely due to instability in the estimation of σ^2.

Figure 3.14 compares the power results for DM tests, $\log(RM)$ tests, t-tests, and Wilcoxon rank sum tests when $p_0 = 0.5$ and $\nu^{-1} = 0.5$. From this figure it is clear that for this small value of coefficient of variation, power for DM and $\log(RM)$ tests under both ZIG and ZILN distributional assumptions, as well as for t-tests, are very similar. There are slight differences, but nothing substantial. When the data are truly ZIG, powers for tests assuming ZILN were slightly lower than powers for those assuming ZIG; most particularly when only τ_1 is varied, the ZILN DM tests

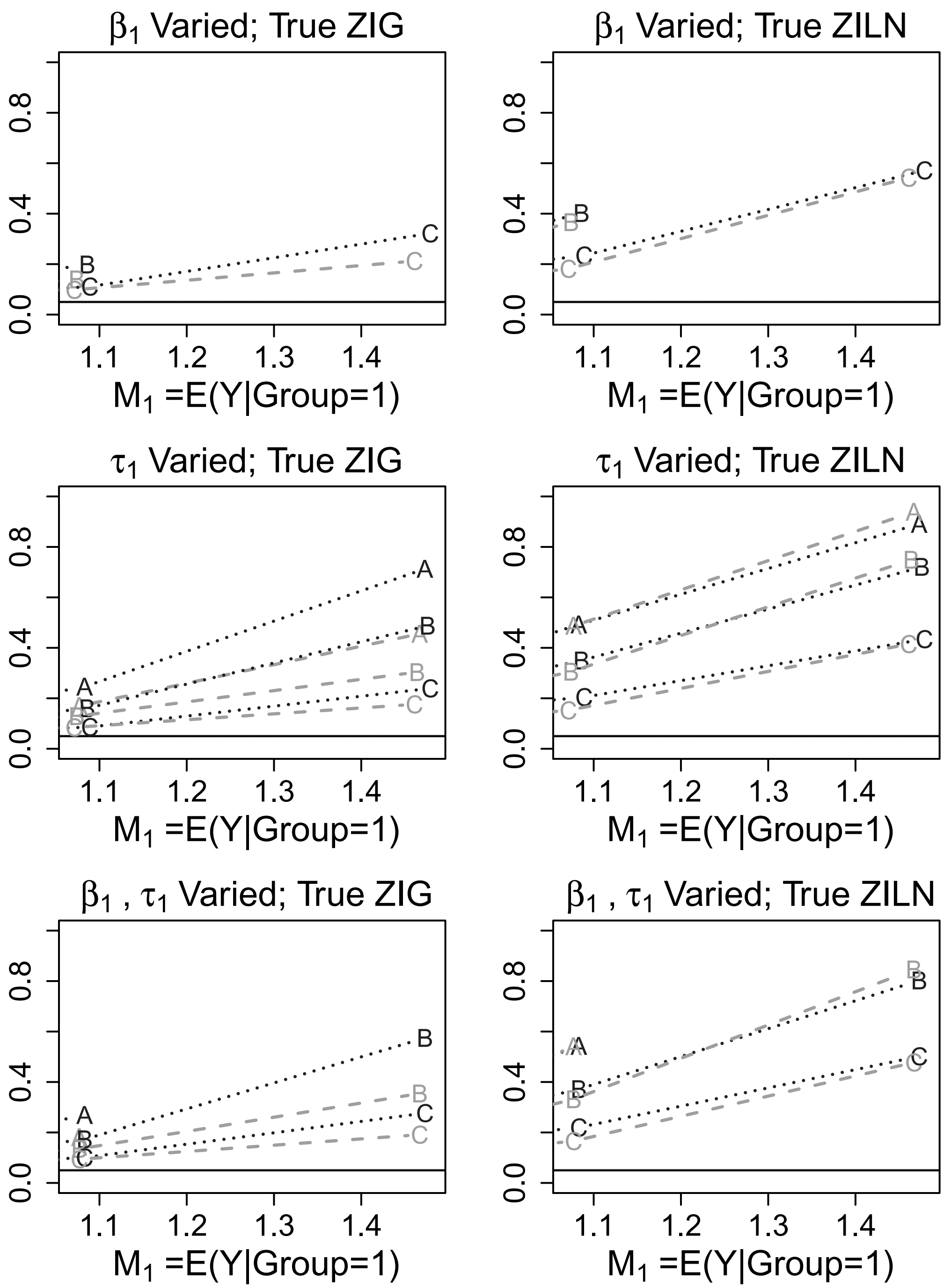

Figure 3.13: Power for $\log(RM)$ based tests for $\nu^{-1} = 2$ and n=200; two group comparison 'without covariate adjustment'. Symbols correspond to the group 0 settings with 'A' indicates $P(Y > 0|group = 0) = 0.75$, 'B' indicates $P(Y > 0|group = 0) = 0.5$, and 'C' indicates $P(Y > 0|group = 0) = 0.25$. The overall mean of the second group is on the x-axis. See Table 3.3 for more setting details. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

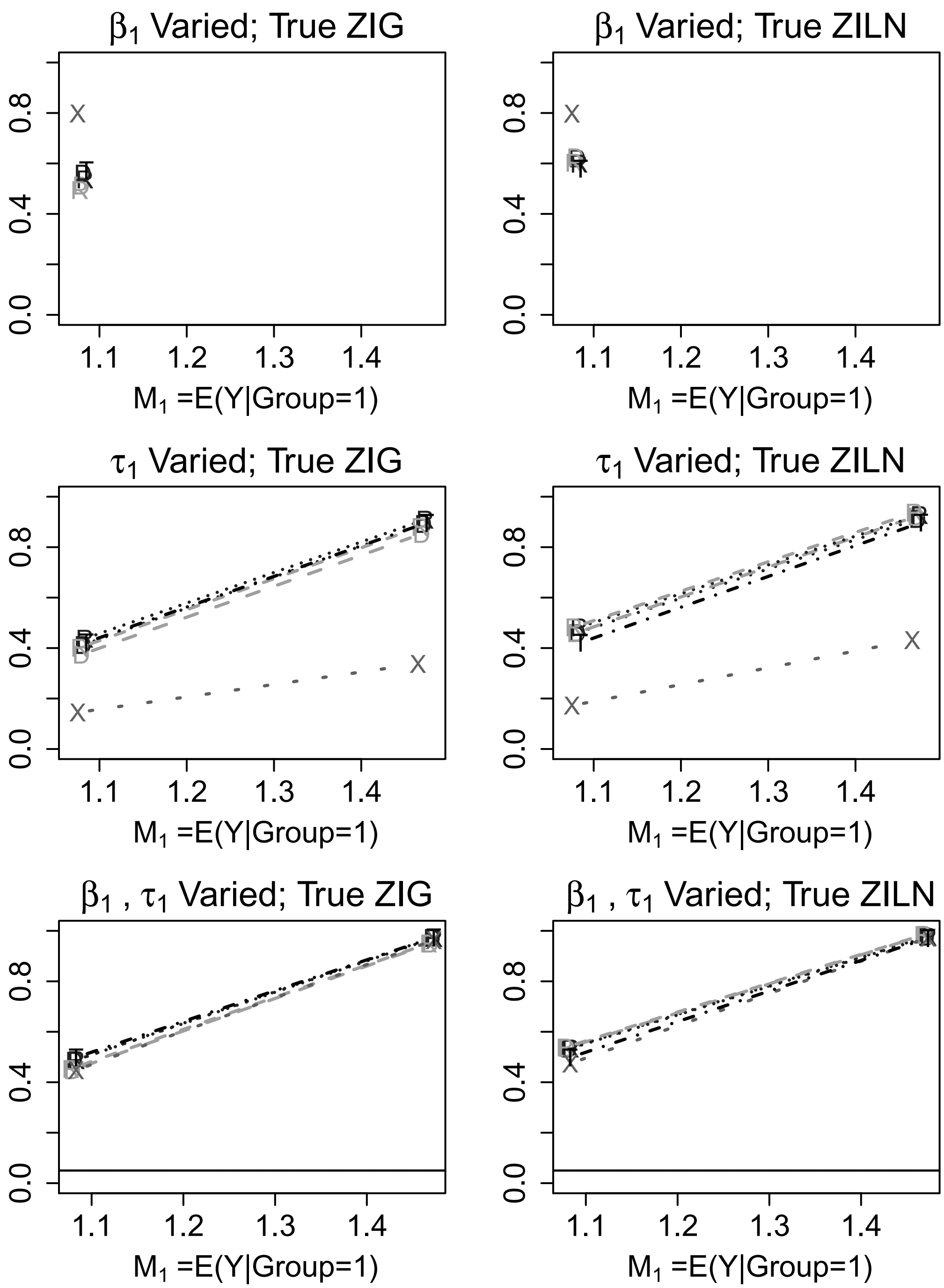

Figure 3.14: Power for mean-based tests using setting 'B' and $\nu^{-1} = 0.5$ where $P(Y > 0|group = 0) = 0.5$; two group comparison 'without covariate adjustment'. 'D' refers to DM test, 'R' refers to RM test, 'T' refers to the t-test, and 'X' refers to the Wilcoxon rank sum test. For DM and RM, darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

had lower power levels than the ZIG DM and $\log(RM)$ tests and the t-tests. For data simulated from a ZILN distribution, t-tests had slightly lower power than the ZIG and ZILN based tests especially when the effect was in the continuous part of the model. The Wilcoxon rank sum test had very different power results from those seen for the mean-based tests. When the effect of group was in the binomial part of the model, power for the Wilcoxon rank sum test was higher than for the other tests. When the effect of group was only in the continuous part, the Wilcoxon rank sum test yielded much lower power levels than the other tests. Finally, when the effect was in both parts of the model the power levels observed for the Wilcoxon rank sum test were closer to the powers obtained by the mean-based tests, although slightly on the lower end.

From Figure 3.15 it is seen that when $\nu^{-1} = 2$ the differences in power between the various tests become more pronounced. When data are simulated from ZIG distributions, analyses assuming ZIG have higher power levels than analyses assuming ZILN, and DM-based tests have higher power levels than RM-based tests. For such data, the misspecified ZILN analyses have almost no power. When the data are simulated from ZILN distributions, the differences are less pronounced than for ZIG, and and the inappropriate ZIG analyses tend to have slightly higher power for the smaller effects and the ZILN analyses have slightly higher power for the larger effects. The power for finding group differences using Wilcoxon rank sum tests is higher than the other tests when there is a nonzero β_1 effect, and lower than all but the misspecified ZILN DM tests when the nonzero effect of group is only in the continuous part of the model. T-tests have power levels that are closer to those of the two-part model mean-based tests,tending to be slightly higher when the data are simulated from ZIG analyses and slightly lower when data are simulated from ZILN analyses.

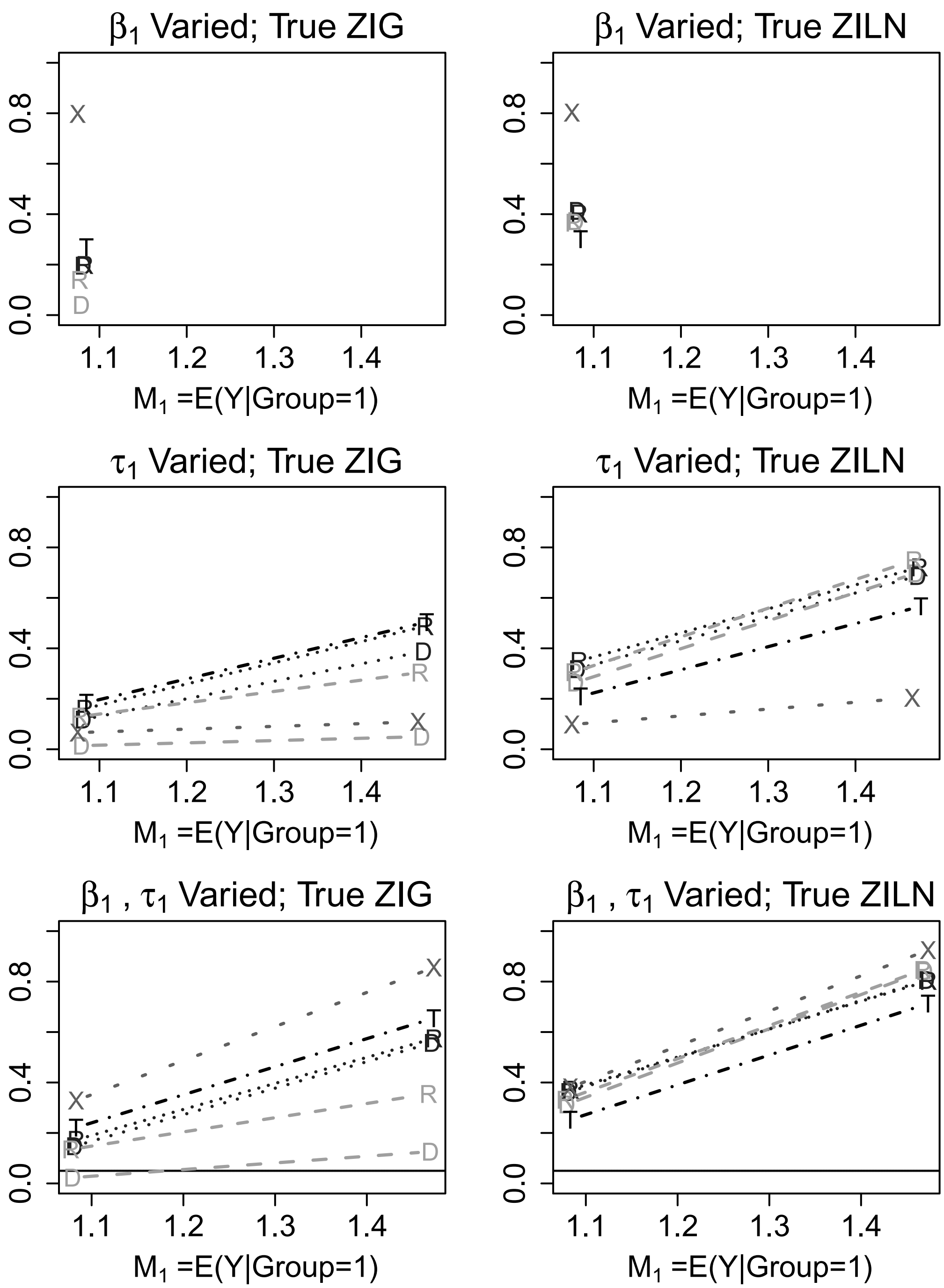

Figure 3.15: Power for mean-based tests using setting 'B' and $\nu^{-1} = 2$ where $P(Y > 0|group = 0) = 0.5$; two group comparison 'without covariate adjustment'. 'D' refers to DM test, 'R' refers to RM test, 'T' refers to the t-test, and 'X' refers to the Wilcoxon rank sum test. For DM and RM, darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

3.7 Summary

In this chapter, the overall mean of group in terms of the parameters of the ZIG and ZILN models were defined. Difference in mean and ratio of mean tests for finding a difference between the overall group means were proposed. Using the same simulated data sets as in Chapter 2 Type 1 error rates using settings with no group differences and settings with dissonant group effects that canceled out in terms of the group mean were examined. A final set of simulations (also the same as in Chapter 2) was used to examine power levels.

The simulation study shows that when DM-based tests are used to analyse data simulated as ZIG, Type 1 error rates are very conservative. This is especially true for analyses assuming ZILN and simulations where ν^{-1} is high. For analyses correctly assuming ZIG, Type 1 error rates are conservative but improve slowly with sample size. This slow improvement continues with even larger sample sizes, indicating that slow asymptotic properties of this Wald test may be at least one reason for this over conservativeness. When data are simulated from ZILN distributions, Type 1 error rates are conservative for analyses assuming ZILN when ν^{-1} is high. These Type 1 error rates come closer to nominal as sample size increases. When data are simulate from ZILN distributions and ZIG distributions are assumed, Type 1 error rates are elevated and become increasingly so as sample size increases. Type 1 error rates under the dissonant effect settings exhibit somewhat similar patterns. The main exception being slightly elevated Type 1 error rates for correctly specified ZIG analyses and correctly specified ZILN analyses when the setting has the largest dissonant effects and the sample size is small.

Tests based on RM are within the appropriate Type 1 error range for analyses assuming ZILN except for a small sample size case. When assuming ZIG, Type 1 error rates are slightly conservative (again except for a small sample size case) when the data are truly ZIG, and elevated when the data are truly ZILN. These RM results hold for the settings with no group differences and the settings with no overall group mean differences but with dissonant effects.

Type 1 error rates for t-tests are within range to slightly conservative when the two groups are equal with respect to both the probability of a non-zero outcome and the conditional mean. On the other hand, when there are dissonant effects Type 1 error rates for t-tests can be inappropriately high. However, the Type 1 error rates even for strong dissonant effects improve with increased sample size. Wilcoxon rank sum tests have appropriate Type 1 error rates when there are no differences between the groups. On the other hand, when there are dissonant effects, Wilcoxon rank sum tests have substantial power to find differences between the two groups, increasing with sample size.

Power was examined at two levels of DM and RM. The DM and RM values were achieved in three different ways. Namely, through group differences in the probability of a non-zero outcome, group differences in the conditional mean, and through group differences stemming from both the probability of a non-zero outcome and the conditional mean. These were also examined at different levels of ν^{-1}. When $\nu^{-1} = 0.5$, DM and RM tests assuming ZIG and ZILN as well as the t-test, had similar power levels for data simulated from both ZIG and ZILN distributions. When $\nu^{-1} = 2$ and the data were simulated from ZIG distributions, tests assuming ZIG and t-tests had the highest power levels, and the DM tests assuming ZILN had essentially no power. When $\nu^{-1} = 2$ and the data were simulated from ZIG distributions, analyses assuming both ZIG and ZILN distributions had slightly higher power than t-tests. The power of the Wilcoxon rank sum test changed greatly with the location of the group effect. Wilcoxon rank sum tests had the highest power levels when one group has more non-zero values than the other, and the lowest power levels for detecting differences in the group means of the non-zero values.

CHAPTER 4
MEAN-BASED TWO GROUP COMPARISON TESTS WITH COVARIATE ADJUSTMENT UNDER ZERO INFLATED GAMMA AND ZERO INFLATED LOG NORMAL DISTRIBUTIONS

4.1 Adjusting for Covariates in Mean-Based Tests

Most of the research involving mean-based tests created from two-part models, both for the difference of means (DM) and the ratio of means (RM), has not addressed the issue of adjusting for covariates. An exception to this was a paper by Dominici and Zeger [20] where they proposed several methods of adjusting for covariates when calculating DM under a ZILN framework. Additionally, Tooze et al.[13] used a two-part ZILN model in a repeated measures framework. They noted that the ratio of means varies with the values of the other covariates in the model. Neither of these papers addressed methods for estimation of DM or tests relating to DM under ZIG, and we are unaware of any authors who have looked at RM estimation or tests under ZIG or ZILN models.

Covariates can be easily included in two-part models but add some complication to the creation of mean-based tests. The ratio of group means or the difference in group means resultant from two-part models will differ depending on the covariate values. Therefore, to create an overall mean comparison, the difference of mean or ratio of means must be marginalized over the values of the covariates. The ratio of means could be marginalized in one of two ways, either through marginalizing the subject specific ratios of means (RM_{SS}), or through marginalizing the means and then taking the ratio (RM_{MAR}). In this chapter, we propose Wald-type tests for DM, RM_{SS}, and RM_{MAR}. Type 1 error rates and power levels for these tests will be compared under two frameworks, adjusting for a dichotomous covariate and adjusting for a continuous covariate. In reference to our recurring example of driving data, the methods in this chapter correspond to comparing average lane departure severity for

Parkinson's disease and control subjects when adjusting for gender or when adjusting for age.

4.2 Marginal Means of Zero Inflated Gamma and Zero Inflated Log-Normal Distributions Adjusting for Covariates

In the context of adjusting for covariates, two of the metrics defined later in this chapter (DM and RM_{MAR}) stem from a null hypothesis that the mean of the semicontinuous outcome given group 0 marginalized over the covariates is equal to the mean of the semicontinuous outcome given group 1 marginalized over the covariates. We begin here by defining these marginal means. We will define X_1 to be the indicator of group, and Z to be the adjusting covariate. The marginal group means can then be defined as:

$$E(Y|X_1 = x_1) = E_Z(E(Y|X_1 = x_1, Z)). \tag{4.1}$$

The marginalization over Z can be over either a theoretical distribution of Z or over an observed distribution of Z. In practice, the distribution used should always be presented (using, for example, summary statistics or histograms) so that the generalizability of the mean with respect to the adjusting covariate is clear. For this dissertation, we will marginalize over the observed distribution of the covariates. In doing so, the marginal group mean can be expressed as follows:

$$\begin{aligned} M_j &= E(Y|X_1 = j) = E_Z(E(Y|X_1 = x_1, Z)) \\ &= \frac{1}{n}\sum_i^n E(Y|X_1 = x_1, Z_i) \\ &= \frac{1}{n}\sum_i^n P(Y > 0|X_1 = x_1, Z_i)E(Y|Y > 0, X_1 = x_1, Z_i) \end{aligned} \tag{4.2}$$

where $x_1 = j$ represents the group for which the mean is being calculated for $j = 0, 1$, z_i equals the value of the adjusting covariates for subject i with $i = 1, 2, \ldots,$ n, and n is the total number of subjects.

To understand the implication of marginalizing over the observed distribution, consider the following example. When adjusting for gender, marginalizing over the observed distribution of males and females leads a marginal mean that is a weighted average of the gender-specific means with the weights for each corresponding to the proportion of males and females in the entire sample. On the other hand, marginalizing in terms of a theoretical distribution assuming equal numbers of each gender would make those weights equal regardless of the number of males and females in the sample. In other instances, however, determining an appropriate theoretical distribution may not be feasible and it may make sense to consider a sample distribution as representative of the true distribution of interest and use it for the marginalization. In this dissertation, we marginalize over the observed covariate distributions for both continuous and dichotomous covariate outcomes.

These methods could be generalized to adjust for any number of covariates. However, the scope of this dissertation will include only adjusting for one covariate at a time. Following the marginalization framework in Equation 4.2 and the ZIG and ZILN regression models outlined in Section 2.2, we define τ_0 as an intercept term, τ_1 as a group effect term, and τ_2 as the effect of the adjusting covariate term for the continuous part of the model. Similarly, we define β_0 as an intercept term, β_1 as a group effect term, and β_2 as the term corresponding to the adjusting covariate for the logistic regression portion of the model. Then the marginal mean given group membership for ZIG regression model is

$$M_j = E(Y|X_1 = x_1) = \frac{1}{n}\sum_{i}^{n} \frac{e^{\beta_0+\beta_1 x_1+\beta_2 z_i}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 z_i}} e^{\tau_0+\tau_1 x_1+\tau_2 z_i} \tag{4.3}$$

where $j = 0, 1$ is the indicator of group membership such that $x = j$.

For ZILN regression, the marginal mean given group membership is:

$$M_j = E(Y|X_1 = x_1) = \frac{1}{n}\sum_{i}^{n} \frac{e^{\beta_0+\beta_1 x_1+\beta_2 z_i}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 z_i}} e^{\tau_0+\tau_1 x_1+\tau_2 z_i+\frac{\sigma^2}{2}}. \tag{4.4}$$

These marginal means are found by averaging over the expected outcome for each subject regardless of true group membership given the observed covariate values and membership in group j.

4.3 Mean Difference

When a two-part model includes covariates other the than variable of interest, DM must be calculated by marginalizing over the distribution of the adjusting covariate. Because of the distributive property of the expectation function with respect to subtraction, the DM calculated is identical whether found by either averaging over the subject specific differences of means or by subtracting one marginal group mean from the other. Hence, the difference of means when adjusting for covariates can be defined as:

$$\begin{aligned} DM &= M_1 - M_0 = E(Y|X_1 = 1) - E(Y|X_1 = 0) \\ &= E_Z(E(Y|X_1 = 1, Z)) - E_Z(E(Y|X_1 = 0, Z)) \\ &= E_Z\left(E(Y|X_1 = 1, Z) - E(Y|X_1 = 0, Z)\right) \end{aligned} \tag{4.5}$$

where Z represents the covariates that are being adjusted for.

Section 4.3.1 and Section 4.3.2 will define DM in terms of ZIG and ZILN regression model parameters. In each of these sections, the asymptotic distribution of DM resulting from an application of the delta method will be presented. The proofs of these results (not shown) follow a pattern similar to that explored for the tests under the non-adjusting framework which is shown in the appendices. For the DM tests, the null hypothesis being tested is that the $M_1 = M_0$ or alternatively that $DM = 0$. Also, note that these estimates depend on the distribution of the covariates observed in the data set.

4.3.1 Mean Difference Assuming a Zero Inflated Gamma Distribution

The difference of two group means adjusted for and marginalized over covariates, Z, resulting from a ZIG regression model follows from the definition of the marginalized group means assuming ZIG as defined in Equation 4.3, and the definition of marginalized DM in Equation 4.5. Specifically,

$$\begin{aligned}
DM(\beta,\tau) &= E_Z(E(Y|X_1=1,Z) - E_Z(E(Y|X_1=0,Z)) \\
&= M_1 - M_0 \\
&= \frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_1+\beta_2 z_i}}{1+e^{\beta_0+\beta_1+\beta_2 z_i}} e^{\tau_0+\tau_1+\tau_2 z_i} - \frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_2 z_i}}{1+e^{\beta_0+\beta_2 z_i}} e^{\tau_0+\tau_2 z_i} \\
&= \frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_1+\beta_2 z_i}}{1+e^{\beta_0+\beta_1+\beta_2 z_i}} e^{\tau_0+\tau_1+\tau_2 z_i} - \frac{e^{\beta_0+\beta_2 z_i}}{1+e^{\beta_0+\beta_2 z_i}} e^{\tau_0+\tau_2 z_i} \\
&= \frac{1}{n}\sum_i^n (p_{1i}\mu_{1i} - p_{0i}\mu_{0i}),
\end{aligned} \tag{4.6}$$

where $p_{ji} = P(Y > 0|X_1 = j, Z_i)$ and $\mu_{ji} = E(Y|Y > 0, X_1 = j, Z_i)$ with $j = 0, 1$ as an indicator of group membership and $i = 0, 1, \ldots, n$ where n is the total number of subjects. Note that for each subject i, p_{ji} and μ_{ji} are calculated (given the observed covariates, z_i) for both $j = 0$ and $j = 1$.

Delta method techniques can be used to find an estimator for the variance of DM. Specifically, it can be shown (using methods similar to those in the appendices) that

$$\sqrt{n}\left(DM(\hat{\beta},\hat{\tau}) - DM(\beta,\tau)\right) \xrightarrow{d} N(0, \Delta'_{DM}\Sigma_{\beta,\tau}\Delta_{DM}). \tag{4.7}$$

To define Δ_{DM} and $\Sigma_{\beta,\tau}$ we first define some more notation. Let X be an n $\times$ 3 matrix consisting of a column of 1's, a column containing the observed X_1 (group membership) values, and a column containing the observed covariate values. Let p_i be the estimated probability of a nonzero outcome for subject i given the observed group membership and covariate values, $p_i = P(Y > 0|X_1 = x_{1i}, Z_i)$. Let D_p be an n × n diagonal matrix with p_i on the diagonals. Finally, let $D_{p(1-p)}$ being an n × n diagonal matrix composed of $p_i(1-p_i)$ on the diagonals. Then Δ_{DM} is a vector of derivatives of $DM = DM(\beta, \tau)$ such that:

$$\Delta_{DM} = \left(\frac{\delta DM}{\delta \beta_0} \quad \frac{\delta DM}{\delta \beta_1} \quad \frac{\delta DM}{\delta \beta_2} \quad \frac{\delta DM}{\delta \tau_0} \quad \frac{\delta DM}{\delta \tau_1} \quad \frac{\delta DM}{\delta \tau_2} \right)'$$
$$= \begin{pmatrix} \frac{1}{n}\sum_i^n \left(p_{1i}(1-p_{1i})\mu_{1i} - p_{0i}(1-p_{0i})\mu_{0i}\right) \\ \frac{1}{n}\sum_i^n \left(p_{1i}(1-p_{1i})\mu_{1i}\right) \\ \frac{1}{n}\sum_i^n \left(z_i p_{1i}(1-p_{1i})\mu_{1i} - z_i p_{0i}(1-p_{0i})\mu_{0i}\right) \\ \frac{1}{n}\sum_i^n \left(p_{1i}\mu_{1i} - p_{0i}\mu_{0i}\right) \\ \frac{1}{n}\sum_i^n p_{1i}\mu_{1i} \\ \frac{1}{n}\sum_i^n \left(z_i p_{1i}\mu_{1i} - z_i p_{0i}\mu_{0i}\right) \end{pmatrix}, \tag{4.8}$$

and

$$\Sigma_{\beta,\tau} = \begin{pmatrix} n(X'D_{p(1-p)}X)^{-1} & \mathbf{0} \\ \mathbf{0} & \nu^{-1}(X'D_pX)^{-1} \end{pmatrix}. \tag{4.9}$$

Using the estimate of $DM(\beta, \tau)$ defined in Equation 4.6 and the estimates of Δ_{DM} and $\Sigma_{\beta,\tau}$ defined in Equations 4.9 and 4.8, a Wald-type test statistics can be created as

$$D(\hat{\beta}, \hat{\tau}) = \frac{DM(\hat{\beta}, \hat{\tau})}{\sqrt{n\Delta'_{\hat{DM}}\Sigma_{\hat{\beta},\hat{\tau}}\Delta_{\hat{DM}}}}. \tag{4.10}$$

This Wald-type test statistic can be compared to a standard normal distribution.

4.3.2 Mean Difference Assuming a Zero Inflated Log-Normal Distribution

The marginalized DM for ZILN regression can be created using Equations 4.4 and 4.5 as starting point. DM is then defined as:

$$\begin{aligned}
DM(\beta,\tau,\sigma^2) &= E_Z(E(Y|X_1=1,Z)) - E_Z(E(Y|X_1=0,Z)) = M_1 - M_0 \\
&= \frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_1+\beta_2 z_i}}{1+e^{\beta_0+\beta_1+\beta_2 z_i}} e^{\tau_0+\tau_1+\tau_2 z_i+\frac{\sigma^2}{2}} - \frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_2 z_i}}{1+e^{\beta_0+\beta_2 z_i}} e^{\tau_0+\tau_2 z_i+\frac{\sigma^2}{2}} \\
&= \frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_1+\beta_2 z_i}}{1+e^{\beta_0+\beta_1+\beta_2 z_i}} e^{\tau_0+\tau_1+\tau_2 z_i+\frac{\sigma^2}{2}} - \frac{e^{\beta_0+\beta_2 z_i}}{1+e^{\beta_0+\beta_2 z_i}} e^{\tau_0+\tau_2 z_i+\frac{\sigma^2}{2}} \\
&= \frac{1}{n}\sum_i^n (p_{1i}\mu_{1i} - p_{1i}\mu_{1i}).
\end{aligned} \tag{4.11}$$

As in Chapter 3 we assume a homogenous $\frac{\sigma^2}{2}$. This is equivalent to assuming a constant coefficient of variation. Using the delta method, it can be shown that:

$$\sqrt{n}\left(DM(\hat{\beta},\hat{\tau},\hat{\sigma^2}) - DM(\beta,\tau,\sigma^2)\right) \xrightarrow{d} N(0,\Delta'_{DM}\Sigma_{\beta,\tau,\sigma^2}\Delta_{DM}). \tag{4.12}$$

The terms in Equation 4.12 are defined in Equations 4.13 and 4.14. For the sake of space, we relax the notation to represent $DM(\beta,\tau,\sigma^2)$ with DM. Then,

$$\Delta_{DM} = \begin{pmatrix} \frac{\delta DM}{\delta\beta_0} & \frac{\delta DM}{\delta\beta_1} & \frac{\delta DM}{\delta\beta_2} & \frac{\delta DM}{\delta\tau_0} & \frac{\delta DM}{\delta\tau_1} & \frac{\delta DM}{\delta\tau_2} & \frac{\delta DM}{\delta\sigma^2} \end{pmatrix}'$$

$$\begin{pmatrix}
\frac{1}{n}\sum_i^n (p_{1i}(1-p_{1i})\mu_{1i} - p_{0i}(1-p_{0i})\mu_{0i}) \\
\frac{1}{n}\sum_i^n (p_{1i}(1-p_{1i})\mu_{1i}) \\
\frac{1}{n}\sum_i^n (z_i p_{1i}(1-p_{1i})\mu_{1i} - z_i p_{0i}(1-p_{0i})\mu_{0i}) \\
\frac{1}{n}\sum_i^n (p_{1i}\mu_{1i} - p_{0i}\mu_{0i}) \\
\frac{1}{n}\sum_i^n p_{1i}\mu_{1i} \\
\frac{1}{n}\sum_i^n (z_i p_{1i}\mu_{1i} - z_i p_{0i}\mu_{0i}) \\
\frac{1}{n}\sum_i^n (p_{1i}\mu_{1i} - p_{0i}\mu_{0i})
\end{pmatrix}, \tag{4.13}$$

and

$$\Sigma_{\beta,\tau,\sigma^2} = \begin{pmatrix} n(X'D_{p(1-p)}X)^{-1} & \mathbf{0} & 0 \\ \mathbf{0} & \sigma^2(X'D_pX)^{-1} & 0 \\ \mathbf{0} & \mathbf{0} & \frac{2\sigma^4}{n_0p_0+n_1p_1} \end{pmatrix}. \tag{4.14}$$

where $D_{p(1-p)}$ is a diagonal matrix composed of $p_i(1-p_i)$ on the diagonals where p_i is the probability of a nonzero outcome given subject i's covariates, X is a matrix containing group and covariate data. The proof of this result (not shown) follows the same delta method approach as is used in the appendices for the non-adjusting cases.

Using $DM(\beta, \tau, \sigma^2)$ from Equation 4.11 and Δ_{DM} and $\Sigma_{\beta,\tau,\sigma^2}$ the from Equations 4.13 and 4.14 a Wald-type test statistic can be defined as

$$D(\hat{\beta}, \hat{\tau}, \hat{\sigma}^2) = \frac{DM(\hat{\beta}, \hat{\tau}, \hat{\sigma}^2)}{\sqrt{n\Delta'_{\hat{DM}}\Sigma_{\hat{\beta},\hat{\tau},\hat{\sigma}^2}\Delta_{\hat{DM}}}} \tag{4.15}$$

which can then be compared to a standard normal distribution. This formulation can also be used to create confidence intervals.

4.4 Subject Specific Ratio of Means vs. Marginal Ratio of Means

When calculating ratios and expectations the order of operations matters. The ratio of an expectation does not equal the expectation of a ratio. Because of this, a ratio of means adjusting for covariates could be defined in one of two ways depending on the desired interpretation. One definition we will call the average subject specific ratio of means (RM_{SS}), the other we will name the marginal ratio of means (RM_{MAR}).

RM_{SS} is defined as the average of the individual subject ratios of means. From the two-part model, a ratio of means can be constructed for each subject which equals the expected value of the semicontinuous outcome given their observed covariate values and assuming they had been in group 1 divided by the expected value of the semicontinuous outcome given their observed covariate values and assuming they had been in group 0. In other words, a subject with the covariate values of subject i would have a RM_{SSi} times higher mean if they were in group 1 than if they were in group 0 (with all other covariates held constant). RM_{SS} is then the average of these subject specific mean ratios and can be written as follows:

$$RM_{SS} = E_Z\left(\frac{E(Y|X_1 = 1, Z)}{E(Y|X_1 = 0, Z)}\right) = \frac{1}{n}\sum_{1}^{n} RM_{SSi} \tag{4.16}$$

where n is the total number of subjects in the sample.

RM_{MAR} may be more applicable in addressing population based group differences. The marginal ratio of means is defined as follows:

$$RM_{MAR} = \frac{E_Z(E(Y|X_1 = 1, Z)}{E_Z(E(Y|X_1 = 1, Z))} = \frac{M_1}{M_0} \tag{4.17}$$

RM_{MAR} is constructed as a ratio of the marginalized group means M_j. The marginalized group means are, as defined in Section 4.2, the average of the semicontinuous outcome given group 0 or 1 averaged over the distribution of the adjusting covariates Z. One interpretation of RM_{MAR} is that if all subjects had been given treatment 1 the average outcome would be RM_{MAR} higher than if all of the subject had been given treatment 0.

4.4.1 Average Subject Specific Ratio of Means

In testing for the effect of group in terms of the average subject specific ratio of means, the null hypothesis we consider is that the average subject specific ratio of means (RM_{SS}) is equal to one. Unlike for DM and RM_{MAR}, this cannot be translated into an equal means comparison as the expectation over the covariates is taken after the individual ratio of means are calculated. As was the case for RM in the non-adjusted framework outlined in Chapter 3, this ratio also leads to simplification through cancelation of terms that are in both the numerator and denominator. Starting with the definition of RM_{SS} in Equation 4.17 for ZIG regression RM_{SS} is defined and simplified as follows:

$$\begin{aligned}
RM_{SS}(\beta,\tau) &= E_Z\left(\frac{E(Y|X_1=1,Z)}{E(Y|X_1=0,Z)}\right)\\
&= \frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_1+\beta_2 z_i}}{1+e^{\beta_0+\beta_1+\beta_2 z_i}} e^{\tau_0+\tau_1+\tau_2 z_i} \Big/ \frac{e^{\beta_0+\beta_2 z_i}}{1+e^{\beta_0+\beta_2 z_i}} e^{\tau_0+\tau_2 z_i}\\
&= e^{\beta_1+\tau_1}\frac{1}{n}\sum_i^n \frac{1+e^{\beta_0+\beta_2 z_i}}{1+e^{\beta_0+\beta_1+\beta_2 z_i}}\\
&= e^{\beta_1+\tau_1}\frac{1}{n}\sum_i^n \frac{1-p_{1i}}{1-p_{0i}},
\end{aligned} \tag{4.18}$$

where $p_{ji} = P(Y > 0|X_1 = j, Z_i)$. From Equation 4.18 we can see that after simplification, RM_{SS} is only dependent on the adjusting covariates through the probability portion of the model; any effect of covariate seen in the continuous part of the model does not effect RM_{SS}.

The null hypothesis of $RM_{SS} = 1$ mentioned above can be equivalently states as $\log(RM_{SS}(\beta,\tau)) = 0$. Using this result, we base our Wald-type $RM_{SS}(\beta,\tau)$ tests on $\log(RM_{SS}(\beta,\tau))$ where

$$\begin{aligned}
\log(RM_{SS}(\beta,\tau)) &= \beta_1+\tau_1+\log\left(\frac{1}{n}\sum_i^n \frac{1+e^{\beta_0+\beta_2 z_i}}{1+e^{\beta_0+\beta_1+\beta_2 z_i}}\right)\\
&= \beta_1+\tau_1+\log\left(\frac{1}{n}\sum_i^n \frac{1-p_{1i}}{1-p_{1i}}\right).
\end{aligned} \tag{4.19}$$

Starting with $\log(RM_{SS}(\beta,\tau))$ from Equation 4.19 we use the delta method to create a Wald-type test statistic for the null hypothesis that $log(RM_{SS}(\beta,\tau)) = 0$. (Note that this is not equivalent to $M_1 = M_0$.) To create this test, it can be shown that

$$\sqrt{n}\left(\log(RM_{SS}(\hat\beta,\hat\tau)) - log(RM_{SS}(\beta,\tau))\right) \xrightarrow{d} N(0, \Delta'_{\log(RM_{SS})}\Sigma_{\beta,\tau}\Delta_{\log(RM_{SS})}) \tag{4.20}$$

where $\Sigma_{\beta,\tau}$ is defined as in Equation 4.9, and $\Delta_{\log(RM_{SS})}$ if found by taking derivatives of $\log(RM_{SS} = \log(RM_{SS}(\beta,\tau))$ and simplifying, such that

$$\Delta_{\log(RM_{SS})} =$$

$$\left(\frac{\delta\log(RM_{SS})}{\delta\beta_0} \quad \frac{\delta\log(RM_{SS})}{\delta\beta_1} \quad \frac{\delta\log(RM_{SS})}{\delta\beta_2} \quad \frac{\delta\log(RM_{SS})}{\delta\tau_0} \quad \frac{\delta\log(RM_{SS})}{\delta\tau_1} \quad \frac{\delta\log(RM_{SS})}{\delta\tau_2}\right)'$$

$$= \begin{pmatrix} \frac{1}{n}\sum_i^n \left(e^{\beta_0+\beta_2 z_i}(1-p_{1i}) - \frac{p_{1i}}{p_{0i}}\right) \Big/ \frac{1}{n}\sum_i^n \frac{1-p_{1i}}{1-p_{0i}} \\ \frac{1}{n}\sum_i^n \frac{(1-p_{1i})^2}{1-p_{0i}} \Big/ \frac{1}{n}\sum_i^n \frac{1-p_{1i}}{1-p_{0i}} \\ \frac{1}{n}\sum_i^n \left(z_i e^{\beta_0+\beta_2 z_i}(1-p_{1i}) - \frac{p_{1i}}{p_{0i}}\right) \Big/ \frac{1}{n}\sum_i^n \frac{1-p_{1i}}{1-p_{0i}} \\ 0 \\ 1 \\ 0 \end{pmatrix}. \tag{4.21}$$

The proof of this result (not shown) follows the same delta method approach as is used in the appendices for the non-adjusting cases.

Finally, for finite sample sizes, a test for an effect of group in terms of $\log(RM_{SS}(\beta,\tau))$ can be constructed with the test statistic

$$S(\hat{\beta},\hat{\tau}) = \frac{\log(RM_{SS}(\hat{\beta},\hat{\tau}))}{\sqrt{n\Delta'_{\log(\hat{RM}_{SS})}\Sigma_{\hat{\beta},\hat{\tau}}\Delta_{\log(\hat{RM}_{SS})}}} \tag{4.22}$$

which will be asymptotically distributed as a standard normal.

When assuming a ZILN distribution with a constant coefficient of variation, the form for RM_{SS} and the null hypothesis tested are identical to those for when a ZIG distribution is assumed. For the ZILN regression, RM_{SS} is defined and can be simplified as follows:

$$\begin{aligned} RM_{SS}(\beta,\tau) &= E_Z\left(\frac{E(Y|X_1=1,Z)}{E(Y|X_1=0,Z)}\right) \\ &= \frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_1+\beta_2 z_i}}{1+e^{\beta_0+\beta_1+\beta_2 z_i}} e^{\tau_0+\tau_1+\tau_2 z_i+\frac{\sigma^2}{2}} \Big/ \frac{e^{\beta_0+\beta_2 z_i}}{1+e^{\beta_0+\beta_2 z_i}} e^{\tau_0+\tau_2 z_i+\frac{\sigma^2}{2}} \\ &= e^{\beta_1+\tau_1}\frac{1}{n}\sum_i^n \frac{1+e^{\beta_0+\beta_2 z_i}}{1+e^{\beta_0+\beta_1+\beta_2 z_i}} \\ &= e^{\beta_1+\tau_1}\frac{1}{n}\sum_i^n \frac{1-p_{1i}}{1-p_{0i}}. \end{aligned} \tag{4.23}$$

A Wald-type hypothesis test can be created by starting with $\log(RM_{SS}(\beta,\tau))$ which will be of the same final form as it was for ZIG (Equation 4.19) because of the canceling out of a homogeneous σ^2. Using delta method techniques as in previous sections, it can be shown that

$$\sqrt{n}\left(\log(RM_{SS}(\hat{\beta},\hat{\tau})) - \log(RM_{SS}(\beta,\tau))\right) \xrightarrow{d} N(0, \Delta'_{\log(RM_{SS})}\Sigma_{\beta,\tau,\sigma^2}\Delta_{\log(RM_{SS})}) \tag{4.24}$$

where $\Sigma_{\beta,\tau,\sigma^2}$ is defined as in Equation 4.14 and $\Delta_{\log(RM_{SS}(\beta,\tau))}$ is defined in Equation 4.25. Since σ^2 cancels out when creating RM_{SS}, its contribution to $\Delta_{\log(RM_{SS}(\beta,\tau))}$ will be the addition of an entry that equals zero. Consequently,

$$\begin{aligned}
&\Delta_{\log(RM_{SS}(\beta,\tau))} = \\
&\left(\frac{\delta\log(RM_{SS})}{\delta\beta_0} \quad \frac{\delta\log(RM_{SS})}{\delta\beta_1} \quad \frac{\delta\log(RM_{SS})}{\delta\beta_2} \quad \frac{\delta\log(RM_{SS})}{\delta\tau_0} \quad \frac{\delta\log(RM_{SS})}{\delta\tau_1} \quad \frac{\delta\log(RM_{SS})}{\delta\tau_2} \quad \frac{\delta\log(RM_{SS})}{\delta\sigma^2}\right)' \\
&= \begin{pmatrix}
\frac{1}{n}\sum_i^n \left(e^{\beta_0+\beta_2 z_i}(1-p_{1i}) - \frac{p_{1i}}{p_{0i}}\right) \Big/ \frac{1}{n}\sum_i^n \frac{1-p_{1i}}{1-p_{0i}} \\
\frac{1}{n}\sum_i^n \frac{(1-p_{1i})^2}{1-p_{0i}} \Big/ \frac{1}{n}\sum_i^n \frac{1-p_{1i}}{1-p_{0i}} \\
\frac{1}{n}\sum_i^n \left(z_i e^{\beta_0+\beta_2 z_i}(1-p_{1i}) - \frac{p_{1i}}{p_{0i}}\right) \Big/ \frac{1}{n}\sum_i^n \frac{1-p_{1i}}{1-p_{0i}} \\
0 \\
1 \\
0 \\
0
\end{pmatrix}.
\end{aligned} \tag{4.25}$$

The proof of this result (not shown) follows the same delta method approach as is used in the appendices for the non-adjusting cases. Using this result, for a finite sample a Wald-type test can be constructed for $RM_{SS} = 0$ which is

$$S(\hat{\beta},\hat{\tau},\hat{\sigma^2}) = \frac{\log(RM_{SS}(\hat{\beta},\hat{\tau}))}{\sqrt{n\Delta'_{\log(\hat{RM}_{SS})}\Sigma_{\hat{\beta},\hat{\tau},\hat{\sigma}^2}\Delta_{\log(\hat{RM}_{SS})}}} \tag{4.26}$$

which can compared to a standard normal distribution.

4.4.2 Marginal Ratio of Means

The ratio of the marginalized group means, RM_{MAR}, follows from the same hypothesis as DM, which is that the marginal mean of group 1 equals the marginal mean of group 0. The definition of RM_{MAR} utilizes the marginal mean definitions in Equations 4.3 for ZIG regression and 4.4 for ZILN regression to create a ratio of marginal means given the group assignments.

When a ZIG model is assumed, RM_{MAR} is defined as:

$$\begin{aligned}
RM_{MAR}(\beta,\tau) &= \frac{E_Z(E(Y|X_1=1,Z))}{E_Z(E(Y|X_1=1,Z))} \\
&= \frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_1+\beta_2 z_i}}{1+e^{\beta_0+\beta_1+\beta_2 z_i}} e^{\tau_0+\tau_1+\tau_2 z_i} \Bigg/ \frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_2 z_i}}{1+e^{\beta_0+\beta_2 z_i}} e^{\tau_0+\tau_2 z_i} \\
&= e^{\tau_0+\tau_1}\frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_1+\beta_2 z_i}}{1+e^{\beta_0+\beta_1+\beta_2 z_i}} e^{\tau_2 z_i} \Bigg/ e^{\tau_0}\frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_2 z_i}}{1+e^{\beta_0+\beta_2 z_i}} e^{\tau_2 z_i} \qquad (4.27) \\
&= e^{\tau_1}\frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_1+\beta_2 z_i}}{1+e^{\beta_0+\beta_1+\beta_2 z_i}} e^{\tau_2 z_i} \Bigg/ \frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_2 z_i}}{1+e^{\beta_0+\beta_2 z_i}} e^{\tau_2 z_i} \\
&= e^{\tau_1}\frac{\frac{1}{n}\sum_i^n p_{1i} e^{\tau_2 z_i}}{\frac{1}{n}\sum_i^n p_{0i} e^{\tau_2 z_i}}.
\end{aligned}$$

Unlike what was seen in defining RM_{SS}, for RM_{MAR}, the adjusting covariates do not cancel out with regards to either part of the model; both $\tau_2 z_i$ and $\beta_2 z_i$ are involved in calculating RM_{MAR}.

The null hypothesis of $M_1 = M_0$ or $RM_{MAR} = 1$ mentioned above can be equivalently states as $\log(RM_{MAR}) = 0$. In this section, we create Wald-type test for $\log(RM_{MAR})$ using the delta method where $\log(RM_{MAR})$ is defined as follows:

$$\begin{aligned}
\log(RM_{MAR}(\beta,\tau)) &= \tau_1 + \log(\sum_i^n \frac{e^{\beta_0+\beta_1+\beta_2 z_i}}{1+e^{\beta_0+\beta_1+\beta_2 z_i}} e^{\tau_2 z_i}) - \log(\sum_i^n \frac{e^{\beta_0+\beta_2 z_i}}{1+e^{\beta_0+\beta_2 z_i}} e^{\tau_2 z_i}) \\
&= \tau_1 + \log(\sum_i^n p_{1i} e^{\tau_2 z_i}) - \log(\sum_i^n p_{0i} e^{\tau_2 z_i}).
\end{aligned} \qquad (4.28)$$

Using the delta method, it can be shown that

$$\sqrt{n}\left(\log(RM_{MAR}(\hat{\beta},\hat{\tau})) - \log(RM_{MAR}(\beta,\tau))\right) \xrightarrow{d} N(0, \Delta'_{\log(RM_{MAR})}\Sigma_{\beta,\tau}\Delta_{\log(RM_{MAR})}) \tag{4.29}$$

where $\Sigma_{\beta,\tau}$ is defined as in Equation 4.9 and $\Delta_{\log(RM_{MAR})}$ is a matrix of derivatives of $\log(RM_{MAR}(\beta,\tau))$ in terms of β_0, β_1, τ_0 and, τ_1. Specifically,

$$\Delta_{\log(RM_{MAR}(\beta,\tau))} = \left(\frac{\delta\log(RM_{MAR})}{\delta\beta_0} \quad \frac{\delta\log(RM_{MAR})}{\delta\beta_1} \quad \frac{\delta\log(RM_{MAR})}{\delta\beta_2} \quad \frac{\delta\log(RM_{MAR})}{\delta\tau_0} \quad \frac{\delta\log(RM_{MAR})}{\delta\tau_1} \quad \frac{\delta\log(RM_{MAR})}{\delta\tau_2}\right)'$$
$$= \begin{pmatrix} \frac{\frac{1}{n}\sum_i^n(p_{1i}(1-p_{1i})\mu_{1i})}{\frac{1}{n}\sum_i^n(p_{1i}\mu_{1i})} - \frac{\frac{1}{n}\sum_i^n(p_{0i}(1-p_{0i})\mu_{0i})}{\frac{1}{n}\sum_i^n(p_{0i}\mu_{0i})} \\ \frac{\frac{1}{n}\sum_i^n(p_{1i}(1-p_{1i})\mu_{1i})}{\frac{1}{n}\sum_i^n(p_{1i}\mu_{1i})} \\ \frac{\frac{1}{n}\sum_i^n(z_ip_{1i}(1-p_{1i})\mu_{1i})}{\frac{1}{n}\sum_i^n(p_{1i}\mu_{1i})} - \frac{\frac{1}{n}\sum_i^n(z_ip_{0i}(1-p_{0i})\mu_{0i})}{\frac{1}{n}\sum_i^n(p_{0i}\mu_{0i})} \\ 0 \\ 1 \\ \frac{\frac{1}{n}\sum_i^n(z_ip_{1i}\mu_{1i})}{\frac{1}{n}\sum_i^n(p_{1i}\mu_{1i})} - \frac{\frac{1}{n}\sum_i^n(z_ip_{0i}\mu_{0i})}{\frac{1}{n}\sum_i^n(p_{0i}\mu_{0i})} \end{pmatrix}. \tag{4.30}$$

The proof of this result (not shown) follows the same delta method approach as that used in the appendices. Using this result for finite sample sizes a test can be defined for the effect of group in terms of RM_{MAR} as

$$M = \frac{\log(R\hat{M_{MAR}}(\beta,\tau))}{\sqrt{n\Delta'_{\log(R\hat{M}_{MAR})}\Sigma_{\hat{\beta},\hat{\tau}}\Delta_{\log(R\hat{M}_{MAR})}}} \tag{4.31}$$

which can be compared to a standard normal distribution.

Similarly, when assuming ZILN regression, RM_{MAR} can be defined as:

$$
\begin{aligned}
RM_{MAR}(\beta,\tau) &= \frac{E_Z(E(Y|X_1=1,Z))}{E_Z(E(Y|X_1=1,Z))} \\
&= \frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_1+\beta_2 z_i}}{1+e^{\beta_0+\beta_1+\beta_2 z_i}} e^{\tau_0+\tau_1+\tau_2 z_i+\frac{\sigma^2}{2}} \Bigg/ \frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_2 z_i}}{1+e^{\beta_0+\beta_2 z_i}} e^{\tau_0+\tau_2 z_i+\frac{\sigma^2}{2}} \\
&= e^{\tau_0+\tau_1+\frac{\sigma^2}{2}} \frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_1+\beta_2 z_i}}{1+e^{\beta_0+\beta_1+\beta_2 z_i}} e^{\tau_2 z_i} \Bigg/ e^{\tau_0+\frac{\sigma^2}{2}} \frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_2 z_i}}{1+e^{\beta_0+\beta_2 z_i}} e^{\tau_2 z_i} \\
&= e^{\tau_1} \frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_1+\beta_2 z_i}}{1+e^{\beta_0+\beta_1+\beta_2 z_i}} e^{\tau_2 z_i} \Bigg/ \frac{1}{n}\sum_i^n \frac{e^{\beta_0+\beta_2 z_i}}{1+e^{\beta_0+\beta_2 z_i}} e^{\tau_2 z_i} \\
&= e^{\tau_1} \frac{\frac{1}{n}\sum_i^n p_{1i} e^{\tau_2 z_i}}{\frac{1}{n}\sum_i^n p_{0i} e^{\tau_2 z_i}}.
\end{aligned}
\tag{4.32}
$$

In Equation 4.32 it is seen that when σ^2 is assumed to be homogeneous across group and adjusting covariate, it can be factored out of the expectations and then canceled out of the ratio leading to a similar form for $RM_{MAR}(\beta,\tau)$ regardless of distribution choice (compare Equations 4.27 and 4.32). To create a Wald test based on the hypothesis that $RM_{MAR}=1$, we will test the equivalent $\log(RM_{MAR})=0$. Since RM_{MAR} is of the same form for both ZIG and ZILN analyses, the form for $\log(RM_{MAR})$ is the same as that seen in Equation 4.28.

Using delta method techniques, the asymptotic distribution of $\log(RM_{MAR}(\hat{\beta},\hat{\tau}))$ can be found as

$$
\sqrt{n}\left(\log(RM_{MAR}(\hat{\beta},\hat{\tau})) - \log(RM_{MAR}(\beta,\tau))\right) \xrightarrow{d} N(0, \Delta'_{\log(RM_{MAR})}\Sigma_{\beta,\tau,\sigma^2}\Delta_{\log(RM_{MAR})})
\tag{4.33}
$$

where $\Sigma_{\beta,\tau,\sigma^2}$ is defined as in Equation 4.14 and where the matrix of derivatives

$\Delta_{\log(RM_{MAR})}$ given a ZILN distribution is:

$$\Delta_{\log(RM_{MAR}} =$$

$$\begin{pmatrix} \frac{\delta \log(RM_{MAR})}{\delta\beta_0} & \frac{\delta \log(RM_{MAR})}{\delta\beta_1} & \frac{\log(RM_{MAR})}{\delta\beta_2} & \frac{\delta \log(RM_{MAR})}{\delta\tau_0} & \frac{\delta \log(RM_{MAR})}{\delta\tau_1} & \frac{\delta \log(RM_{MAR})}{\delta\tau_2} \end{pmatrix}'$$

$$= \begin{pmatrix} \frac{\frac{1}{n}\sum_i^n (p_{1i}(1-p_{1i})\mu_{1i})}{\frac{1}{n}\sum_i^n (p_{1i}\mu_{1i})} - \frac{\frac{1}{n}\sum_i^n (p_{0i}(1-p_{0i})\mu_{0i})}{\frac{1}{n}\sum_i^n (p_{0i}\mu_{0i})} \\ \frac{\frac{1}{n}\sum_i^n (p_{1i}(1-p_{1i})\mu_{1i})}{\frac{1}{n}\sum_i^n (p_{1i}\mu_{1i})} \\ \frac{\frac{1}{n}\sum_i^n (z_i p_{1i}(1-p_{1i})\mu_{1i})}{\frac{1}{n}\sum_i^n (p_{1i}\mu_{1i})} - \frac{\frac{1}{n}\sum_i^n (z_i p_{0i}(1-p_{0i})\mu_{0i})}{\frac{1}{n}\sum_i^n (p_{0i}\mu_{0i})} \\ 0 \\ 1 \\ \frac{\frac{1}{n}\sum_i^n (z_i p_{1i}\mu_{1i})}{\frac{1}{n}\sum_i^n (p_{1i}\mu_{1i})} - \frac{\frac{1}{n}\sum_i^n (z_i p_{0i}\mu_{0i})}{\frac{1}{n}\sum_i^n (p_{0i}\mu_{0i})} \\ 0 \end{pmatrix}. \tag{4.34}$$

The proof of this result (not shown) follows the same delta method approach as that used in the appendices for the non-adjusting cases. From this results a Wald-type test for $log(RM_{MAR}) = 0$ can be created such that

$$M = \frac{\log(RM_{MAR}(\hat{\beta}, \hat{\tau}))}{\sqrt{n\Delta'_{\log(R\hat{M}_{MAR})}\Sigma_{\hat{\beta},\hat{\tau},\hat{\sigma}^2}\Delta_{\log(R\hat{M}_{MAR})}}} \tag{4.35}$$

which can be compared to a standard normal distribution.

4.5 Simulation Methods for Comparing Two Groups When Adjusting for a Dichotomous Covariate

The simulation study outlined in this section explores Type 1 error and power rates for tests based on DM, RM_{SS}, and RM_{MAR} mean-based comparisons between two groups while adjusting for a dichotomous covariate. Data were simulated from both ZIG and ZILN regression frameworks with various sample sizes, parameter values, and group by gender combinations. For subject i, let x_i be the group identifier and z_i be the identifier for the adjusting covariate (e.g. gender). Also, let β_0, β_1,

Correlation between group and gender	Group 0		Group 1	
	Female	Male	Female	Male
0	50	50	50	50
0.2	60	40	40	60
0.4	70	30	30	70
0.6	80	20	20	80

Table 4.1: Correlation and number of subjects in each group by gender cohort

and β_2 be the true effects in the binomial part of the model for intercept, group, and gender respectively. Finally, let τ_0, τ_1, and τ_2 be the true effects in the continuous part of the model for intercept, group, and gender with ν as the dispersion parameter for gamma and σ^2 as the dispersion parameter for log-normal. Simulating from ZIG or ZILN distributions requires a two-step process. First, for subject i we simulate Y_i^* from a binomial distribution such that $\text{logit}(P(Y_i^* = 1|X_{1i} = x_{1i})) = \beta_0+\beta_1 x_{1i}+\beta_2 z_i$. Then, if $Y_i^* = 0$, we set $Y_i = 0$; otherwise we simulate Y_i from a gamma distribution with $Y_i \sim Gamma(e^{\tau_0+\tau_1 x_{1i}+\tau_2 z_i}, \nu)$ for ZIG regression or from a log-normal distribution with $\log(Y_i) \sim N(\tau_0+\tau_1 x_1 i+\tau_2 z_i, \sigma^2)$ for ZILN regression. For our simulations we utilized four design matrices which allowed for different correlations between group and gender. The resulting sample size per group by gender combination for these design matrices are outlined in Table 4.1 for an overall sample size of 200.

At each setting, 10,000 data sets were simulated. Overall sample sizes of n=100 and 200 were used, however, we will primarily report results where n=200. All models were estimated via the Newton-Raphson technique outlined in Section 2.4. Data sets for which the maximization technique did not converge were excluded from the Type 1 error and power calculations. Settings where less than 99% of data sets had estimable parameters are excluded from this report.

Tests based on DM, RM_{SS}, and RM_{MAR} were calculated as described earlier in this chapter. In addition to these mean-based tests, simple linear regression and rank transformed regression were used as comparison tests. Rank transformed regression was performed by simple linear regression on the ranks of the outcome variable.

4.5.1 Simulation Settings

Four groups of settings were used to explore Type 1 error and power for DM, RM_{SS}, and RM_{MAR} tests. These are outlined in Tables 4.2 - 4.5 which display for each group by gender cohort ($X_1 = j$ by $Z = k$) the probability of a nonzero outcome p_{jk}, the mean of the nonzero outcomes μ_{jk}, and the resultant overall group by gender mean M_{jk}.

Table 4.2 outlines settings designed to examine Type 1 error rates when the two group comparison is for groups that have exactly the same settings. For some of these settings, there are gender effects but for all of the settings there are no group effects. These settings are designed to explore Type 1 error for DM, RM_{SS}, and RM_{MAR} in terms of group, allowing for different effects in terms of gender. We will call these 'general null settings' because within each setting the two groups are simulated from the same distribution with the same means, probabilities, and dispersion parameter values. Since there are no group differences $RM_{SS} = RM_{MAR} = 1$ and $DM = 0$ for these settings. Setting 'B' has no gender or group effects; settings 'G', 'H', and 'I' have no group effects, but have gender effects designed such that for the gender variable $RM_{SS} = RM_{MAR} = 1.44$ and $DM = 0.33$ with the gender effect coming entirely from the binomial portion of the model for setting 'G', coming entirely from the continuous part of the model for setting 'I', and coming from both parts of the model equally for setting 'H'. The effect of this can be seen by noticing that the probabilities, conditional means, and overall means differ across genders (indexed by k) but not across groups (indexed by j), e.g, $M_{00} = M_{10}$ and $M_{01} = M_{11}$.

In contrast to the general null settings, 'metric-based dissonant null settings'

	p_{00}	μ_{00}	M_{00}	p_{01}	μ_{01}	M_{01}	p_{10}	μ_{10}	M_{10}	p_{11}	μ_{11}	M_{11}
B	0.5	1.5	0.75	0.5	1.5	0.75	0.5	1.5	0.75	0.5	1.5	0.75
G	0.5	1.5	0.75	0.72	1.5	1.08	0.5	1.5	0.75	0.72	1.5	1.08
H	0.5	1.5	0.75	0.6	1.8	1.08	0.5	1.5	0.75	0.6	1.8	1.08
I	0.5	1.5	0.75	0.5	2.16	0.75	0.5	1.5	0.75	0.5	2.16	1.08

Table 4.2: General null regression settings for a two group comparison adjusting for a dichotomous covariate.

[1] Given $X_1 = j$ and $Z = k$, p_{jk} is the probability of a nonzero outcome, μ_{jk} is the mean of the nonzero outcomes, and M_{jk} is the overall mean.

[2] Setting 'B' has no Z effects, 'G' has a Z effect only in the binomial portion of the model, 'H' has Z effects in both parts of the model, and 'I' has an Z effect only in the continuous part of the model.

have nonzero group effects in both parts of the model which cancel out in terms of the metric of interest. Because RM_{SS} is an average of subject specific ratios but RM_{MAR} and DM are based on the marginalized group means, the dissonant null settings for which $RM_{SS} = 1$ will differ from those where $RM_{MAR} = 1$ and $DM = 0$. Table 4.3 details the settings at the metric-based null where $RM_{SS} = 1$. For the settings presented here, the effects of gender (β_2 and τ_2) are set to be equal to the effects which correspond to group (β_1 and τ_1). Other settings explored included smaller effects of the adjusting covariate gender relative to the size of the effects of group. The settings outlined in Table 4.3 were determined by first setting τ_1 and τ_2 effects equal to the τ_1 setting used for the dissonant effects in Table 3.2 from Section 3.5.1. Then using these values and the planned X matrices (those outlined in Table 4.1), $\beta_1 = \beta_2$ were found such that $RM_{SS} = 1$. This occurs when

$$e^{\tau_1} = \frac{1}{n}\sum_{1}^{n} \frac{p_{0i}}{p_{1i}} \tag{4.36}$$

where $p_{0i} = e^{\beta_0+\beta_2 z_i}/(1+e^{\beta_0+\beta_2 z_i})$ and $p_{1i} = e^{\beta_0+\beta_1+\beta_2 z_i}/(1+e^{\beta_0+\beta_1+\beta_2 z_i})$ for $i = 1, 2, \ldots,$ n. For the X matrices defined in Table 4.1 we have equal numbers of male and females and equal numbers of cases and controls. Because of this, the parameter settings determined will apply to all of the X matrices proposed.

	p_{00}	μ_{00}	M_{00}	p_{01}, p_{10}	μ_{01}, μ_{10}	M_{01}, M_{10}	p_{11}	μ_{11}	M_{11}
b	0.75	1	0.75	0.5487	1.5	0.8231	0.3302	2.25	0.7429
c	0.75	1	0.75	0.3300	3	0.9900	0.0748	9	0.6732
d	0.5	1.5	0.75	0.2733	3	0.8199	0.1239	6	0.7435

Table 4.3: Metric-based dissonant null settings for $RM_{SS} = 1$ adjusting for a dichotomous covariate.

[1] Given $X_1 = j$ and $Z = k$, p_{jk} is the probability of a nonzero outcome, μ_{jk} is the mean of the nonzero outcomes, and M_{jk} is the overall mean.

	p_{00}	μ_{00}	M_{00}	p_{01}, p_{10}	μ_{01}, μ_{10}	M_{01}, M_{10}	p_{11}	μ_{11}	M_{11}
b	0.75	1	0.75	0.5505	1.5	0.8258	0.3333	2.25	0.75
c	0.75	1	0.75	0.3431	3	1.0292	0.0833	9	0.75
d	0.5	1.5	0.75	0.2743	3	0.8229	0.1250	6	0.75

Table 4.4: Null regression settings based on equality of marginal means ($RM_{MAR} = 1$, $DM = 0$) adjusting for a dichotomous covariate.

[1] Given $X_1 = j$ and $Z = k$, p_{jk} is the probability of a nonzero outcome, μ_{jk} is the mean of the nonzero outcomes, and M_{jk} is the overall mean.

The group of settings outlined in Table 4.4 are used to explore Type 1 error rates with dissonant effects under a metric-based null where the marginal mean for group 0 equals the marginal mean for group 1 leading to $RM_{MAR} = 1$ and $DM = 0$. In a similar method to that described above for RM_{SS}, the parameter settings from

Section 3.5.1 were used as a starting point with new β_1 values determined such that $RM_{MAR} = 1$.

Table 4.5 and Table 4.6 outline settings used to explore power. For Table 4.5, settings were determined such that $RM_{SS} = 1.44$ for both the group effect and the gender effect. Table 4.6 outlines settings with weaker gender effects. These settings were found such that $RM_{SS} = 1.44$ for group, but the effects of gender were smaller than those for group, specifically they were chosen such that $\beta_2 = \sqrt{\beta_1}$. Most of the results outlined in Section 4.8.3 refer to Table 4.6 settings which include the weaker effects for the adjusting covariate, gender. These strong effect settings were also examined using all of the metrics, but will only be displayed in the results when the power results differ substantially from those of the stronger adjusted effect settings.

	p_{00}	μ_{00}	M_{00}	p_{01}, p_{10}	μ_{01}, μ_{10}	M_{01}, M_{10}	p_{11}	μ_{11}	M_{11}
G^2	0.5	1.5	0.75	0.8839	1.5	1.3259	0.9831	1.5	1.4746
H	0.5	1.5	0.75	0.6156	1.8	1.1081	0.7195	2.16	1.5541
I	0.5	1.5	0.75	0.5	2.16	1.08	0.5	3.1104	1.5552

Table 4.5: Power settings for two-group comparisons adjusting for a dichotomous covariate.

[1] Given $X_1 = j$ and $Z = k$, p_{jk} is the probability of a nonzero outcome, μ_{jk} is the mean of the nonzero outcomes, and M_{jk} is the overall mean.

[2] When the correlation between group and gender is 0.4 or 0.6 less than 99% of the data sets had estimable parameters.

	p_{01}	μ_{01}	M_{01}	p_{10}	μ_{10}	M_{10}	p_{11}	μ_{11}	M_{11}
G	0.6497	1.5	0.9745	0.7747	1.5	1.1621	0.8644	1.5	1.2966
H	0.5540	1.6431	0.9104	0.6069	1.8	1.0923	0.6573	1.9717	1.2959
I	0.5	1.8	0.9	0.5	2.16	1.08	0.5	2.5920	1.2960

Table 4.6: Power settings for two-group comparison adjusting for a dichotomous covariate with smaller adjusting effect sizes.

[1] Given $X_1 = j$ and $Z = k$, p_{jk} is the probability of a nonzero outcome, μ_{jk} is the mean of the nonzero outcomes, and M_{jk} is the overall mean. For all settings shown here, $p_{01} = 0.5$, $\mu_{01} = 1.5$, and $M_{01} = 0.75$.

4.6 Simulation Results for Comparing Two Groups When Adjusting for a Dichotomous Covariate

4.6.1 Type 1 Error for Consonant Tests Under General Null Setting

Figures 4.1 - 4.5 show the Type 1 error rates for the settings described in Table 4.2. These settings have no differences across groups, but most have effects in the adjusting covariate gender. Various levels of correlation between the group variable and the adjusting covariate were included (see Table 4.1). Within each plot included in Figures 4.1 - 4.5, the correlation between group and the adjusting covariate are on the x-axis; the observed Type 1 error rate is on the y-axis. Each figure contains a set of four plots. The plots on the left contain settings where the data was simulated from a ZIG distribution and the plots on the right are for settings where the data were simulated from a ZILN distribution. The top plots in each set have coefficient of variation of $\nu^{-1} = 0.5$ and the bottom plots in each set have a coefficient of variation of $\nu^{-1} = 2$. Figures 4.1 - 4.2 include all of the settings outlined in Table 4.2 with 'B' representing no effect of the adjusting covariate, 'G' the adjusting covariate having an effect in the binomial part, 'I' in the continuous part only, and 'H' in both parts. A solid black line shows the nominal level of 0.05 that was used for these analyses with dotted lines at 0.0457 and 0.0543 reflecting $\pm$ 1.96 standard errors of a binomial distributions when p=0.5 and 10,000 different data sets are run. Thus, any observed Type 1 error rates that are outside these bounds are different than would be expected if the Type 1 error were truly equal to the nominal 0.5 level.

Figure 4.1 shows the Type 1 error results for the DM Wald test. When the data are simulated from a ZIG distribution and $\nu^{-1} = 0.5$, all DM Type 1 error rates are within the appropriate range or only slightly conservative. However, when $\nu^{-1} = 2$, the Type 1 error rates are extremely low. This is especially true for data incorrectly analyzed as ZILN. Results not pictured found that this extreme conservationism is somewhat mitigated as sample size increases, so, part of this extreme conservativeness may be due to slow convergence of the DM Wald test. For the ZIG data, the effect of the adjusting covariate (seen in the different setting labels) does not impact the Type 1 error rates for the effect of group. The graphs on the right in Figure 4.1 show that when ZILN data are analyzed as ZILN, Type 1 error rates are within the appropriate range to slightly conservative. On the other hand, when ZILN data are analyzed as ZIG, the Type 1 error rates are elevated. This is especially true when ν^{-1} is high. Simulations not included show that, as was seen in the univariate case, Type 1 error rates become increasingly elevated as sample size increases. Type 1 error rates for DM when ZILN data are analyzed as ZIG differ by adjusting covariate setting. When there is an effect of the adjusting covariate in terms of the probability of a nonzero outcome ('G' and 'H') Type 1 error rates are higher than when there is not ('B' and 'I'). One potential reason for this is that the effects used are positive, leading to a higher probability of nonzero outcomes and therefore smaller levels of zero-inflation and higher expected sample sizes. Figure 4.1 also shows that Type 1 error decreases as correlation increases when $\nu^{-1} = 2$, especially when ZIG analyses are used.

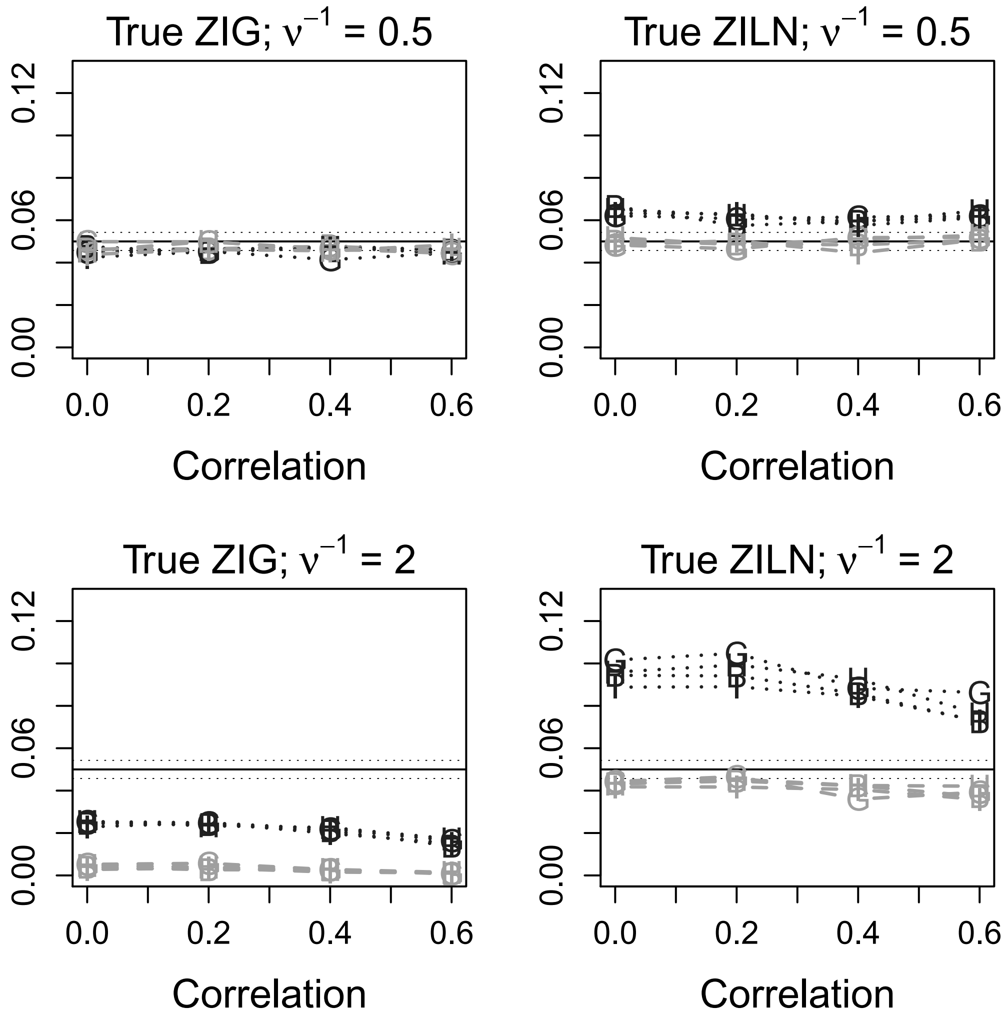

Figure 4.1: Type 1 error for tests based on DM; two group comparison adjusting for a dichotomous covariate. Symbols correspond to the settings labels in Table 4.2. The settings have no group effects, but for some settings there are significant adjusting (gender) effects. Setting 'B' has no gender effect, 'G' has a gender effect only in the binomial portion of the model, 'H' has gender effects in both parts of the model, and 'I' has a gender effect only in the continuous part of the model. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

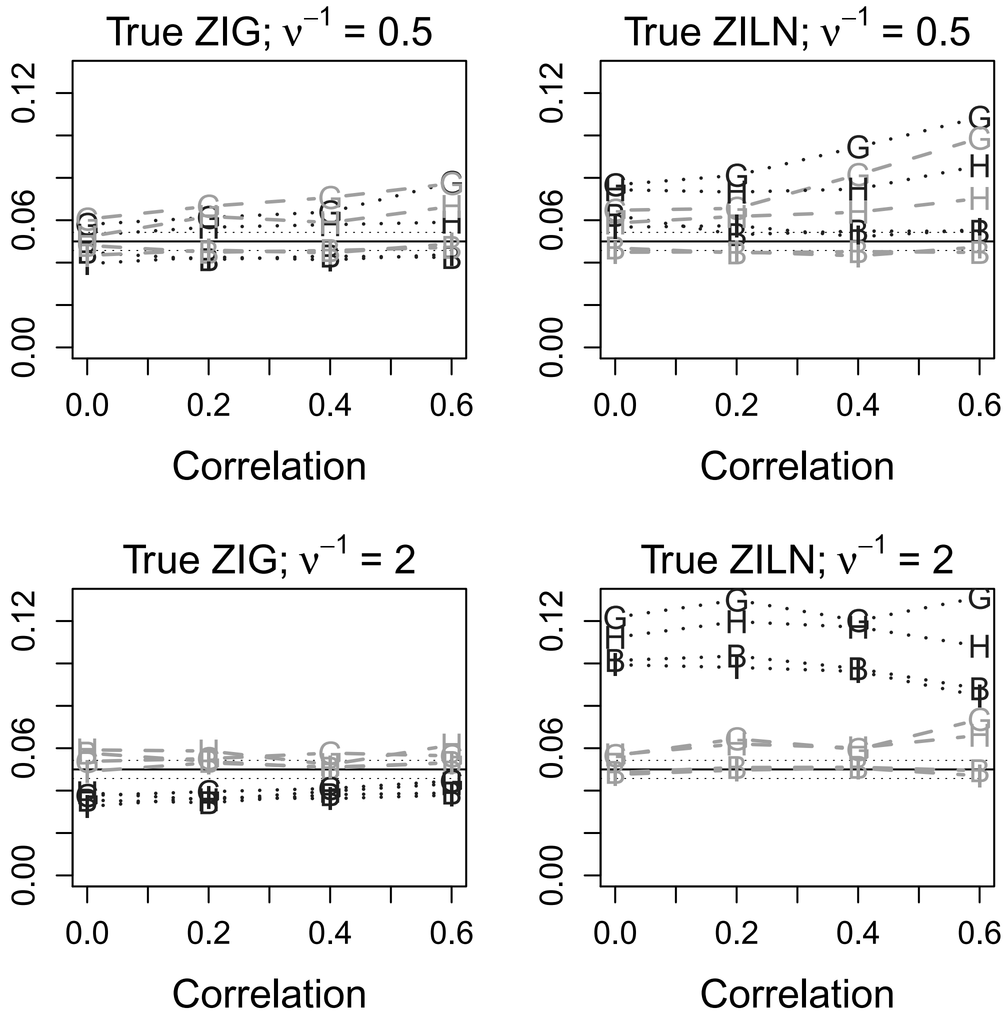

Figure 4.2: Type 1 error for tests based on RM_{SS}; two group comparison adjusting for a dichotomous covariate. Symbols correspond to the settings labels in Table 4.2. The settings have no group effects, but for some settings there are significant adjusting (gender) effects. Setting 'B' has no gender effect, 'G' has a gender effect only in the binomial portion of the model, 'H' has gender effects in both parts of the model, and 'I' has a gender effect only in the continuous part of the model. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

Type 1 error rates for RM_{SS} are seen in Figure 4.2. For ZIG data with $\nu^{-1} = 0.5$, Type 1 error is higher than nominal when the adjusting covariate has a nonzero effect in the binomial part of the model ('G' and 'H'), but is slightly lower than nominal when the adjusting effect is either equal to zero in both parts of the model ('B') or is nonzero in the continuous part only ('I'). Also, Type 1 error rates are slightly higher for incorrectly specified ZILN analyses than they are for correctly specified ZIG analyses. For ZIG data with $\nu^{-1} = 2$, Type 1 error rates are close to the appropriate range, but are slightly elevated for ZILN analyses and slightly conservative for ZIG analyses.

The graphs on the right in Figure 4.2 show the Type 1 error rates for RM_{SS} when data are simulated from a ZILN distribution. When $\nu^{-1} = 0.5$, Type 1 error rates were elevated for settings with a positive effect in the binomial part of the model ('G' and 'H') and within the appropriate range for those with a binomial effect equal to zero ('B' and 'I'). This Type 1 error elevation is slightly higher when the data are analyzed as ZIG, and increases with correlation between group and the adjusting covariate. For ZILN data with $\nu^{-1} = 2$, there is a much greater difference in Type 1 error between ZIG and ZILN analyses. ZIG analyses yield Type 1 error rates that are more than double the nominal 0.05. For correctly specified ZILN analyses with $\nu^{-1} = 2$, Type 1 error rates are within the appropriate range when there are no nonzero gender effects in the binomial part of the model ('B' and 'I'), and slightly elevated Type 1 error rates when there is a nonzero adjusting effect in the binomial part of the model ('G' and 'H'). Type 1 error increases with the strength of the adjusting effect in the binomial part of the model. These differences are more pronounced for RM_{SS} than they were for DM.

Figure 4.3 shows the Type 1 error results for RM_{MAR} based tests. With the exception of ZIG data when $\nu^{-1} = 2$ (bottom left graph), these results are very similar to the DM results. These similarities are likely due to the fact that both are

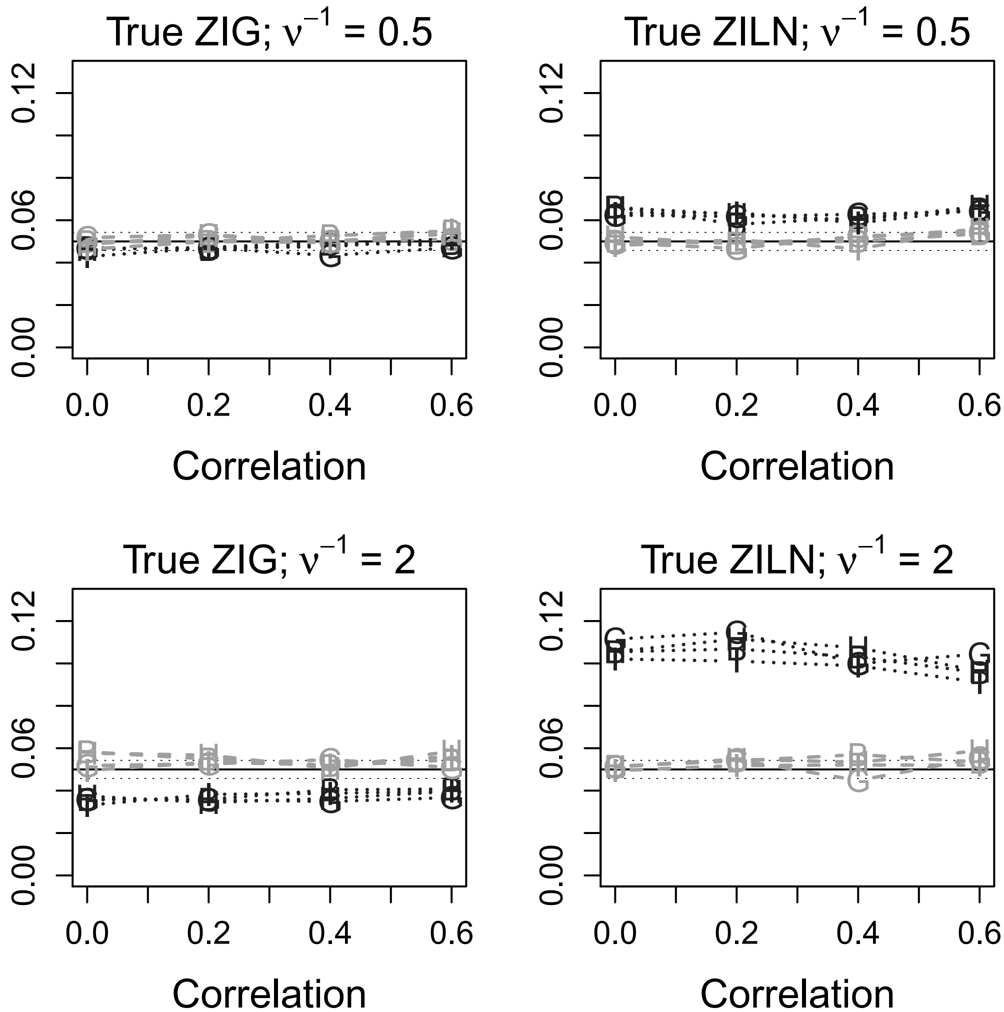

Figure 4.3: Type 1 error for tests based on RM_{MAR}; two group comparison adjusting for a dichotomous covariate. Symbols correspond to the settings labels in Table 4.2. The settings have no group effects, but for some settings there are significant adjusting (gender) effects. Setting ‘B’ has no gender effect, ‘G’ has a gender effect only in the binomial portion of the model, ‘H’ has gender effects in both parts of the model, and ‘I’ has a gender effect only in the continuous part of the model. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

based upon the marginalized group means. When the data are ZIG and $\nu^{-1} = 0.5$, Type 1 error rates are within the appropriate range. Type 1 error rates are also appropriate when ZILN data are analyzed as ZILN. Type 1 error rates for ZILN data analyzed as ZIG are slightly high when $\nu^{-1} = 0.5$ and quite a bit higher when $\nu^{-1} = 2$ with only slight differences between the various settings for the adjusting covariate. On the other hand, when the data are from a ZIG distribution and $\nu^{-1} = 0.5$, the excessive conservativeness seen in the DM analyses is no longer present. Instead, the ZILN analyses are in the appropriate range and the ZIG analyses are only slightly conservative.

Figures 4.4 and 4.5 compare all of the mean-based tests to each other and to simple linear regression and rank-based regression test [21] at one adjusted covariate setting each. Figure 4.4 reports the simulations where there is an effect only in the binomial part (setting 'G'), and Figure 4.5 does the same when there is a nonzero effect only in the continuous part (setting 'I'). Settings 'B' and 'H' are not pictured. In setting 'B' there are no adjusting effects and the observed results are nearly identical to those pictured for 'I'. In setting 'H' there are adjusting effects in both parts of the two-part models and the differences between tests are similar to those seen for setting 'G', albeit somewhat muted. For these two figures, 'S' represents Type 1 error for the RM_{SS} based tests, 'M' represents Type 1 error for the RM_{MAR} based tests, 'D' for DM based tests, 'L' for simple linear regression, and 'R' for rank transformed regression. For all figures, light gray with dashes refer ZILN based analyses, dark grey with dots refers to ZIG based analyses and black with an alternating dot-dash pattern is used to show the simple linear regression results.

When the adjusting covariate has an effect in the binomial part of the model (pictured in Figure 4.4) or in both parts (not pictured), the RM_{SS} Wald test has the highest Type 1 error rates followed by RM_{MAR} and DM. As noted in Figure 4.1 - 4.3, when the data are ZILN but analyzed as ZIG, these Type 1 error rates are excessively

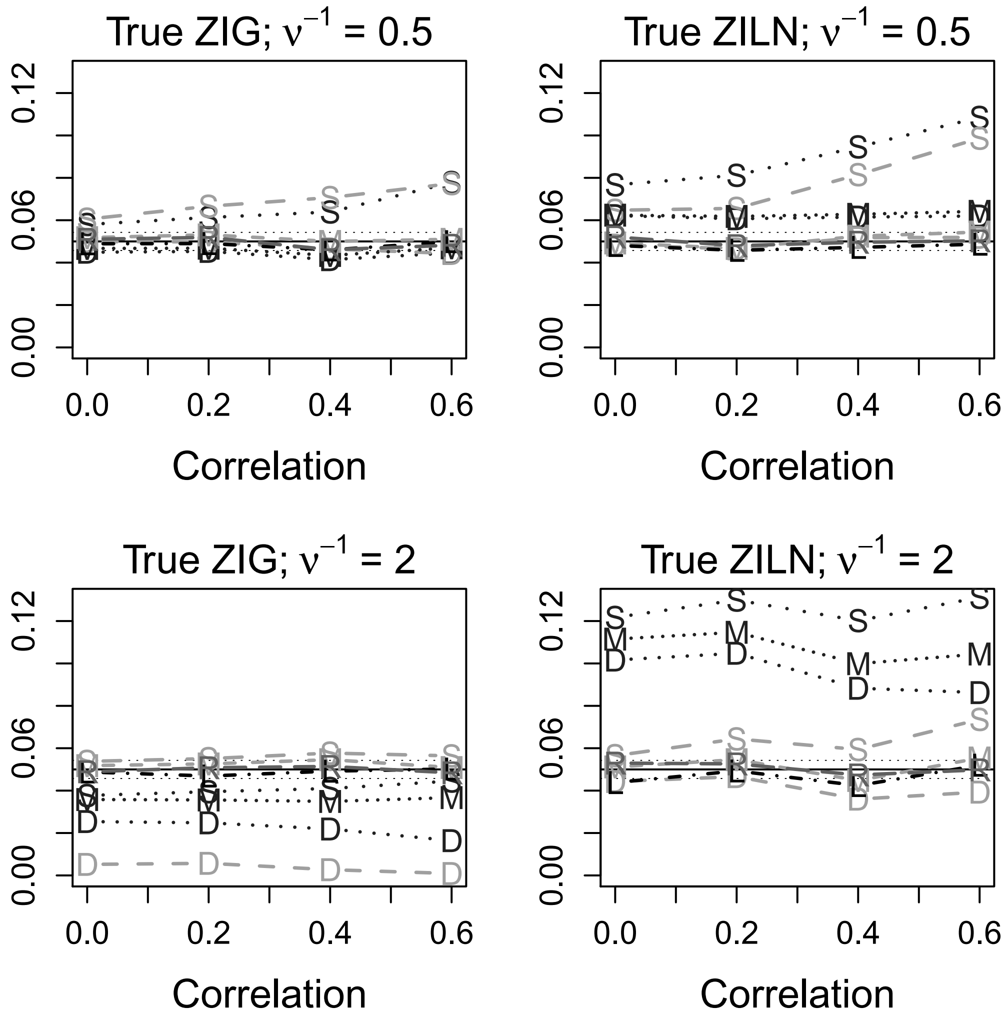

Figure 4.4: Comparison of Type 1 error for mean-based tests, SLR, and RTR; adjusting effect in the binomial part; two group comparison adjusting for a dichotomous covariate. The symbols in the figure correspond to the test being used. 'D' represents the Type 1 error rates for tests based on DM, 'S' those based on RM_{SS}, and 'M' those based on RM_{MAR} with darker grey symbols with dotted lines for ZIG analyses and lighter grey symbols with dashed lines for ZILN analyses. SLR analysis results are shown with the symbol 'L' and black dash-dot lines; RTR with the symbol 'R' and medium grey lines with long dashes.

high. In Figure 4.4 it can be seen that DM has the most excessively conservative Type 1 error rates when data simulated as ZIG are analyzed as ZILN. Simple linear regression and rank transformed regression both typically have appropriate Type 1 error rates for this setting, except that for some correlations when the data are simulated from a ZILN distribution and $\nu^{-1} = 2$, where simple linear regression had slightly low Type 1 errors.

Figure 4.5 compares the Type 1 error rates for the mean-based tests, simple linear regression, and rank transformed regression when the only nonzero effect of the adjusting covariate is in the continuous part of the model. When the true adjusting effect is zero in both parts of the model, similar Type 1 error rates are found. For these settings, RM_{SS} and RM_{MAR} have nearly identical Type 1 error rates and DM has lower Type 1 error rates than both of them. This is likely due to the fact that the DM estimate includes the estimate of σ^2 which greatly increases the $Var(DM)$ estimate whereas in RM_{MAR} and RM_{SS}, σ^2 cancels out and does not effect $Var(\log(RM_{MAR}))$ or $Var(\log(RM_{SS}))$. When the coefficient of correlation is high (e.g. $\nu^{-1} = 2$), simulations from a gamma distribution have some very small values which do not fit well in log-normal distributions. This results in poor estimates of σ^2 and larger $Var(DM)$ which leads to poorer Type 1 error rates. Simple linear regression and rank transformed regression typically exhibit appropriate Type 1 error rates for the general null settings examined in this section with only slightly conservative Type 1 error rates for some settings where the data are ZILN and $\nu^{-1} = 2$.

4.6.2 Type 1 Error for Metric Based Null Settings

In the previous section, the same settings were able to lead to $RM_{SS} = 1$, $RM_{MAR} = 1$, and $DM = 0$, because outcome values for the two groups were drawn from the same distribution. These setting were also at the assumed null for simple linear regression and rank transformed regression since there were no group effects.

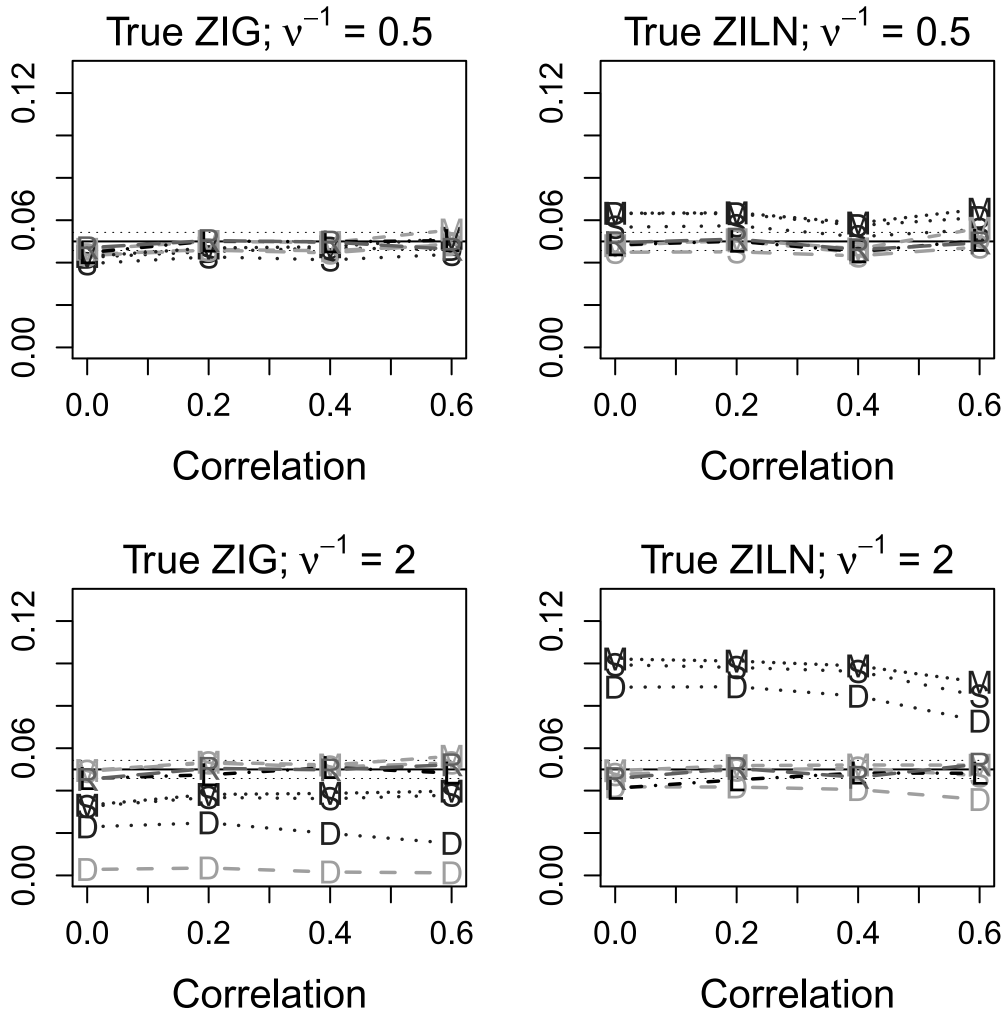

Figure 4.5: Comparison of Type 1 error for mean-based tests, SLR, and RTR; adjusting effect in the binomial part; two group comparison adjusting for a dichotomous covariate. The symbols in the figure correspond to the test being used. 'D' represents the Type 1 error rates for tests based on DM, 'S' those based on RM_{SS}, and 'M' those based on RM_{MAR} with darker grey symbols with dotted lines for ZIG analyses and lighter grey symbols with dashed lines for ZILN analyses. SLR analysis results are shown with the symbol 'L' and black dash-dot lines; RTR with the symbol 'R' and medium grey lines with long dashes.

In contrast this section explores settings where the two groups differ in terms of probability of a nonzero outcome and the conditional mean given a nonzero outcome, but where these effects cancel out in terms of the metric of interest. We call these settings 'metric-based null' settings and they are outlined for $RM_{SS} = 1$ in Table 4.3 and for $RM_{MAR} = 1$ and $DM = 0$ in Table 4.4.

Figures 4.6 and 4.8 show the Type 1 error rates for DM and RM_{MAR} using the marginal mean-based null settings 'b', 'c', and 'd' outlined in Table 4.4. Figure 4.7 shows the Type 1 error rates for RM_{SS} using the settings outlined in Table 4.3 for which $RM_{SS} = 1$. In these figures, dark gray symbols with dashed lines represent ZILN analyses and light gray symbols with dotted lines are used for ZIG analyses. As in previous sets of figures, the plots on the left represent true ZIG data, those on the right represent true ZILN data, those on the top represent data simulated such that $\nu^{-1} = 0.5$, and those on the bottom include data simulated with $\nu^{-1} = 2$.

The Type 1 error results shown in Figure 4.6 for the DM based test under the equality of marginal mean settings exhibit similar trends as those that were seen in the general null settings. When data are ZIG and $\nu^{-1} = 0.5$, Type 1 error rates are within the appropriate range or slightly conservative whereas when $\nu^{-1} = 2$, Type 1 error rates are excessively conservative, especially when the data are analyzed as ZILN. For ZILN data, ZILN analyses are within range when $\nu^{-1} = 0.5$ and slightly conservative when $\nu^{-1} = 2$ and the correlation between group and gender is high. Misspecified ZIG analyses have elevated Type 1 error rates which increase with ν^{-1}. For ZIG data with $\nu^{-1} = 2$, setting 'c' has higher (closer to nominal) Type 1 error rates than settings 'b' and 'd'; setting 'c' has stronger dissonant effects than the other settings which may be counteracting slightly the tendency of analyses where $\nu^{-1} = 2$ to be extremely conservative. The bottom right plot shows that for DM based tests with the ZILN data analyzed as ZIG, a factor driving elevated Type 1 error is the level of zero-inflation which impacts the expected sample size for the continuous part

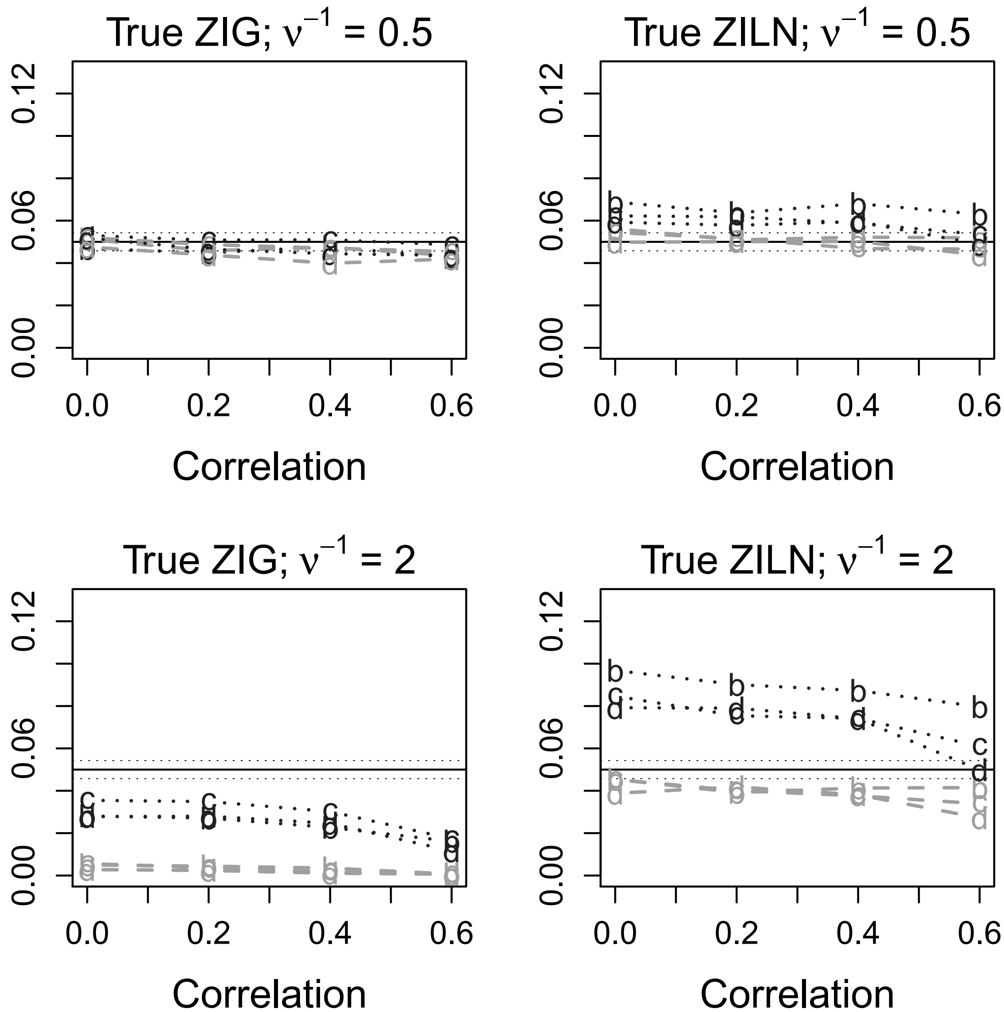

Figure 4.6: Type 1 error for DM at metric-based null adjusting for a dichotomous covariate. Labels refer to settings outlined in Table 4.4 and darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

of the model. Lower levels of zero-inflation lead to higher Type 1 error rates; setting 'b' which has the highest probabilities of nonzero values also has the highest Type 1 error rates.

Figure 4.7 contains the Type 1 error results for RM_{SS} for the settings described in Table 4.3 where $RM_{SS} = 1$. The plots on the left show results for ZIG simulated data. For these scenarios, ZILN analyses have higher Type 1 error rates than do the ZIG analyses. For settings 'b' and 'c' when $\nu^{-1} = 0.5$ and the correlation between group and gender was low, Type 1 error rates are slightly elevated for both ZIG and ZILN analyses. At higher levels of correlation, Type 1 error rates are within the appropriate range. Setting 'd' when $\nu^{-1} = 0.5$ and the data are truly ZIG has conservative Type 1 error rates for all levels of correlation between age and gender. When $\nu^{-1} = 2$ and the data are truly ZIG, ZILN analyses have Type 1 error rates that are within the acceptable range and ZIG analyses have slightly conservative Type 1 error rates. For data simulated as ZILN, there are larger differences between the Type 1 error rates among the various settings using RM_{SS} tests than there are for the DM tests. A comparison of the settings plotted shows that the Type 1 error rate increases with the probability of a nonzero outcome. When $\nu^{-1} = 0.5$, setting 'd', which has the greatest amount of zero-inflation, has slightly conservative Type 1 error rates and settings 'b' and 'c' have Type 1 error rates that are somewhat higher than the nominal 0.05 level. When $\nu^{-1} = 2$, Type 1 error rates are near the appropriate levels when ZILN is assumed but still exhibit a spread such that Type 1 error for 'b' is greater than Type 1 error for 'c' which is greater than Type 1 error for 'd'. Where ZIG is the assumed model, the same ordering is found, but the Type 1 error rates are extremely high. Type 1 error for ZILN data analyzed as ZIG declines slightly with increased correlation between group and gender.

Figure 4.8 shows the Type 1 error results for RM_{MAR} for the marginal mean-based null settings outlined in Table 4.4. When ZILN is assumed, Type 1 error

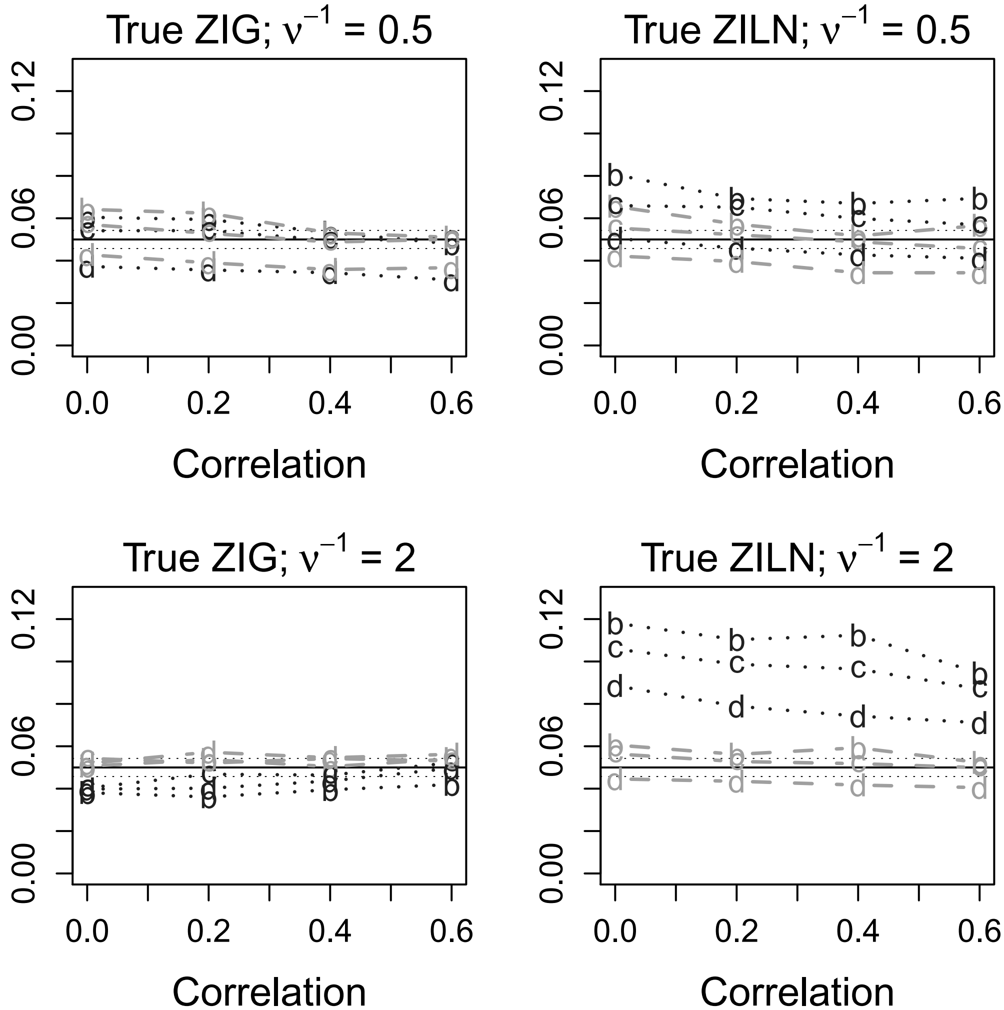

Figure 4.7: Type 1 error for RM_{SS} at metric-based null adjusting for a dichotomous covariate. Labels refer to settings outlined in Table 4.3 and darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

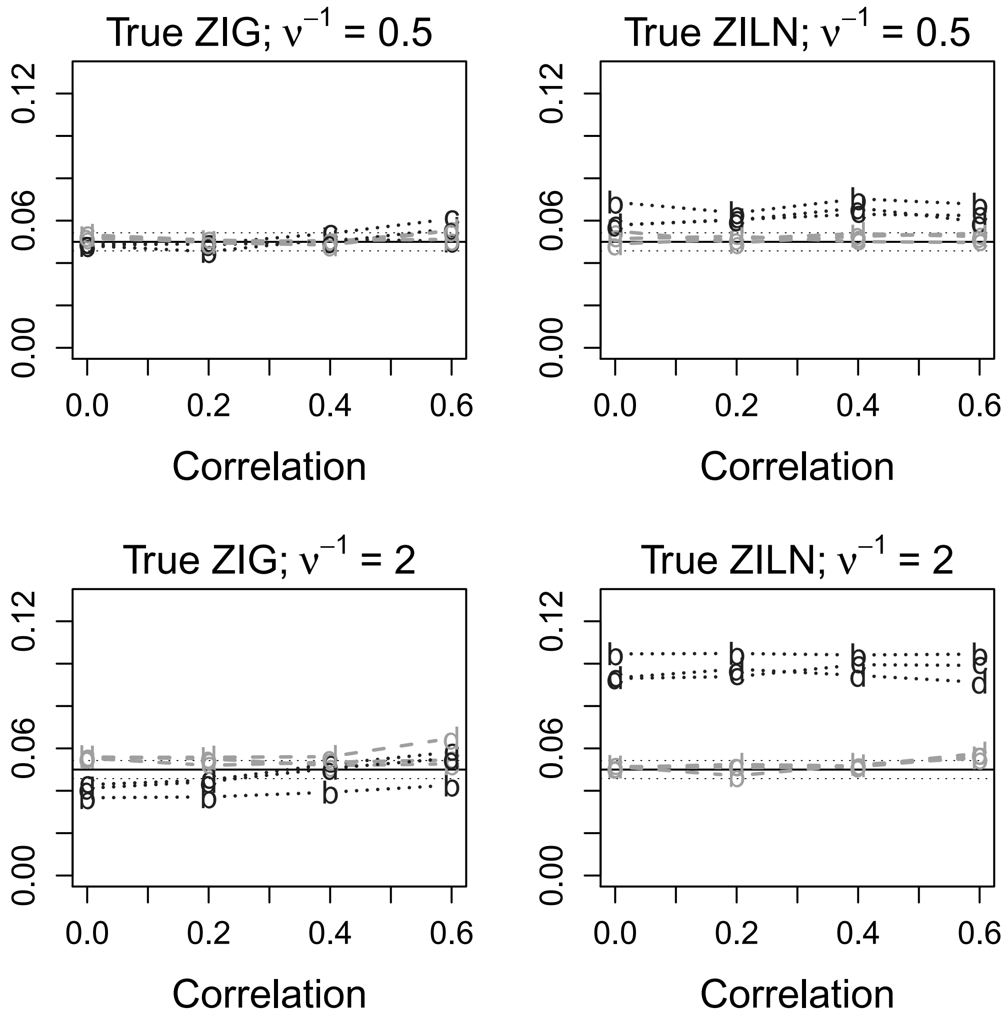

Figure 4.8: Type 1 error for RM_{MAR} at metric-based null adjusting for a dichotomous covariate. Labels refer to settings outlined in Table 4.4 and darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

rates tend to be close to nominal with some very slightly elevated. When the data are simulated from a ZIG distribution and $\nu^{-1} = 0.5$, Type 1 error rates for analyses assuming ZIG are within the appropriate range for most dissonant effect settings, but are slightly elevated for setting 'c' which has the largest dissonant effect. When $\nu^{-1} = 2$ and data are simulated from a ZIG distribution, ZIG analyses are conservative with setting 'b' (smallest dissonant effect size and the smallest amount of zero inflation) being the most conservative. When the data are simulated from a ZILN distribution, Type 1 error rates for ZIG analyses are elevated and setting 'b' has the highest Type 1 errors. Type 1 error rates for ZILN data analyzed with ZIG regression increase with ν^{-1}.

Rank transformed regression (RTR) is simple linear regression using the ranks of the observed data. Specifically, we used methods similar to those proposed by Iman and Conover [21] where both the outcome and the covariates are ranked. The dissonant effect settings from this section were designed to cancel out in terms of the overall mean-based tests. However, they do not necessarily cancel out in terms of ranks. Within each of these settings, the rank distributions differ between the two groups as one group has a higher probability of non-zero values and a higher conditional mean than does the other group. These cancel out so as to have an equal overall group mean, but these settings by definition lead to differences in rank. Therefore, analyses of these setting using RTR could rightly be considered to be exploring the power for finding dissonant effects. The caveat that follows is that if one is interested in mean-based comparisons, RTR will give very incorrect results. Figure 4.9 shows the power of RTR for finding dissonant effects which cancel out in terms of the marginal mean as in Table 4.4 for a sample size of n = 200. Similar results were found for the $RM_{SS} = 1$ settings. Setting 'c' (which has the largest dissonant effects) has the highest powers followed by settings 'd' and then 'b'. The power for the RTR tests increases as ν^{-1} increases. Power decreases as the correlation

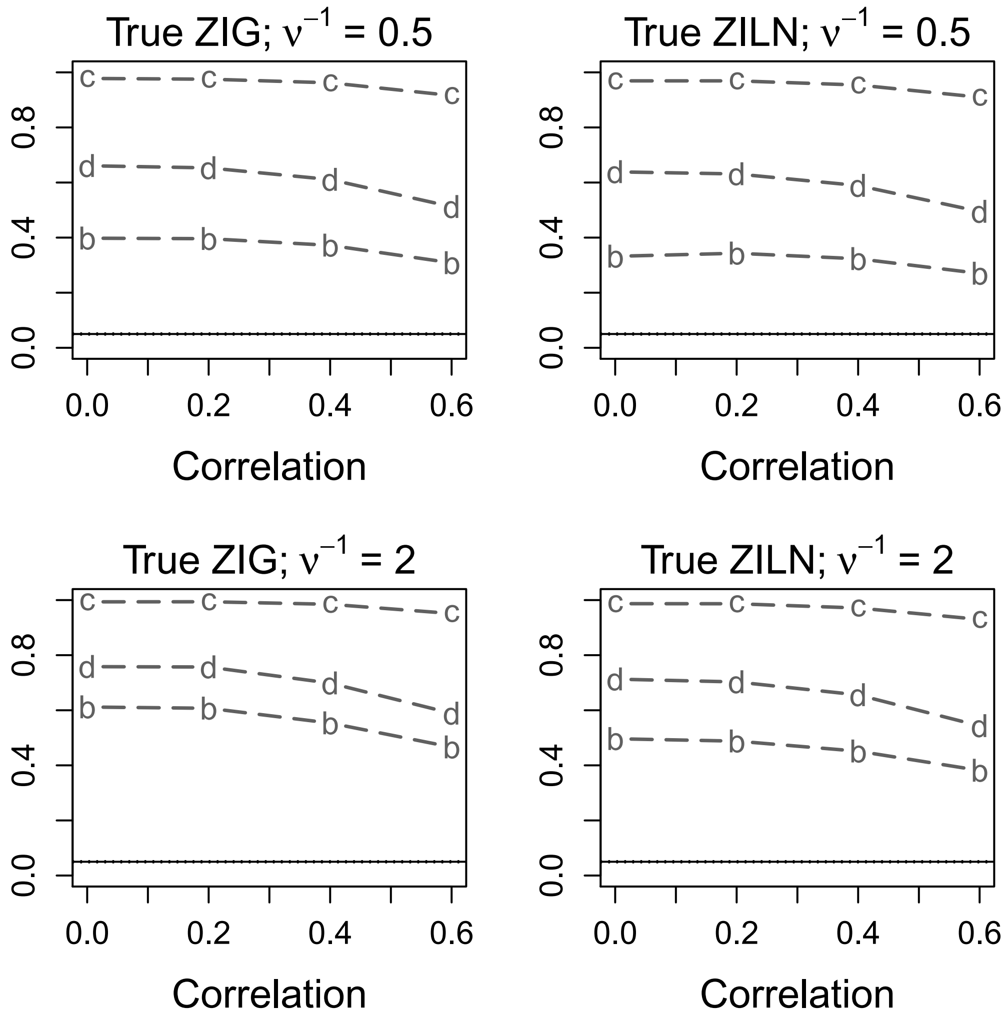

Figure 4.9: Power for Rank Transform Regression at marginal mean based null adjusting for a dichotomous covariate. Labels refer to settings outlined in Table 4.4 and darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

between group and gender increases.

Figures 4.10 to 4.12 compare Type 1 error rates for DM, RM_{SS}, RM_{MAR}, and SLR. For these comparisons, settings are shown which place each metric at its own null; i.e. RM_{SS} is at the settings from Table 4.3 and all of the other metrics are at the closely related settings from Table 4.4. This is done to show true Type 1 error rates. However, since the settings in the two tables are very close in value to each other, little change in the difference between the Type 1 error rates observed for the various metrics would be seen if identical settings were used for all metrics. In these figures, light gray with dashed lines refer ZILN-based analyses, dark gray with dotted lines refers to ZIG based analyses and black with an alternating dot-dash pattern is used for the simple linear regression results. 'S' represents Type 1 error for the RM_{SS} based tests, 'M' for the RM_{MAR} based tests, 'D' for DM based tests, and 'L' for simple linear regression.

Figure 4.10 compares Type 1 error rates for setting 'b' which has the smallest amount of zero inflation and the smallest dissonant effects relative to the other settings examined in this section. For this setting, SLR remains within the appropriate Type 1 error range when $\nu^{-1} = 0.5$ and in most cases even when $\nu^{-1} = 2$ except for the setting where the data are simulated from a ZILN distribution and the correlation between group and gender is 0.6 where it is slightly elevated. Among the mean-based tests however, there are more Type 1 error discrepancies. When the data are simulated from a ZIG distribution with $\nu^{-1} = 0.5$, RM_{SS} is elevated at low correlations, but all other tests are at appropriate levels. For ZIG data with $\nu^{-1} = 2$ the main departure from nominal Type 1 errors is in the direction of extreme conservativeness. DM assuming ZILN is the worst offender with Type 1 error rates close to 0, ranging from 0.0007 to 0.005. The second most conservative metric for ZIG data with $\nu^{-1} = 2$ is DM assuming ZIG. For such data, RM_{SS} and RM_{MAR} both have slightly conservative Type 1 error rates when assuming ZIG, but when assuming ZILN their

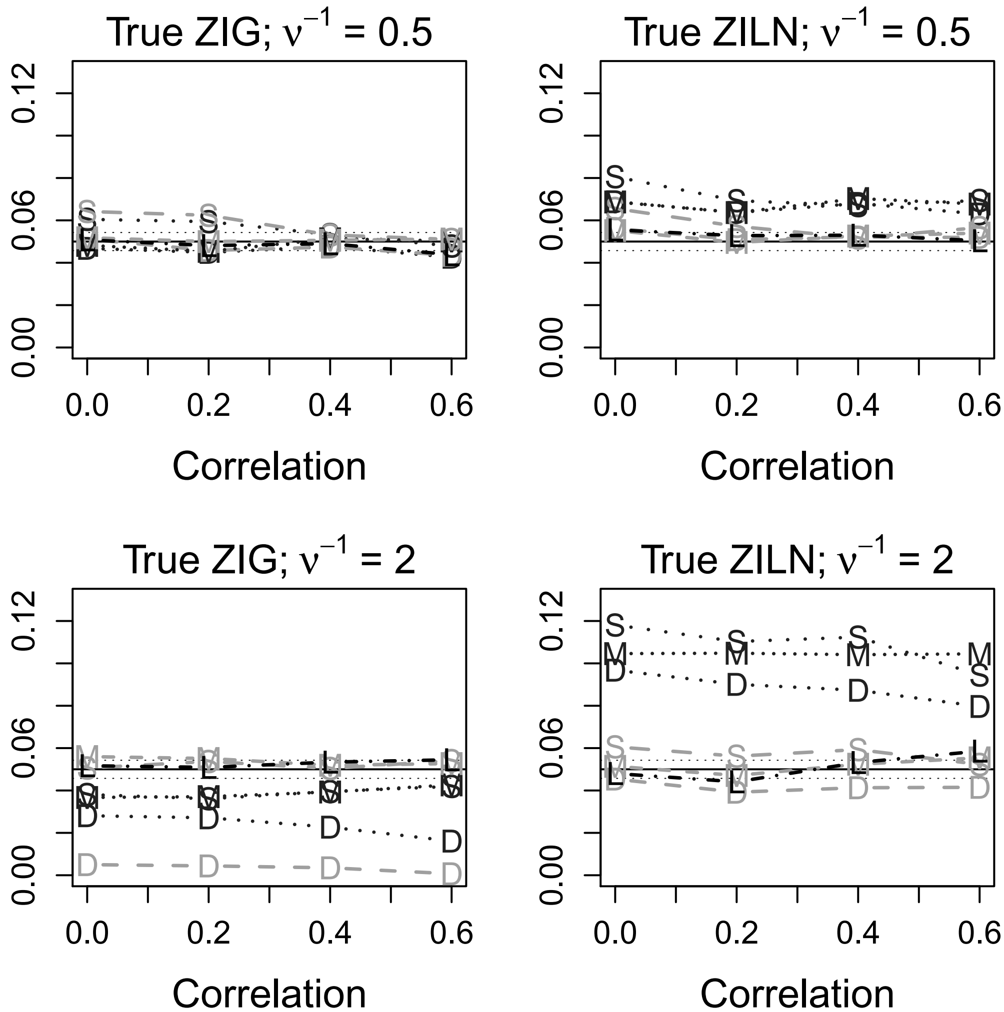

Figure 4.10: Type 1 error for all mean-based tests at metric-based null adjusting for a dichotomous covariate; setting 'b'. The symbol 'D' represents the Type 1 error rates for tests based on DM, 'S' those based on RM_{SS}, and 'M' those based on RM_{MAR} with darker grey symbols with dotted lines for ZIG analyses and lighter grey symbols with dashed lines for ZILN analyses. SLR analysis results are shown with the symbol 'L' and black dash-dot lines; RTR with the symbol 'R' and medium grey lines with long dashes.

Type 1 error rates are within or very close to the appropriate range.

When data are simulated from a ZILN distribution and $\nu^{-1} = 0.5$, all mean-based tests have slightly elevated Type 1 error rates when a ZIG distribution is assumed with RM_{SS} having slightly higher Type 1 error rates than the other tests when the correlation between group and gender is small. When ZILN regression is used to analyze ZILN data where $\nu^{-1} = 0.5$, Type 1 error rates are lower than when ZIG is assumed, but RM_{SS} still has a slightly elevated Type 1 error rates when adjusting for gender when there is no correlation between group and gender. When data simulated as ZILN with $\nu^{-1} = 2$ are correctly analyzed as ZILN, RM_{SS} tends to have slightly higher than nominal Type 1 error rates, DM has slightly lower than nominal Type 1 error rates, and RM_{MAR} has Type 1 error rates that are within or close to the acceptable Type 1 error range. The largest departures from the nominal Type 1 error occur when ZILN data with $\nu^{-1} = 2$ are analyzed as ZIG. For most levels of correlation under these scenarios, RM_{SS} has higher Type 1 error rates than the other metrics, followed by RM_{MAR} and then DM. However, Type 1 error for RM_{SS} decreases with correlation between group and gender while RM_{MAR} has similar levels of Type 1 error across correlation levels leading to RM_{MAR} having a higher Type 1 error rate when correlation = 0.6.

Figure 4.11 compares Type 1 errors across metrics for setting 'c'. Setting 'c' has the largest dissonant effect sizes and a zero inflation level in between those of the other two settings. One thing that stands out for these scenarios is that simple linear regression (labeled 'L') is elevated for for ZIG and ZILN data and for all correlations. When $\nu^{-1} = 2$, Type 1 error rates for SLR increase as the correlation between the variable of interest and the adjusting covariate increases. When the data are simulated from a ZIG distribution, Type 1 error rates for SLR are generally higher than those of the mean-based tests, except when $\nu^{-1} = 0.5$ and the correlation between group and gender is 0.6 where RM_{MAR} has similar Type 1 errors to SLR.

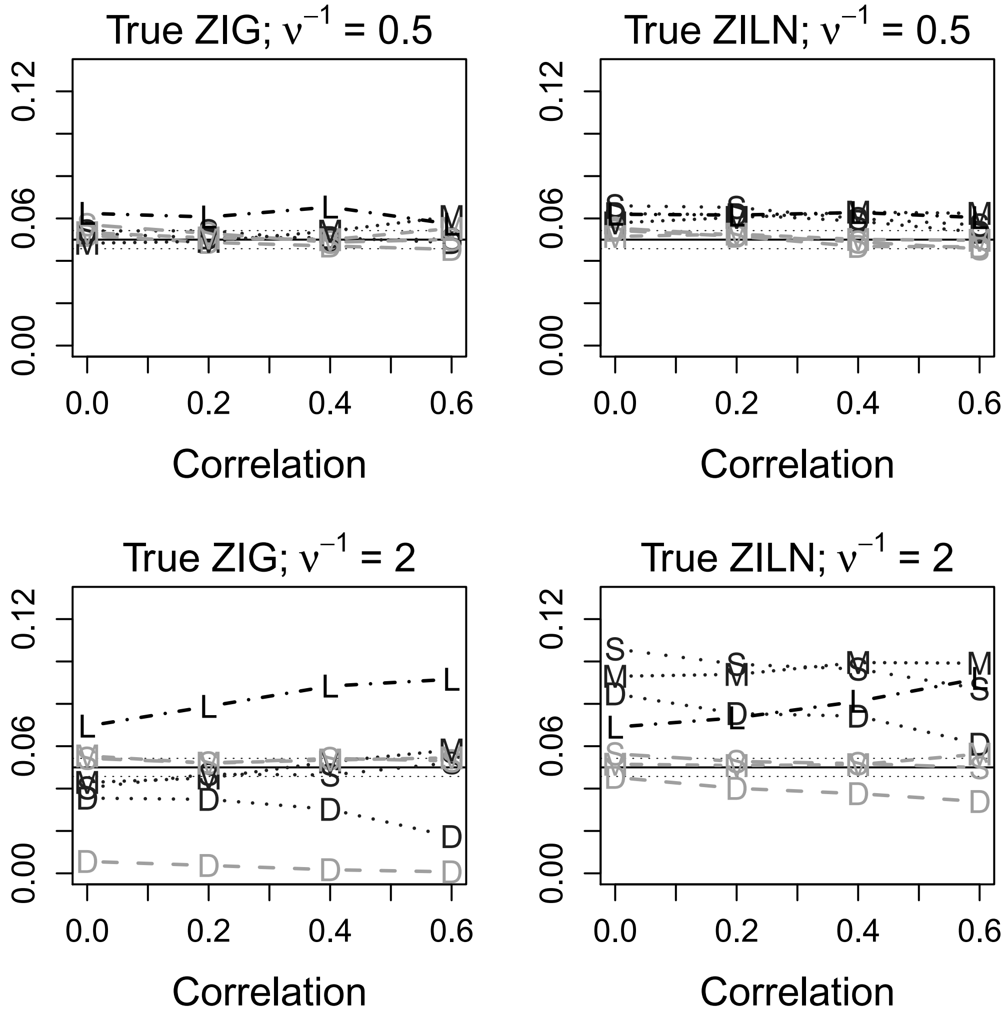

Figure 4.11: Type 1 error for all mean-based tests at metric-based null adjusting for a dichotomous covariate; setting 'c'. 'D' represents the Type 1 error rates for tests based on DM, 'S' those based on RM_{SS}, and 'M' those based on RM_{MAR} with darker grey symbols with dotted lines for ZIG analyses and lighter grey symbols with dashed lines for ZILN analyses. SLR analysis results are shown with the symbol 'L' and black dash-dot lines; RTR with the symbol 'R' and medium grey lines with long dashes.

When data are simulated from ZILN distributions and $\nu^{-1} = 0.5$, SLR yields similar Type 1 errors to DM, RM_{SS}, and RM_{MAR} assuming ZIG and higher Type 1 errors than when ZILN is assumed. When the data comes from a ZILN distribution and $\nu^{-1} = 2$, Type 1 error for SLR increases as the correlation between the covariates in the model increases whereas Type 1 error decreases as correlation increases for RM_{SS} and DM when ZIG is assumed leading to SLR having lower Type 1 errors than the other tests when the correlation is low and a higher Type 1 error than RM_{SS} and DM assuming ZIG when the correlation is high. The mean-based tests have similar comparisons to each other as they did for setting 'b'. For ZILN data RM_{SS} assuming ZIG has the highest Type 1 errors at low correlations and RM_{MAR} assuming ZIG has the highest Type 1 errors when correlation is high. When the data are simulated from ZIG, DM assuming ZILN has the lowest Type 1 errors followed by DM assuming ZIG and then by RM_{SS} and RM_{MAR} assuming ZIG with RM_{SS} and RM_{MAR} assuming ZILN having appropriate Type 1 errors except when correlation is low and $\nu^{-1} = 0.5$ where RM_{SS} has slightly elevated Type 1 error rates.

Figure 4.12 compares Type 1 error rates across tests for setting 'd' as outlined in Table 4.3 for RM_{SS} and in Table 4.4 for the other metrics. Simple linear regression has appropriate Type 1 error when $\nu^{-1} = 0.5$ and there is a nonzero correlation between group and gender, slightly elevated Type 1 errors when $\nu^{-1} = 0.5$ and there is no correlation between group and gender, and more elevated Type 1 errors when $\nu^{-1} = 2$ and group and gender are correlated. When $\nu^{-1} = 2$ and the data are truly ZILN, RM_{MAR} assuming ZIG has the highest Type 1 error rates unlike for the settings presented in the previous figures where RM_{SS} had the highest Type 1 errors. When the data are simulated from a ZIG distribution with $\nu^{-1} = 2$, in a similar as that seen in previous settings, DM assuming ZILN has Type 1 errors of essentially 0, DM assuming ZIG has the next lowest Type 1 errors, RM_{SS} and RM_{MAR} assuming ZIG have slightly conservative to appropriate Type 1 errors, and RM_{SS} and RM_{MAR}

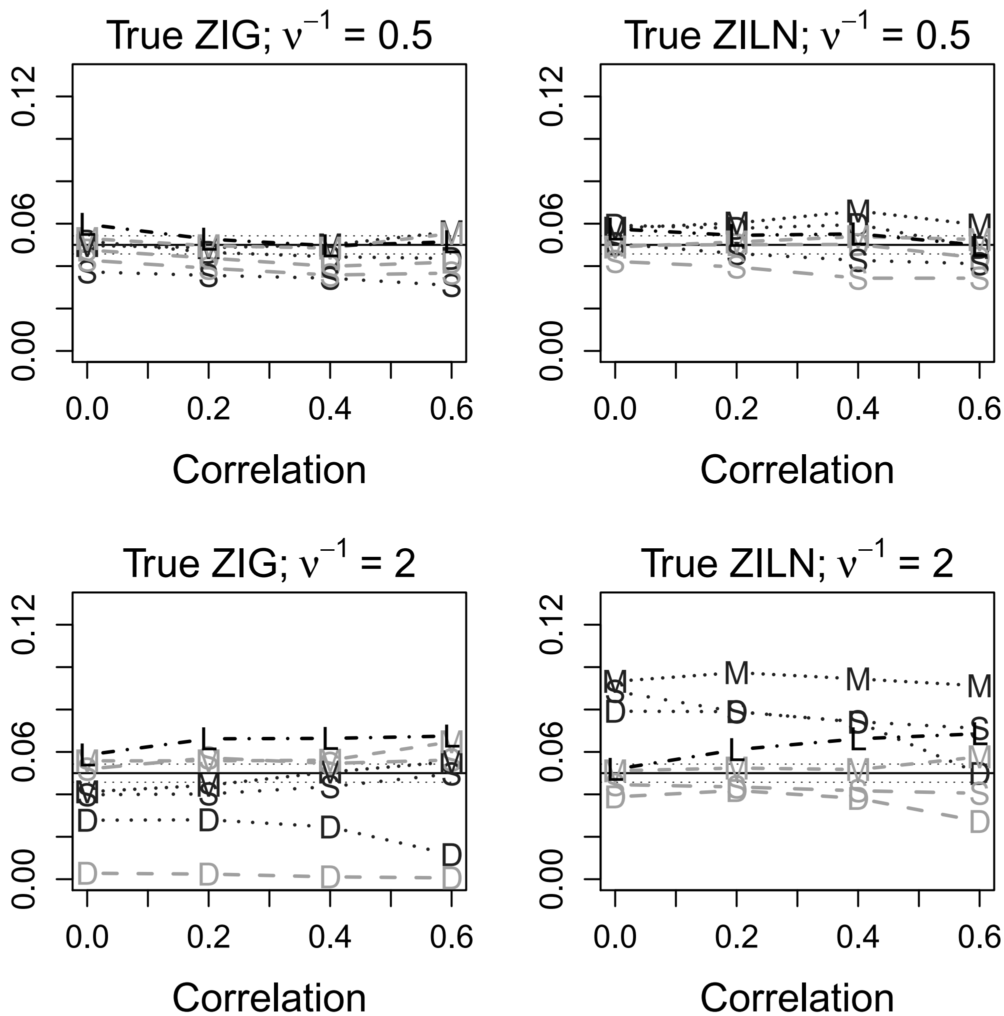

Figure 4.12: Type 1 error for all mean-based tests at metric-based null adjusting for a dichotomous covariate; setting 'd'. 'D' represents the Type 1 error rates for tests based on DM, 'S' those based on RM_{SS}, and 'M' those based on RM_{MAR} with darker grey symbols with dotted lines for ZIG analyses and lighter grey symbols with dashed lines for ZILN analyses. SLR analysis results are shown with the symbol 'L' and black dash-dot lines; RTR with the symbol 'R' and medium grey lines with long dashes.

assuming ZILN have appropriate to slightly elevated Type 1 errors.

In summary, Figures 4.10 to 4.12 show that the relationship of SLR and the mean-based tests have some similarities and some differences across settings. Consistent patterns include the relative order of the Type 1 error rates for the various metrics other than SLR when the data are distributed ZIG, DM being the most extremely conservative in that setting, and the general inflation of Type 1 error with ZILN data are analyzed as ZIG with DM being the least inflated of the mean-based metrics. SLR Type 1 error rate relative to those of the other metrics changed the most from setting to setting, being appropriate when zero-inflation rates are relatively small and too high for the other settings. RM_{SS} assuming ZIG for ZILN simulations with $\nu^{-1} = 2$ also changed across relative to the other metrics, being the highest for some settings, and being lower than RM_{MAR} assuming ZIG for others.

4.6.3 Power for Consonant Tests

Figures 4.13 - 4.16 show the power results for DM, RM_{SS}, and RM_{MAR} tests under the settings described in Tables 4.5 and 4.6. All of these setting have the same conditional mean and the same probability of nonzero values for group 0 females, positive effects for group in one or more parts of the model leading to higher probabilities of nonzero values and/or higher conditional means for the group 1, and positive effects for the adjusting covariate gender resulting in to higher means and/or probabilities for males than for females. The symbol 'G' represents settings where there are nonzero effects of group and gender only in the binomial part, 'H' represents settings where there are nonzero effects in both parts of the two-part models, and 'I' is used for settings where the only nonzero effects are in the continuous part. Figure 4.14 shows power results for RM_{SS} for settings where $RM_{SS} = 1.44$ for group 1 vs group 0 and $RM_{SS} = 1.44$ for male vs female. Figure 4.13, 4.15, and 4.16 show power results for DM, RM_{SS}, and RM_{MAR} where $RM_{SS} = 1.44$ for group 1 vs group

0 and the gender effects are equal to the square root of the group effects.

The power results for DM based tests under the settings in Table 4.6 are shown in Figure 4.13. For DM tests for all true distributions, assumed 0 distributions, and ν^{-1} settings, power is highest when the effect is in the binomial part of the model followed by when there is an effect in both parts of the model. When the data are simulated from a ZIG distribution, ZIG analyses have higher powers than ZILN analyses. This is especially true when $\nu^{-1} = 2$ in which case ZILN analyses have very low power levels. For data simulated from a ZILN distribution with $\nu^{-1} = 0.5$ the power is similar for both ZIG and ZILN analyses. When the data are ZILN with $\nu^{-1} = 2$, ZIG has power levels slightly higher than ZILN analyses. However, such analyses are invalid due to the high Type 1 error rates observed when $\nu^{-1} = 2$. Other simulation results not pictured found that the results had similar patterns when regardless of the values of the adjusting effects.

Unlike the power results seen for DM and those that will be seen for RM_{MAR}, the power of RM_{SS} is effected by the values of the adjusting covariate. Figure 4.14 shows the power results for RM_{SS} when the effect of the gender covariate is equivalent to the effect of the group ($RM_{SS} = 1.44$ for both group and gender effects). These settings are laid out in Table 4.5. Figure 4.15 shows results for a set of simulations where there is a weaker gender effect and are the same settings used for the DM and RM_{MAR} figures. Both RM_{SS} based figures show that regardless of the strength of the adjusting effect, analyses assuming ZIG tend to have higher powers. Where they differ is in terms of the relative powers of settings with effects in the binomial part vs. continuous part vs. in both parts. When there is a strong adjusting effect (Figure 4.14), the setting with nonzero effects in both parts of the model ('H') has the higher power than setting where the effect is only in one part of the model ('G' and 'I'), except when the data are from a ZIG distribution with $\nu^{-1} = 2$ where all settings have roughly equal power. In comparison, when the adjusting effect (the

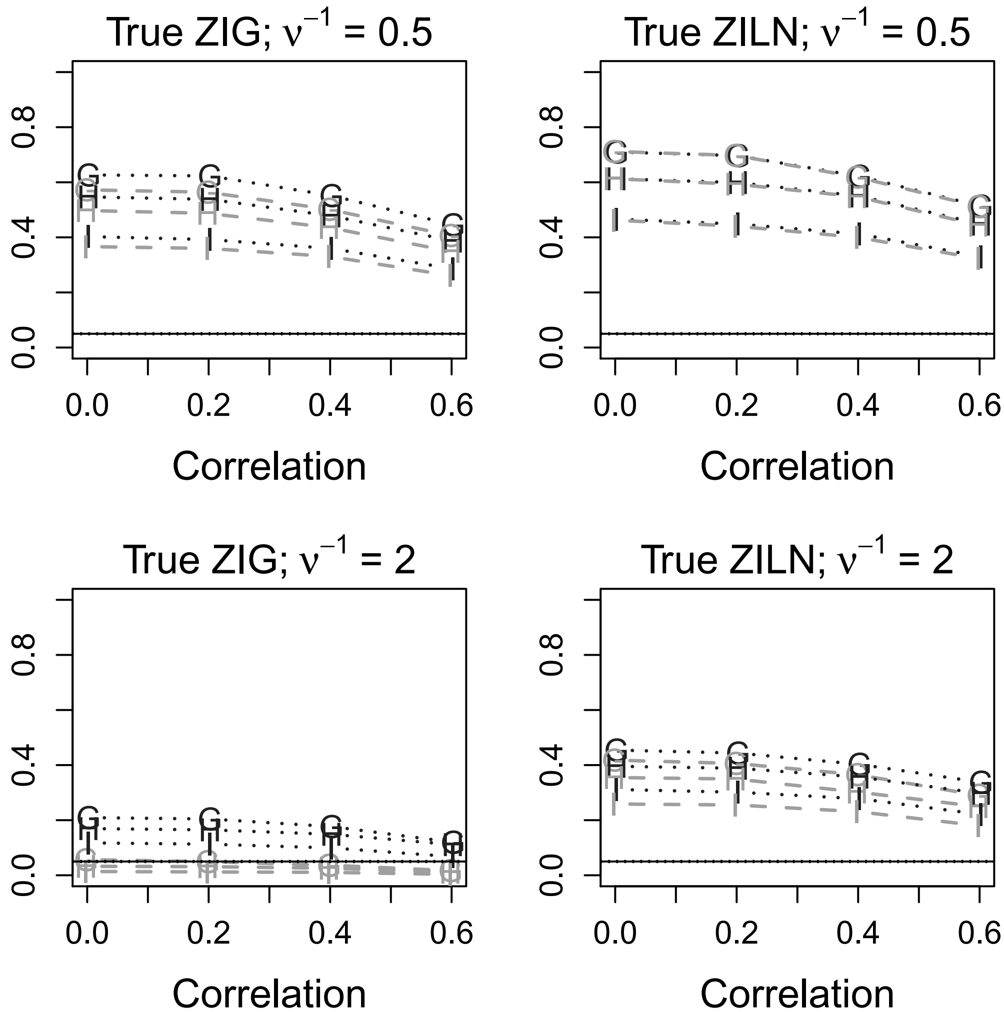

Figure 4.13: Power for DM with a dichotomous adjusting covariate. Setting labels correspond to those outlined in Table 4.6. ‘G’ represents effects only in the binomial part of the model, ‘I’ effect only in the continuous part of the model, and ‘H’ effects which are split between the two parts. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

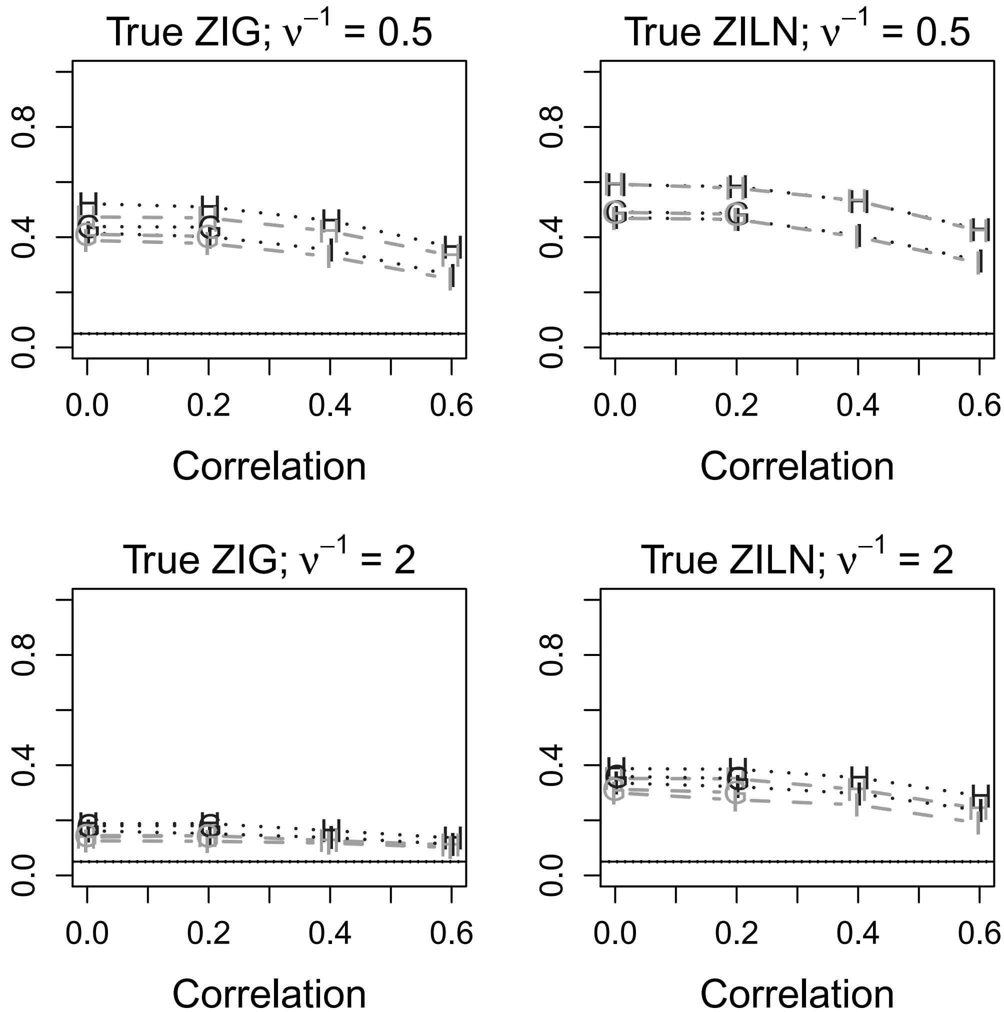

Figure 4.14: Power for RM_{SS} with a strong dichotomous adjusting covariate effect. Setting labels correspond to those outlined in Table 4.5. 'G' represents effects only in the binomial part of the model, 'I' effect only in the continuous part of the model, and 'H' effects which are split between the two parts. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

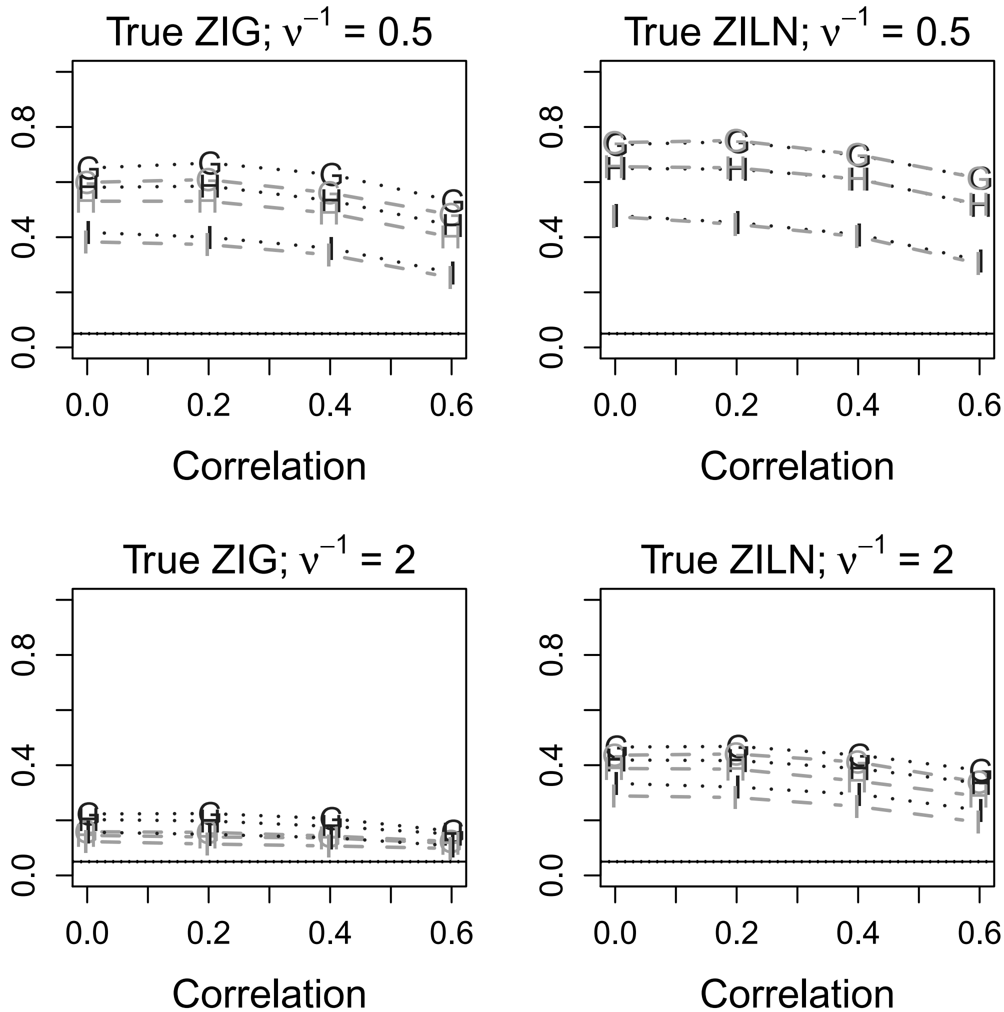

Figure 4.15: Power for RM_{SS} with a weak dichotomous adjusting covariate effect. Setting labels correspond to those outlined in Table 4.6. 'G' represents effects only in the binomial part of the model, 'I' effect only in the continuous part of the model, and 'H' effects which are split between the two parts. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

effect of gender) is small relative to the effect of interest (group) as seen in Figure 4.15, simulations with nonzero effects in the binomial part of the model have the highest power followed by those where the effect is in both parts with settings with the effect only in the continuous part having the lowest powers.

The power results for RM_{MAR} are shown in Figure 4.16. Similar to what was seen for DM, power increases as the effect size in the binomial part of the model increases (e.g. 'G'>'H'>'I'). ZIG analyses tend to have higher powers than ZILN analyses. However, the differences between ZIG and ZILN analyses are smaller than those observed for the DM analyses. This is particularly evident when ZIG data are simulated with $\nu^{-1} = 2$, where powers from ZILN analyses are smaller than those from ZIG analyses. For data simulated from a ZILN distribution with $\nu^{-1} = 2$, it must be noted that although ZIG has higher power this must be discounted due to the high Type 1 error rates observed. Other simulations not pictured here included different values of the adjusting effects but the same the group RM_{SS} values found similar relationships of power to binomial effect size and ZIG versus ZILN model choice as those pictured.

Figures 4.17 - 4.19 compare DM, RM_{SS}, RM_{MAR}, SLR, and RTR power results within various settings. Figure 4.17 shows power results for conditions where $\nu^{-1} = 0.5$ and the adjusting effect is equal to the effect of interest (settings from Table 4.5). Figure 4.18 shows power results for when $\nu^{-1} = 0.5$ but the adjusting effect is weaker than the group effect (settings from Table 4.6). Figure 4.19 includes the power results for $\nu^{-1} = 2$ with a strong adjusting effect (Table 4.5). Each figure includes one plot per mean model setting outlined in Table 4.5 and Table 4.6; the top plots are for setting 'G' where the only nonzero effects are in the binomial part of the model, the middle plots are for setting 'I' where the only nonzero effects are in the continuous part of the model, and the bottom plots include settings where the effects are equal split between the two parts, 'H'. In these plots, the symbol 'D' is used for the power

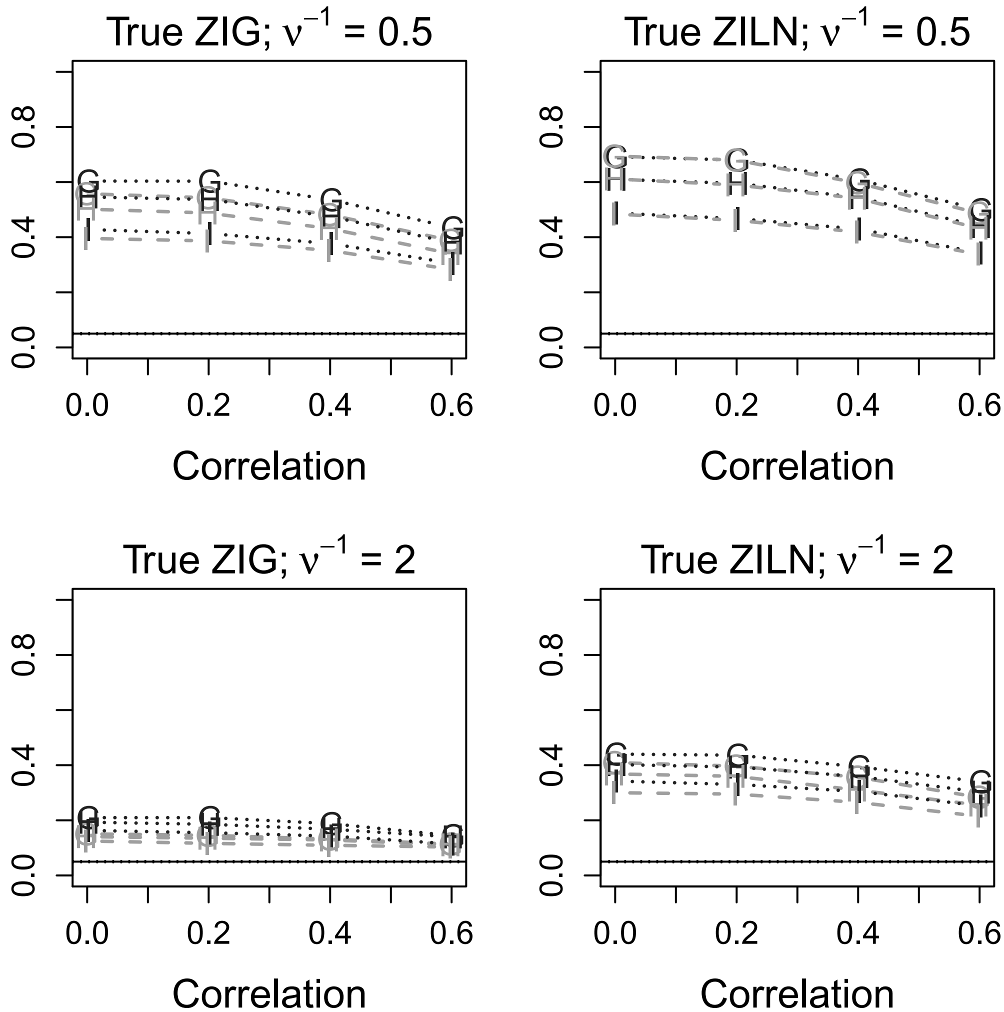

Figure 4.16: Power for RM_{MAR} with a dichotomous adjusting covariate. Setting labels correspond to those outlined in Table 4.6. 'G' represents effects only in the binomial part of the model, 'I' effect only in the continuous part of the model, and 'H' effects which are split between the two parts. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

of DM tests, 'S' is used for subjects specific ratio of means (RM_{SS}), 'M' for the ratio of the marginal means (RM_{MAR}), 'L' is used for simple linear regression, and 'R' for rank transformed regression. ZILN analyses are shown via light gray symbols and dashed lines, ZIG analyses are shown via dark gray symbols and dotted lines, SLR is plotted in black with alternating dash-dot lines, and RTR is plotted with medium gray symbols and long dashed lines.

Figure 4.17 shows the power results for various simulation settings where $\nu^{-1} = 0.5$ and the adjusting effects are of the same magnitudes as the group effects. The relative power levels of the metrics compared are different depending on the group setting and the true distribution form. For setting 'G' under these conditions regardless of the simulated distribution, RTR has the highest powers and RM_{SS} has the lowest powers in between these DM and RM_{MAR} are very close in power levels with DM having slightly higher power at low levels of correlation and RM_{MAR} having slightly higher power at higher levels of correlation; simple linear regression has slightly higher power that DM and RM_{MAR} under a true ZIG scenario and a slightly lower power than DM and RM_{MAR} under a true ZILN scenario. At the highest levels of correlation for setting 'G' less than 99% of the simulated data sets had estimable parameters and are therefore not included in these top plots. The middle plots show that when the nonzero effects are in the continuous part of the model, RTR has very low power relative to the other metrics examined. For this setting ('I'), RM_{MAR} has the highest power of the mean-based tests followed by RM_{SS} and then DM. For ZIG data under setting 'I', the power observed for SLR declines with correlation at a faster rate than the two-part model based tests, leading it to have powers in the middle of the pack when correlation is low and the lowest powers when the correlation is high. For ZILN data under setting 'I', SLR has lower powers than any of the mean-based two-part model tests. The bottom plots show that for this strong adjusting effect case when the effect is in both part of the model (setting 'H') RM_{SS} has the highest

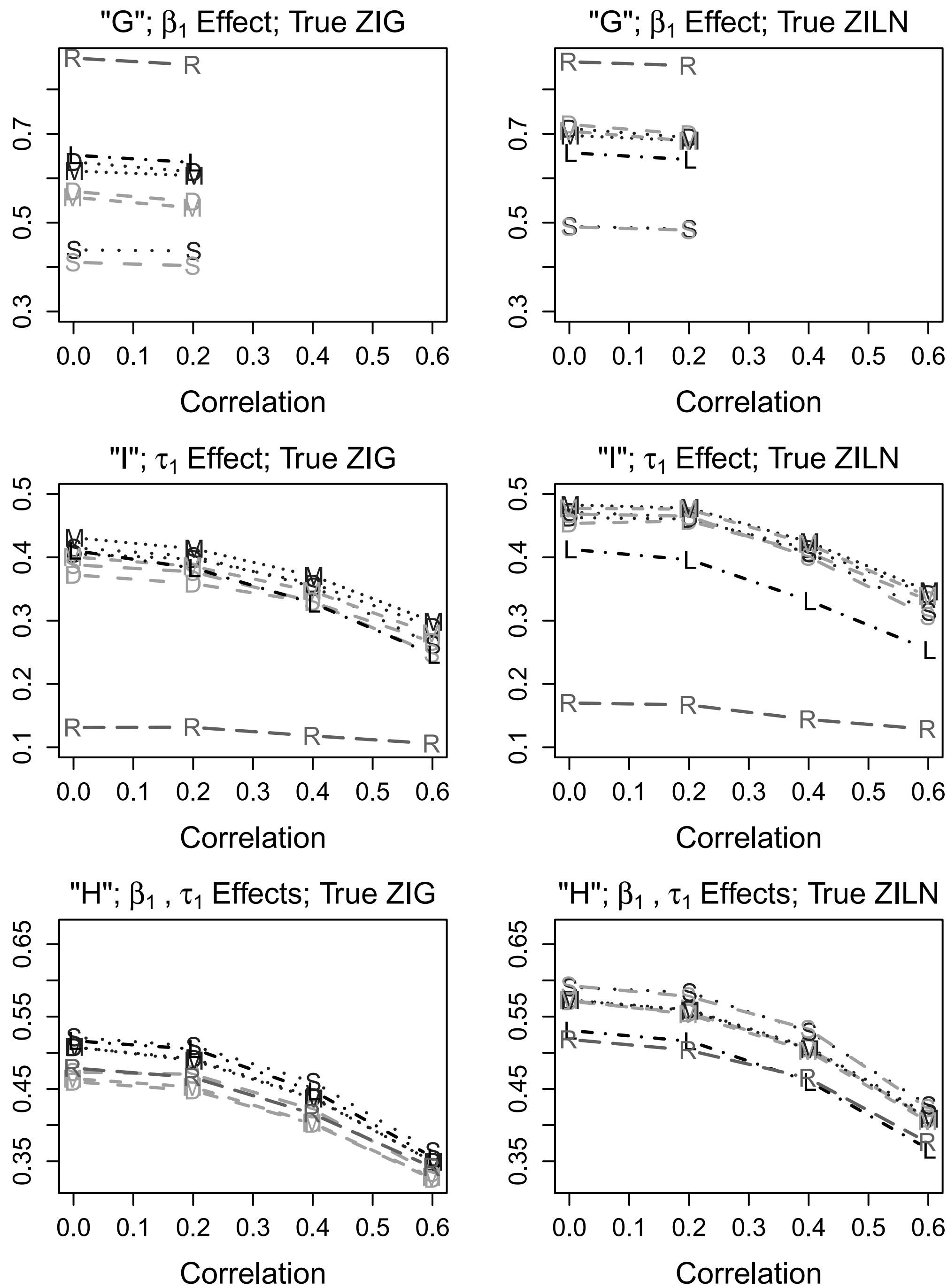

Figure 4.17: Comparison of power for mean-based tests adjusting for a dichotomous covariate with $\nu^{-1} = 0.5$. Graphs titles indicate settings as outlined in Table 4.5. 'D' represents the power for tests based on DM, 'S' those based on RM_{SS}, and 'M' those based on RM_{MAR} with darker grey symbols with dotted lines for ZIG analyses and lighter grey symbols with dashed lines for ZILN analyses. SLR analysis results are shown with the symbol 'L' and black dash-dot lines; RTR with the symbol 'R' and medium grey lines with long dashes.

powers, the powers for DM and RM_{MAR} are slightly lower. When the data are truly ZIG, SLR has powers that are lower than RM_{SS} assuming ZIG but slightly higher than DM and RM_{MAR} assuming ZIG; RTR has powers lower than the ZIG analyses, but higher than the ZILN analyses. When the data are truly ZILN, SLR and RTR have lower powers than the mean-based tests.

In Figure 4.18 the power results are shown for various simulation settings where $\nu^{-1} = 0.5$ and adjusting effects are smaller in magnitude than the group effects. The greatest differences between these weak adjusting covariate effect settings when compared to the strong adjusting effect shown in Figure 4.18 are seen in RM_{SS} when the effect of group is only in one part of the model. In contrast to the previous results, for the weaker adjusting effect settings shown in Figure 4.18 when the effects are only in the binomial part of the model (setting 'G') RM_{SS} has higher powers than do RM_{MAR} or DM. For these settings, RTR shows the same patterns as it did with the stronger effect, power levels are higher than those for the other tests when the effects are in the binomial part, lower when the effects are only in the continuous, in the middle of the pack when effects are in both and the data are ZIG, and lower than the mean-based tests but higher than SLR when effects are in both parts of the model and the data are ZILN. In the settings in Figure 4.18, DM and RM_{MAR} have very similar power levels with DM having slightly higher powers when the effects are in the binomial part of the model ('G') and RM_{MAR} having slightly higher powers when the effects are in the continuous part ('I').

In Figure 4.19, $\nu^{-1} = 2$ and the power settings outlined in Table 4.6 are used to compare the powers of DM, RM_{SS}, RM_{MAR}, SLR, and RTR. For a majority of the tests and settings powers are lower here than they were for the smaller ν^{-1} explored previously. The exception to that is RTR when the effect in is the binomial part of the model. RTR still has powers of around 80% to detect group differences when the true difference has an $RM_{SS} = 1.44$ which comes entirely from the binomial part of

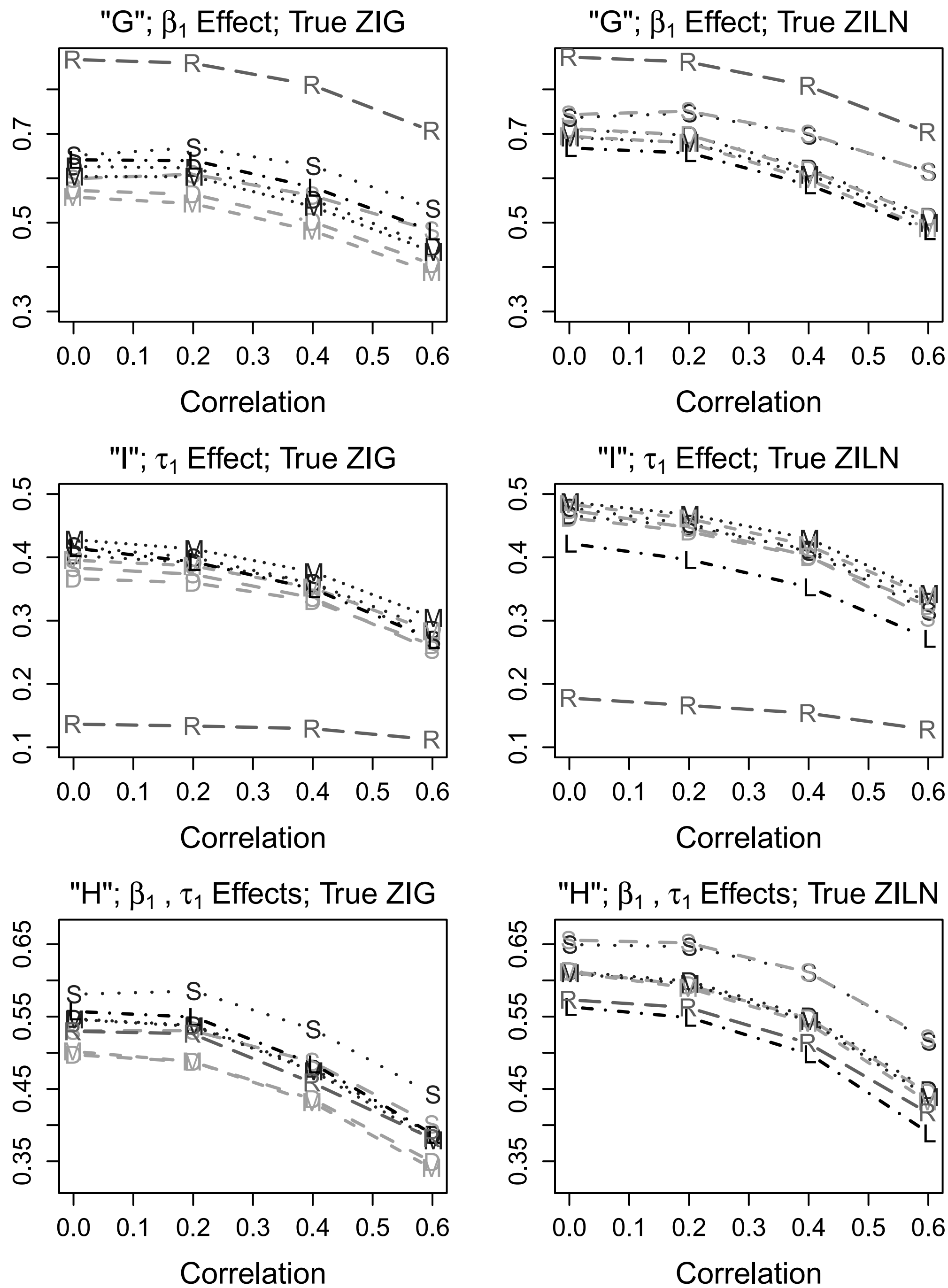

Figure 4.18: Comparison of power for mean-based tests adjusting for a weak dichotomous covariate with $\nu^{-1} = 0.5$. Graphs titles indicate settings as outlined in Table 4.6. 'D' represents the power for tests based on DM, 'S' those based on RM_{SS}, and 'M' those based on RM_{MAR} with darker grey symbols with dotted lines for ZIG analyses and lighter grey symbols with dashed lines for ZILN analyses. SLR analysis results are shown with the symbol 'L' and black dash-dot lines; RTR with the symbol 'R' and medium grey lines with long dashes.

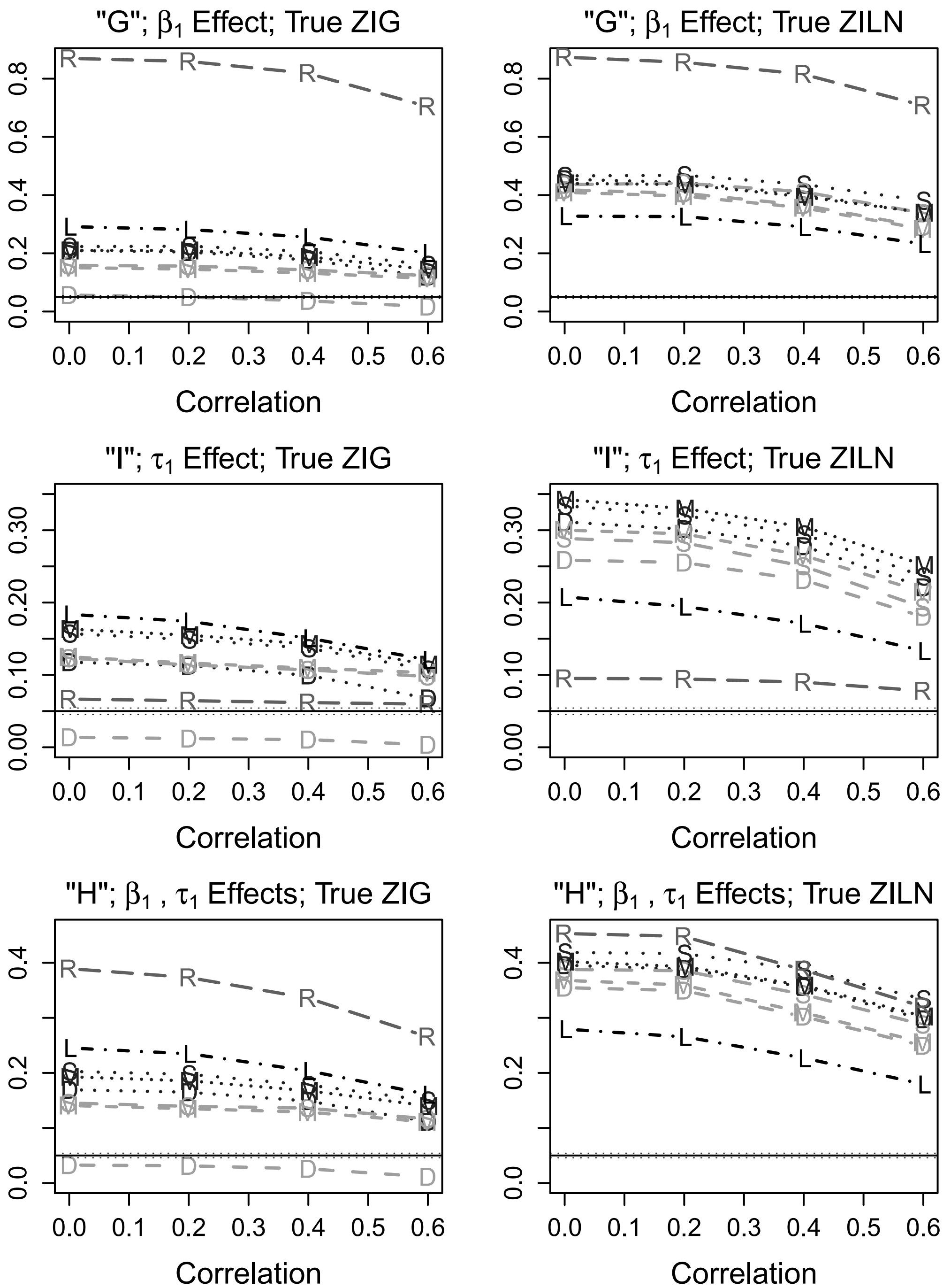

Figure 4.19: Comparison of power for mean-based tests adjusting for a dichotomous covariate with $\nu^{-1} = 2$ and a weak adjusting effect. Graphs titles indicate settings as outlined in Table 4.5. 'D' represents the power for tests based on DM, 'S' those based on RM_{SS}, and 'M' those based on RM_{MAR} with darker grey symbols with dotted lines for ZIG analyses and lighter grey symbols with dashed lines for ZILN analyses. SLR results are shown with the symbol 'L' and black dash-dot lines; RTR with the symbol 'R' and medium grey lines with long dashes.

the model ('G'). For setting 'G', the other tests have much lower power levels than RTR; among them SLR has the highest power when the true data distribution is ZIG and the lowest power when the data are ZILN distributed. For 'G' with a true ZIG, tests assuming ZIG have higher powers than those assuming ZILN; the tests assuming ZIG have very similar powers with RM_{MAR} having slightly higher power only when correlation of group and gender is high, while among the test assuming ZILN DM has the lowest powers (around 0.05) and RM_{MAR} and RM_{SS} have similar power levels. When the data are ZILN with effects in the binomial part of the model, DM and RM_{MAR} have the highest powers of the parametric tests with inappropriate ZIG analyses having higher power levels than the appropriate ZILN analyses.

The middle set of plots show settings where the effect of group is in the continuous part only. Among the mean-based two-part model tests shown in these plots, RM_{MAR} has the highest powers followed by RM_{SS} with DM having the lowest powers of the ZIG and ZILN tests. Also, for setting 'H', with ZIG simulated data SLR has higher power levels than even correctly specified RM_{MAR} and RM_{SS} tests. Power levels obtained under SLR decline more steeply than do the power levels of the other metrics. When the data are ZIG for setting 'I' SLR has the highest power at low group by gender correlations but similar power to RM_{MAR} and RM_{SS} when correlation is high. When the data are from a ZILN distribution and the setting is 'I', SLR has lower power than the mean-based tests, but higher power than RTR.

For the setting in the bottom plots both parts of the model have nonzero effect sizes. Here RTR has the highest powers. When the data are ZIG for these settings, SLR has the next highest power followed by RM_{SS} and RM_{MAR} which are nearly equivalent and DM which has the lowest power which is extremely low when assuming ZILN. For the true ZILN distribution, RM_{SS}, RM_{MAR}, and DM are a bit closer in power, but DM still has the lowest power among them with SLR having the lowest power overall. ZIG analyses have higher powers than ZILN analyses when the data are truly ZILN, but should be disregarded due to the high Type 1 errors when ZILN data with $\nu^{-1} = 2$ are analyzed as ZIG.

4.7 Simulation Methods for Comparing Two Groups When Adjusting for a Continuous Covariate

This section outlines the simulations utilized to explore Type 1 error and power for two-group comparisons using DM, RM_{SS}, and RM_{MAR} while adjusting for a continuous covariate. The data were simulated through the same two-part process described in Section 4.5 but with an adjusting covariate Z that is continuous rather than dichotomous. First Y_i^* is simulated from a binomial distribution such that $\text{logit}(P(Y_i^* = 1|X_{1i} = x_{1i})) = \beta_0 + \beta_1 x_{1i} + \beta_2 z_i$. Then if $Y_i^* = 0$, set $Y_i = 0$ and if $Y_i = 0$ simulate Y_i from a gamma with $Y_i \sim Gamma(e^{\tau_0 + \tau_1 x_{1i} + \tau_2 z_i}, \nu)$ for ZIG regression or from a log-normal distributions with $\log(Y_i) \sim N(\tau_0 + \tau_1 x_1 i + \tau_2 z_i, \sigma^2)$ for ZILN regression. For all simulations, equal group sizes were used. Simulation results will be displayed and discussed which have an overall sample sizes of n = 200.

For the adjusting covariate we are using age in years with ages ranging from

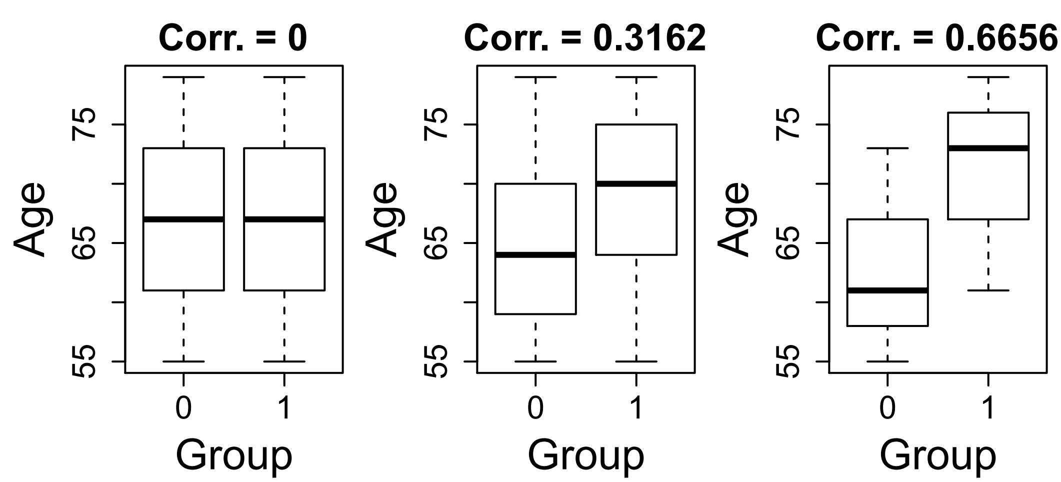

Figure 4.20: Age-by-group correlation settings

55 to 79 year in order to relate to our data example of older adults. We used 3 fixed age-by-group settings which are displayed in box-plots form in Figure 4.20. For the first age-by-group setting, the distribution of age is identical for group 0 and group 1 with 4 subjects at each age from 55 to 79. For the second age-by-group setting, a correlation between age and group equal to 0.31618 was created with group 0 having 6 subjects at each age less than 60, 4 subjects at each age from 61-73, and 2 subjects at ages 74-79 with pattern reversed for group 1 such that group 1 has 2 subjects at the younger ages and 6 subjects at the older ages. The final age-by-group setting was created with even greater difference between the ages of the two groups. The correlation of 0.66564 was created by letting group 0 have 8 subjects at each age from 55 to 60, 6 subjects at 61, 4 subjects at each age from 62 to 72, 2 subjects at age 73, and no subjects older than 73. The age distribution for group 1 was then created such that there were no subjects younger than 61, 2 subject at age 61, 4 at each age from 62 to 72, 6 at age 73, and 8 subjects at each age from 74 to 79. All three age-by-group settings have an overall age distribution that is a fixed discrete distribution with values ranging from 55 to 69. This discrete uniform will be used not only in simulating the data, but also as the distribution over which the means and ratios are marginalized to create the estimates of DM, RM_{SS}, RM_{MAR} and their resultant tests.

For each parameter setting outlined in Section 4.7.1, we simulated 10,000 data sets. Models were estimated via the Newton-Raphson technique as outlined in Section 2.4 and test statistics were computed as described in Sections 4.3 and 4.4. Data sets for which the maximization technique did not converge were excluded from the Type 1 error and power calculations. No settings are being reported on where less than 99% of the data sets had effects which the Newton-Raphson technique could estimate.

4.7.1 Simulation Settings

The first set of simulation studies in this section were performed to examine the Type 1 error rates under 'general null' settings where there are no group effects in terms of either the probability of nonzero values or the mean of the nonzero values. The second set examines Type 1 error under 'dissonant null' settings where any difference in groups cancels out either in terms of the marginalized mean or in terms of RM_{SS}. The final set of simulation studies examines the comparative power for the various metrics proposed in Sections 4.3 and 4.4. The β and τ values used for these four types of settings were determined on the basis of their effect on either RM_{SS} or RM_{MAR}. These determined values were slightly different across the age-by-group settings. Because of this, we will outline in this section the structure of these settings and the method of determining the exact β and τ values rather than creating several tables of the values themselves.

To create the settings for our simulation studies, we first define a way to identify the effect of the continuous variable (e.g. age) on the marginalized means or the marginalized subject specific ratio of means. An individual subject specific ratio of mean for age is defined by examining a one unit effect centered around the data, taking the ratio of the mean value for age + 0.5 over the mean value for age - 0.5. To create a marginalized subject specific ratio of means these individual ratios were marginalized in terms of both the observed values of age and the observed group membership. Consistent with previous sections, let the age variable be Z and the group variable X_1. Then $RM_{SS}(Z)$ can be defined as:

$$RM_{SS}(Z) = E_{X_1,Z}\left(\frac{E(Y|X_1 = x_{1i}, Z = z_i + .5)}{E(Y|X_1 = x_{1i}, Z = z_i - .5)}\right). \tag{4.37}$$

A continuous version of a marginalized mean ratio, RM_{MAR}, will be defined as the ratio of the marginalized mean assuming all subjects are 0.5 years older than they were observed over marginalized mean where all subjects are 0.5 younger than observed. This continuous version of RM_{MAR} is defined as:

$$RM_{MAR}(Z) = \frac{E_{X_1,Z}(E(Y|X_1 = x_{1i}, Z = z_i + .5))}{E_{X_1,Z}(E(Y|X_1 = x_{1i}, Z = z_i - .5))}. \tag{4.38}$$

Table 4.7 outlines the structure of the settings used to compare Type 1 error when there are no group effects in either part of the model, but where for some settings there are significant age effects. For setting 'B', there is no age effect while for settings 'G', 'H', and 'I' there are positive effects of age that yield RM_{SS} values of 1.05 where this $RMss$ value is reached by either an effect in the binomial part of the model ('G'), in the continuous part of the model ('I'), or split between the two parts ('H'). For all of these settings, $RMss(group) = RM_{MAR}(group) = 1$, and $DM(group) = 0$. For setting 'I', $e^{\tau_2} = 1.05$ and $\beta_2 = 0$. For setting 'H', τ_2 was set such that $e^{\tau_2} = \sqrt{1.05}$ and then given the other parameters and the age-by-group distributions, β_2 was found such that $RMss(age) = 1.05$. For setting 'G', $\tau_2 = 0$ and β_2 was found so that $RMss(age) = 1.05$.

Table 4.8 outlines the structure for the dissonant null settings, the situations where the two groups differ in terms of level of zero-inflation and the mean of the nonzero values. This structure is used for two groups of settings, one based on RM_{SS} and the other based on RM_{MAR}. The two different sets of settings are used so that the true Type 1 error rates may be found for each metric. These settings, involving dissonant effects are such that $RM_{SS} \neq RM_{MAR}$. The differences in settings are not large, but must be taken into account to find true Type 1 error rates. setting 'N' has no age effect and is otherwise identical to the setting 'd' from Table 3.2 in Chapter 3. Setting 'G', 'H', and 'I' were created such that when age equals its mean and median of 67, the expected value given a nonzero outcome and the probability of a nonzero outcome for each group were the same as they were for for setting 'N'. These setting

Setting[1]	β_1	τ_1	$RM_{SS}(group)$	β_2	τ_2	$RM_{SS}(age)$
B	0	0	1	0	0	1
G	0	0	1	++	0	1.05
H	0	0	1	+	+	1.05
I	0	0	1	0	++	1.05

Table 4.7: Type 1 error setting patterns for an adjusting for a continuous covariate framework.

[1] Setting 'B' has no age effect,'G' has age effects only in the binomial portion of the model, 'H' has age effects in both parts of the model, and 'I' has age effects only in the continuous part of the model.

[2] '0' implies the parameter equals 0, '+' implies there is some positive effect, and '++' implies a stronger effect

Setting[1]	β_1	τ_1	RM_{SS}(group) or RM_{MAR}(group)	β_2	τ_2	$RM_{SS}(Age)$ or $RM_{MAR}(Age)$
N	-	+	1	0	0	1
G	-	+	1	++	0	1.05
H	-	+	1	+	+	1.05
I	-	+	1	0	++	1.05

Table 4.8: Type 1 error setting patterns with dissonant effects for an adjusting for a continuous covariate framework.

[1] Setting 'N' has no age effects, 'G' has age effects only in the binomial portion of the model, 'H' has age effects in both parts of the model, and 'I' has age effects only in the continuous part of the model.

[2] '0' implies the parameter equals 0, '+' implies there is some positive effect, and '++' implies a stronger effect

are also designed to have an effect of age with an $RM_{SS}(Age)$ or $RM_{MAR}(Age)$ of 1.05 obtained through positive effects in either only the binomial part of the model, 'G', only in the continuous part of the model, or from both parts of the model.

Finally, Table 4.9 outlines the setting used to compare power for the different metrics. For all power settings shown in this dissertation RM_{SS} for group equals 1.44. Power was observed for combinations of the strength of the age effect and location of the group effect. All settings were created such that when age=67 and group=0, the probability of a nonzero outcome was 0.5 and the conditional mean given a nonzero value was 0.75. For settings labeled 'I' $e^{\tau_1} = 1.44$; for settings labeled 'H' $e^{\tau_1} = 1.2$, and for settings labeled 'G' $e^{\tau_1} = 1$. For τ_2, '++' is used to signify that $e^{\tau_2} = 1.1025$, '+' to signify that $e^{\tau_1} = 1.05$, and $\sqrt{+}$ that $e^{\tau_1} = \sqrt{1.05}$. Given the β_0 and τ_0 values that stem from the group 0 at age 67 definition, the τ_1 and τ_2 definitions, and the

		β_1	τ_1	$RM_{SS}(Group)$	β_2	τ_2	$RM_{SS}(Age)$
No Age Effect	G	++	0	1.44	0	0	1
	H	+	+	1.44	0	0	1
	I	0	++	1.44	0	0	1
Weak Age Effect	G	++	0	1.44	+	0	1.05
	H	+	+	1.44	$\sqrt{+}$	$\sqrt{+}$	1.05
	I	0	++	1.44	0	+	1.05
Strong Age Effect	G	++	0	1.44	++	0	1.1025
	H	+	+	1.44	+	+	1.1025
	I	0	++	1.44	0	++	1.1025

Table 4.9: Power setting patterns for an adjusting for a continuous covariate framework.

[1] 'G' has age effects only in the binomial portion of the model, 'H' has age effects in both parts of the model, and 'I' has age effects only in the continuous part of the model.

[2] '0' implies the parameter equals 0, '+' implies there is some positive effect, and '++' implies a stronger effect

age-by-group distribution for the data set being simulated, β_1 and β_2 settings were determined via a grid search such that the RM_{SS} metrics for group and age are as defined in Table 4.9.

4.8 Simulation Results for Comparing Two Groups When Adjusting for a Continuous Covariate

In this section, the Type 1 error and power of DM, RM_{SS}, and RM_{MAR} for find two group comparisons while adjusting for gender are compared to each other and to simple linear regression (SLR) and rank transformed regression (RTR). RTR is performed by transforming the outcome measure and the covariates into ranks and then performing simple linear regression using these ranks.Each group of settings will first be compared within each of DM, RM_{SS}, and RM_{MAR}. Then, for individual settings, the Type 1 error rates or powers levels will be compared among the five proposed tests.

4.8.1 Type 1 Error Under General Null Setting

Figures 4.21 - 4.25 display Type 1 error results for the general null settings outlined in Table 4.7. For these settings, there are no group effects in either the binomial or continuous parts of the model, but there are nonzero effects for the adjusting covariate age. Type 1 error rates are first compared across setting and within metric in Figures 4.21 - 4.23. Then Figures 4.24 and 4.25 compare Type 1 error rates for the various metrics; DM, RM_{SS}, RM_{MAR}, SLR, and RTR.

Figures 4.21, 4.22, and 4.23 show the Type 1 error rates for finding group differences using DM, RM_{SS}, and RM_{MAR} respectively. The labels in these figures correspond to the age effect settings outlined in Table 4.7 with 'G' indicating a nonzero age effect only in binomial part of the model, 'H' a nonzero effect of age only in the continuous part, and 'I' stemming from nonzero age effects relating to

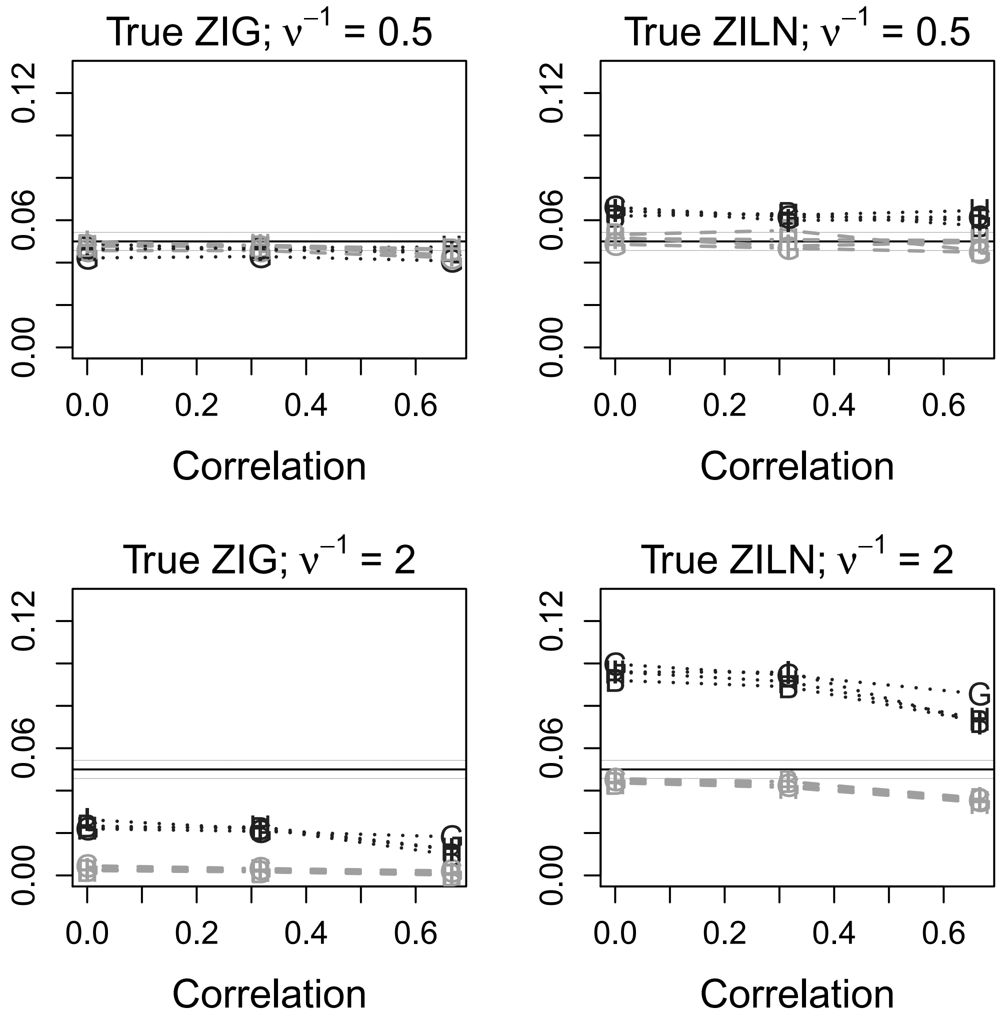

Figure 4.21: Type 1 error for tests based on DM; two group comparison adjusting for a continuous covariate. Symbols correspond to the settings labels in Table 4.7. The settings have no group effects, but for some settings there are significant adjusting (age) effects. Setting 'B' has no age effect, 'G' has an age effect only in the binomial portion of the model, 'H' has age effects in both parts of the model, and 'I' has an age effect only in the continuous part of the model. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

both zero-inflation and the effect on the continuous part of the model. In these figures, dark gray symbols with dotted lines represent analyses using ZIG regression, and light gray symbols with dashed lines show the results for analyses using ZILN regression. Within each figure, the plots on the left contain results for simulations where the data were simulated assuming ZIG and the plots on the right have the results for ZILN simulations. The top plots contain simulations where the coefficient of variation equals 0.5 and the bottom bottom plots show the results for simulations where the coefficient of variation equals 2. A solid black line shows the nominal level of 0.05 that was used for these analyses with dotted lines at 0.0457 and 0.0543 reflecting $\pm$ 1.96 standard errors of a binomial distributions when p=0.5 and 10,000 different data sets are run. Thus, any observed Type 1 error rates that are outside these bounds are different than would be expected if the Type 1 error were truly equal to the nominal 0.5 level.

The Type 1 error results for the DM test using the settings outlined in Table 4.7 are shown in Figure 4.21. The plots on the left side of the figure show that Type 1 error rates for DM when the data are simulated from ZIG distribution are all either low or within the appropriate Type 1 error range. When $\nu^{-1} = 0.5$, both ZIG and ZILN are within the appropriate range or only slightly conservative and there is very little difference between the Type 1 error rates at the various settings. When the data are simulated from a ZIG distribution and $\nu^{-1} = 2$, Type 1 error rates are extremely conservative especially when ZILN is assumed. However, Type 1 error rates do improve with sample size. There is very little difference in Type 1 error across adjusting effect setting. The largest differences are seen when age and group are highly correlated where Type 1 error is slightly higher for the setting 'G', where there is an effect of age in the binomial potion of the model, than it is for the other settings. When the data are simulated from a ZILN distribution, Type 1 error rates from ZIG analyses are elevated and increase with ν^{-1}. When $\nu^{-1} = 2$

and ZILN data are analyzed as ZIG, we can note some slight differences in the Type 1 error rates of the various settings. Setting 'B', with no nonzero adjusting effects, has the lowest Type 1 errors and setting 'G', with nonzero adjusting effects in the binomial part of the model, has the highest Type 1 error rates. When ZILN are appropriately analyzed as ZILN, Type 1 errors are appropriate when $\nu^{-1} = 0.5$ and slightly conservative when $\nu^{-1} = 2$. When $\nu^{-1} = 2$ both ZIG and ZILN analyses have Type 1 error rates that decline as the correlation between age and group increases.

Figure 4.22 shows the Type 1 error results for the RM_{SS} test given the settings outlined in Table 4.7. These RM_{SS} power results exhibit more differences between settings than were seen for DM. For both coefficient of variation levels and all true and assumed distribution combinations, setting 'G' has the lowest Type 1 error rates when there is no correlation between age and group. The Type 1 error rates for setting 'G' (where the only nonzero age effect is in the binomial part of the model) increase with the correlation between group and age. Because of this, when there is no correlation between age and group setting 'G' has the lowest Type 1 error rates. When the correlation is high, setting 'G' has Type 1 error rates that are higher than those of settings 'B' and 'I'. For setting 'H', where the effect of age is split between the continuous and binomial parts of the model, a similar but muted trend of increasing Type 1 with correlation is seen. The Type 1 error rates for settings with no binomial effects ('I' and 'B') stay relatively stable as correlation changes. Analyses assuming ZILN distributions tend to have Type 1 error rates closer to nominal than those assuming ZIG distributions. When the data truly are ZIG, ZIG analyses are slightly conservative and the ZILN analyses are nearer appropriate levels with Type 1 error rates for 'G' and 'H' slightly conservative at low correlations and slightly elevated at higher correlations. On the other hand when the data are truly ZILN, Type 1 error rates from ZILN analyses stay near nominal levels and Type 1 error rates from ZIG analyses are elevated.

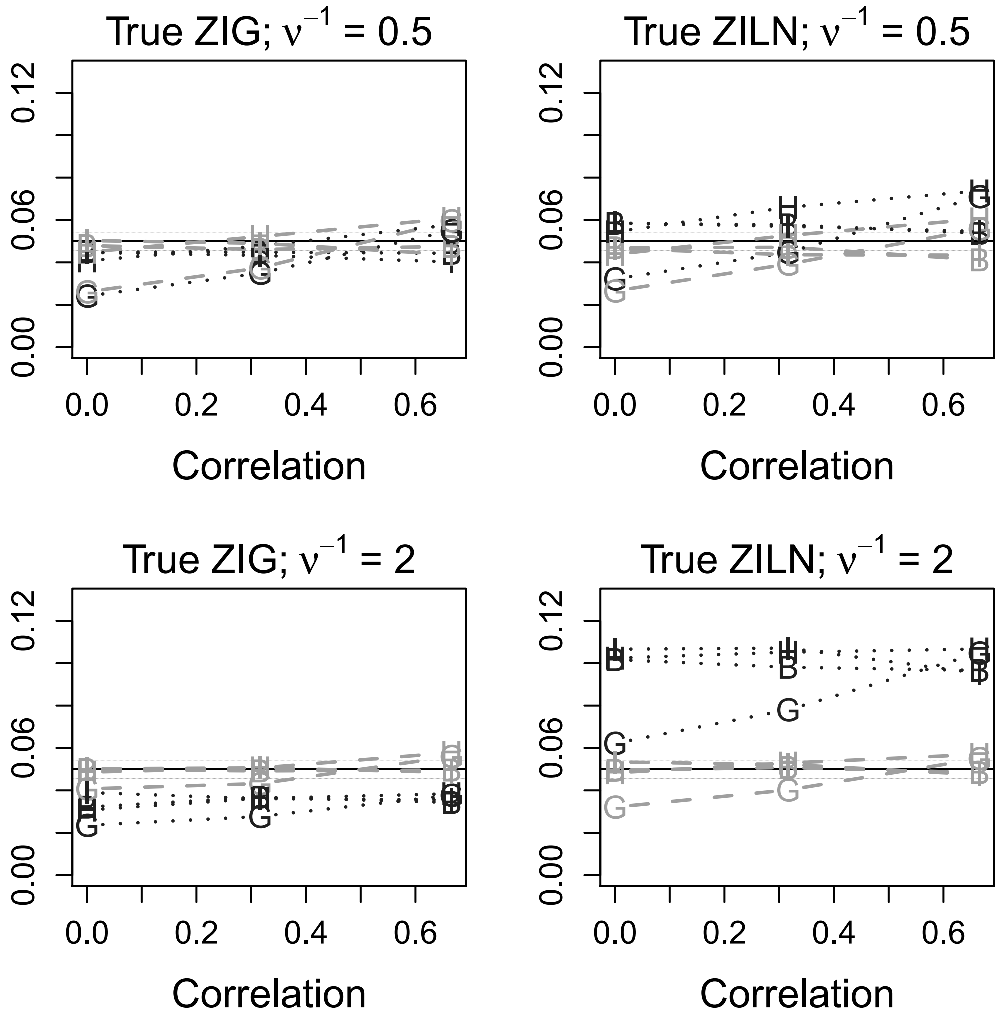

Figure 4.22: Type 1 error for tests based on RM_{SS}; two group comparison adjusting for a continuous covariate. Symbols correspond to the settings labels in Table 4.7. The settings have no group effects, but for some settings there are significant adjusting (age) effects. Setting 'B' has no age effect, 'G' has an age effect only in the binomial portion of the model, 'H' has age effects in both parts of the model, and 'I' has an age effect only in the continuous part of the model. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

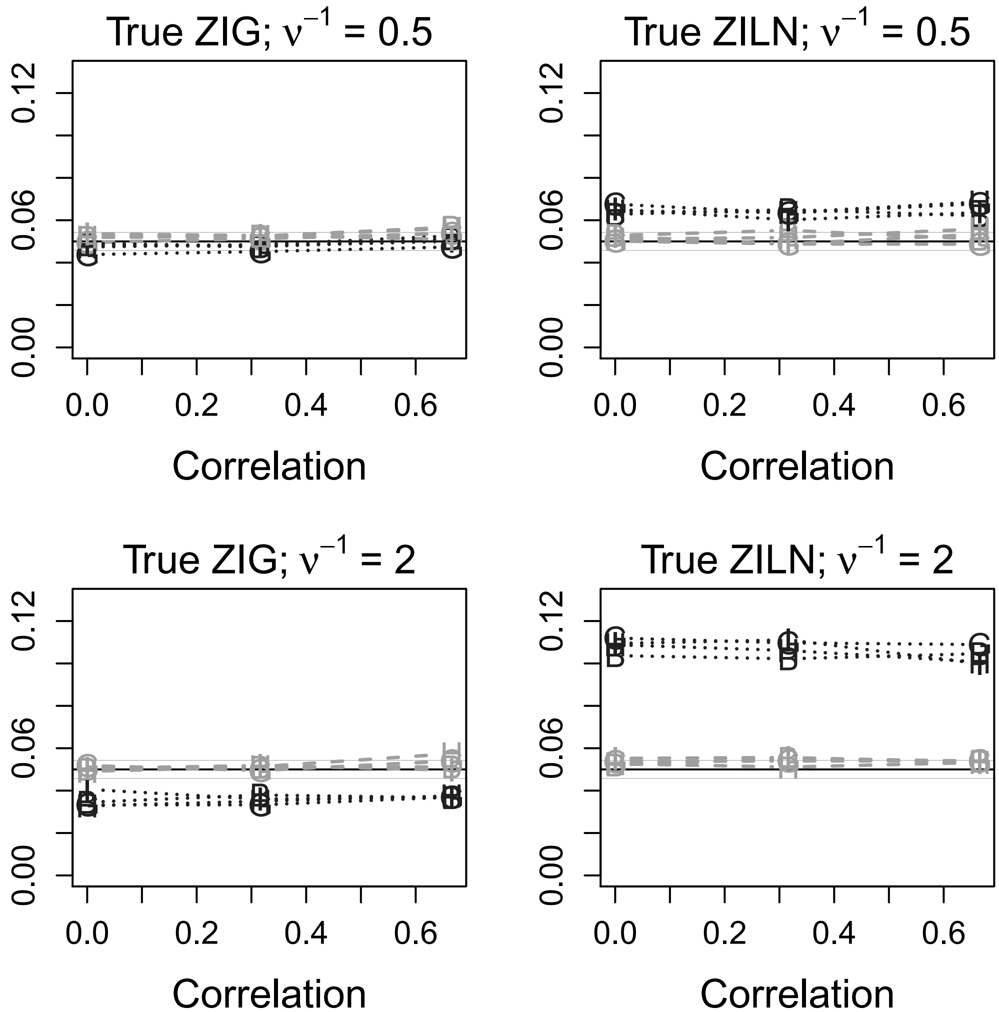

Figure 4.23: Type 1 error for tests based on RM_{MAR}; two group comparison adjusting for a continuous covariate. Symbols correspond to the settings labels in Table 4.7. The settings have no group effects, but for some settings there are significant adjusting (age) effects. Setting 'B' has no age effect, 'G' has an age effect only in the binomial portion of the model, 'H' has age effects in both parts of the model, and 'I' has an age effect only in the continuous part of the model. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

Figure 4.23 shows the Type 1 error rates for RM_{MAR} the adjusting for age scenario outlined in Table 4.7. When $\nu^{-1} = 0.5$, the patterns of Type 1 error rates seen for RM_{MAR} are very similar to those that were seen for DM Type 1 error. When $\nu^{-1} = 0.5$ and the data are simulated from a ZIG distribution, Type 1 error eRWA are near the appropriate range with only slightly elevated Type 1 error rates for misspecified ZILN analyses when the correlation between age and group is high and slightly lower than nominal Type 1 error rates for correctly specified ZIG analyses. When $\nu^{-1} = 0.5$ and ZILN is correctly specified, Type 1 error rates are at or near the appropriate levels. For $\nu^{-1} = 0.5$ and ZILN data misspecified as ZIG, Type 1 error rates are elevated. With the larger coefficient of variation ($\nu^{-1} = 2$) and for data simulated as ZIG, RM_{MAR} based tests have different patterns that those seen for DM. Specifically, when the data are truly ZIG Type 1 error rates for ZILN analyses are near appropriate levels and Type 1 error rates for ZIG analyses are slightly conservative. When ZILN data are appropriately analyzed as ZILN, Type 1 error rates remain near the appropriate levels. When ZILN data are incorrectly analyzed as ZIG Type 1 error rates are elevated.

Figure 4.24 and Figure 4.25 compare the Type 1 error rates for DM, RM_{SS}, RM_{MAR}, SLR, and RTR for the general null settings outlined in Table 4.7. Figure 4.24 shows the Type 1 error results when $\nu^{-1} = 0.5$ and Figure 4.25 when $\nu^{-1} = 0.5$. In these figures, the left plots include Type 1 error rates for data that was simulated from ZIG distributions and the right plots include Type 1 error rates for data that was simulated from ZILN distributions. The plot titles state the setting from Table 4.7 for which the Type 1 error rates are being plotted. Within the plots, the symbols and shades represent the test for which the results are being plotted. For the mean-based tests, DM, RM_{SS}, and RM_{MAR}, the results coming from ZIG analyses are shown in a dark gray with dotted lines with the results from the ZILN analyses are shown with light grey symbols and dashed lines. DM Type 1 error results are represented by the symbol 'D', RM_{SS} by the symbol 'S', and RM_{MAR} by the symbol 'M'. SLR Type 1 error results are show by the symbol 'L' and black dash-dot lines, and the RTR results are show with the symbol 'R' and long dashes that are at a grey in between those used for ZIG and ZILN results.

In Figure 4.24 it can be seen that the relative values of the Type 1 error rates in comparing one test to another is not constant across adjusting effect setting. As such, each graph within the figure will be discussed separately. All simulations presented within Figure 4.24 are for $\nu^{-1} = 0.5$.

When the data are truly ZIG and the age effect is is the binomial part of the two-part model, setting 'G', RM_{SS} has the lowest, most conservative Type 1 error rates when the correlation between age and group is zero or 0.3162 but has higher Type 1 error rates than DM or RM_{MAR} when the correlation is high (0.6656) with the RM_{SS} test under a misspecified ZILN model being slightly above the acceptable Type 1 error range. In this same setting, SLR and RTR have appropriate Type 1 error rates at low correlations between age and group but elevated Type 1 error rates at a high correlation between age and group with RTR having the most elevated Type 1 error rate. The plot on the top right also shows setting 'G' but simulated under ZILN distributions. Here correctly specified ZILN analyses using DM and RM_{MAR} have appropriate Type 1 error rates and incorrectly specified ZIG analyses for DM and RM_{MAR} have elevated Type 1 error rates. For RM_{SS}, Type 1 error rates again increase with the correlation between age and group being conservative at low levels of correlation and elevated (particularly for the ZIG analysis) when the correlation is high. Also, in a similar fashion to the results seen for the ZIG data, SLR and RTR under setting 'G' when the true distribution is ZILN have appropriate Type 1 error rates when the correlation between group and age are low, but elevated Type 1 error rates when the correlation is high, with RTR having the most elevated Type 1 error rates.

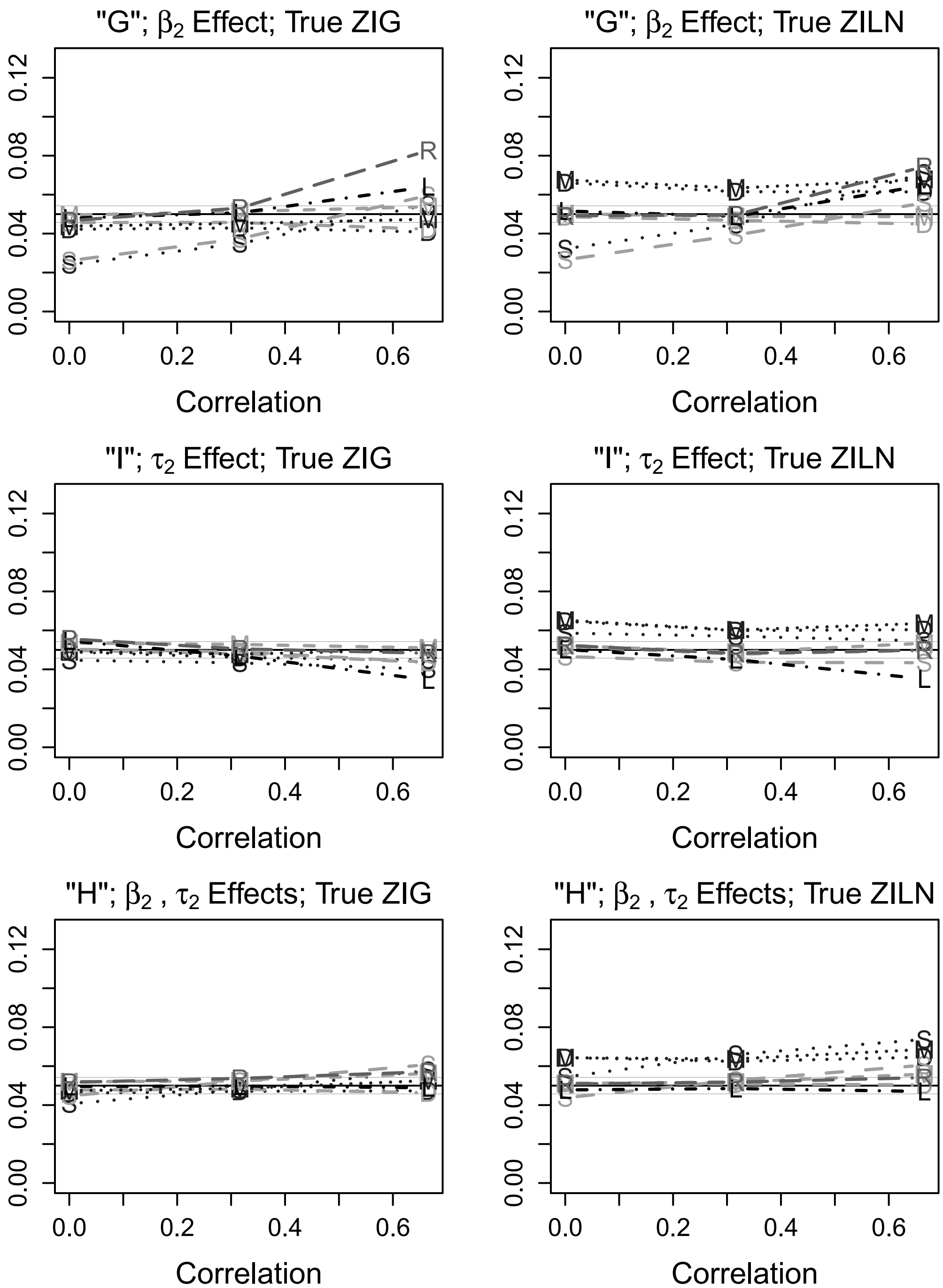

Figure 4.24: Comparison of Type 1 error for mean-based tests and SLR and RTR; $\nu^{-1} = 0.5$; two group comparison adjusting for a continuous covariate. 'D' represents the Type 1 error rates for tests based on DM, 'S' those based on RM_{SS}, and 'M' those based on RM_{MAR} with darker grey symbols with dotted lines for ZIG analyses and lighter grey symbols with dashed lines for ZILN analyses. SLR analysis results are shown with the symbol 'L' and black dash-dot lines; RTR with the symbol 'R' and medium grey lines with long dashes.

For setting ‘I’, where the effect of age is in the continuous part of the model only, when the true distribution is ZIG the differences between the Type 1 error rates for the various tests are slight. Most of the tests have close to nominal Type 1 error rates, though RM_{SS} assuming ZIG has slightly conservative Type 1 errors at all levels of correlation and SLR has slightly conservative Type 1 error rates at high correlation. For setting ‘I’ when the true distribution is ZILN, as is usually the case when the data are analyzed as ZIG the Type 1 error rates are elevated while the analyses assuming ZILN are closer to nominal. Within distributional assumption, Type 1 error rates for RM_{SS} are slightly lower than those for DM and RM_{MAR}. For truly ZILN data under setting ‘I’, RTR regression yields appropriate Type 1 error rates, and SLR demonstrates Type 1 error rates that are near nominal when there is no correlation between age and group and then decrease as the correlation between age and group increases.

The bottom plots show setting ‘H’ where the effect of age is present in both parts of the model. The overall patterns of the relationships between the various tests are similar to those seen for setting ‘G’, but somewhat muted. As such, most of the Type 1 error results for truly ZIG data are within the appropriate bounds, but since the Type 1 error rate for RM_{SS} increases with the correlation between age and group, Type 1 error rates for ZIG RM_{SS} are slightly conservative at the correlation of zero and Type error rates for ZILN RM_{SS} are slightly elevated for the correlation of 0.6656. When the data are truly ZILN, Type 1 error rates for DM and RM_{MAR} assuming ZILN are at appropriate levels, but are elevated when assuming ZIG. SLR and RTR are within appropriate range. Type 1 error rates for RM_{SS} increase with correlation with ZIG analyses have higher Type 1 error rates than ZILN analyses; this leads to ZILN RM_{SS} Type 1 error rates being conservative at correlation equaling zero while ZIG analyses at that level have appropriate Type 1 error rates while at the highest correlation RM_{SS} tests have elevated Type 1 error rates for both ZIG and ZILN analyses with ZIG RM_{SS} tests have the highest Type 1 error rates of all of the tests at the highest correlation level.

Figure 4.25 compares the Type 1 error rates for DM, RM_{SS}, RM_{MAR}, SLR, and RTR when $\nu^{-1} = 2$. Type 1 error rates for these tests under this higher coefficient of variation differ to a larger extent than they did under the lower values of coefficient of variation. When the data are truly ZIG, ZILN DM tests have the lowest Type 1 error rates with all ZILN DM Type 1 error rates close to zero. DM tests correctly assuming ZIG had the next lowest Type 1 error rates followed by ZIG RM_{SS} and ZIG RM_{MAR}. When the correlation between age and group is low and the effect of age is only in the binomial part of the model, RM_{SS} has lower Type 1 error rates than RM_{MAR} otherwise the Type 1 error rates for these tests when correctly assuming ZIG are very close to each other. ZILN DM, RM_{SS}, and RM_{MAR} Type 1 errors as well as SLR and RTR Type 1 errors were typically at appropriate levels with a few exceptions. For setting 'G', ZILN RM_{SS} analyses were slightly conservative when the correlation between group and age was small and slightly elevated at the higher correlation; the Type 1 error rate for RTR was also slightly elevated at the higher correlation of group and age. For setting 'H', both incorrectly specified ZILN RM_{SS} and ZILN RM_{MAR} were slightly elevated at the higher correlation level. SLR was slightly conservative for setting 'I' with high correlation and setting 'H' at all levels of correlation.

The plots on the right compare the Type 1 error rates for these same tests and setting but when the data are simulated from ZILN distributions. For ZILN data simulated under setting 'G', the worst test in terms of elevated Type 1 error rates are the ZIG RM_{MAR} analyses. ZIG DM also has very elevated Type 1 error rates. Type 1 error rates for ZIG RM_{SS} are all elevated, but worsen with the increase of the correlation between group and gender having lower Type 1 error rates than DM at lower levels of correlation and a higher rate at the highest correlation. RTR has

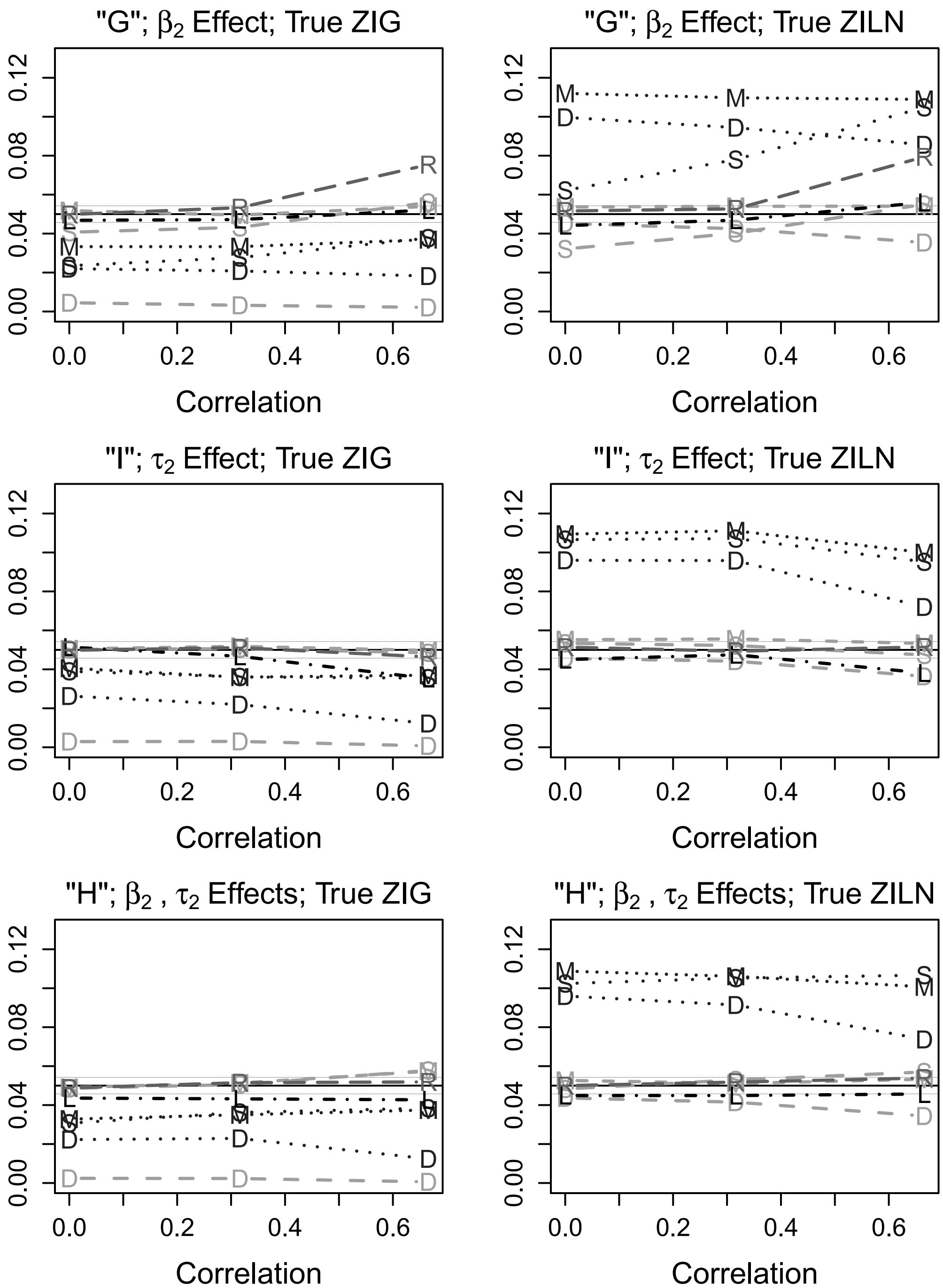

Figure 4.25: Comparison of Type 1 error for mean-based tests, SLR, and RTR; $\nu^{-1} = 2$; two group comparison adjusting for a continuous covariate. 'D' represents the Type 1 error rates for tests based on DM, 'S' those based on RM_{SS}, and 'M' those based on RM_{MAR} with darker grey symbols with dotted lines for ZIG analyses and lighter grey symbols with dashed lines for ZILN analyses. SLR analysis results are shown with the symbol 'L' and black dash-dot lines; RTR with the symbol 'R' and medium grey lines with long dashes.

appropriate Type 1 error rates for low level of correlation, but elevated rates for the highest correlation. SLR-based and ZILN-based analyses have near appropriate Type 1 error rates, with ZILN RM_{SS} having a slightly conservative Type 1 error rate when there is no correlation between age and group and ZILN DM having a slightly conservative Type 1 error rate when the correlation between age and group is high. For settings 'I' and 'H', ZIG RM_{MAR} and ZIG RM_{SS} have the highest Type 1 error rates followed by ZIG DM. For these settings ZILN RM_{SS} and RM_{MAR} have appropriate Type 1 error rates and ZILN DM has appropriate Type 1 error rates when then is no correlation between group and age but slightly conservative Type 1 error rates for higher levels of correlation. For ZILN data simulated under settings 'I' and 'H', RTR has appropriate Type 1 error rates and SLR has appropriate to slightly conservative Type 1 error rates.

4.8.2 Simulation Results Under Metric Based Null Settings

Type 1 Error and Power for Dissonant Effects with Equal Group Means

In this section, Type 1 error results are presented for settings with dissonant effects which cancel out in terms of setting the metric of interest at the null hypothesis. The structure for the settings used here is outlined in Table 4.8. Slightly different setting are required to have dissonant effects that cancel out such that $RM_{SS} = 1$ than are required for $DM = 0$ and $RM_{MAR} = 1$. As such, DM, RM_{SS}, and RM_{MAR} Type 1 error results are shown using the settings for which the metric is at its own null. SLR results are shown at the marginal mean based null used for RM_{MAR} and DM. Figures 4.26, 4.27, and 4.28 show the Type 1 error results for DM, RM_{SS}, and RM_{MAR} respectively for the dissonant null setting in Table 4.8. Figure 4.30 compares the Type 1 error results for DM, RM_{SS}, RM_{MAR}, and SLR for setting 'N' from Table 4.8. Figure 4.31 compares the Type 1 error results for DM, RM_{SS},

RM_{MAR}, and SLR for setting 'I' from Table 4.8.

In Figures 4.26 - 4.28, the top plots include simulations where $\nu^{-1} = 0.5$, the bottom plots include simulations where $\nu^{-1} = 2$. The plots on the left show Type 1 error rates for data simulated as ZIG. And the plots on the right show Type 1 error rates for data simulated as ZILN. Labels used indicate the setting used. As outlined in Table 4.8, 'N' is used for settings with no nonzero adjusting effects, 'G' for settings with a nonzero adjusting effect in the binomial part of the model, 'I' for settings with a nonzero adjusting effect in the continuous part of the model, and 'H' for settings with nonzero adjusting effects in both parts of the model. Dark gray symbols and dotted lines are used for analyses assuming ZIG and light gray symbols and dashed lines are used for analyses assuming ZILN.

Figure 4.26 shows the Type 1 error rates for DM that result from using the metric-based null settings outlined in Table 4.8. The top left graph shows that when the data are simulated from a ZIG distribution and $\nu^{-1} = 0.5$, Type 1 error rates for DM based tests are within the appropriate range regardless of the effect of the adjusting covariate, age. When the data are simulated from a ZIG distribution and $\nu^{-1} = 2$, incorrectly assuming ZILN leads to Type 1 error rates near zero. Correctly assuming ZIG yields Type 1 error rates that are conservative, albeit higher than those seen for ZILN analyses. Within analysis assumption, Type 1 error rates for the different adjusting effect settings explored are very similar. When data are simulated as ZILN and $\nu^{-1} = 0.5$, Type 1 error rates for DM analyses when assuming ZILN are within the appropriate range while Type 1 error rates for DM assuming ZIG are slightly elevated. Presence or absence of adjusting covariate effects has little influence on Type 1 error rates. When $\nu^{-1} = 2$ and the data are simulated from ZILN distributions, Type 1 error rates for analyses assuming ZILN are slightly conservative. Type 1 error rates for DM incorrectly assuming ZIG are elevated. The adjusting effect setting does influence Type 1 error rates for misspecified ZIG analyses with setting ‘G’, where there is an effect for age in the binomial part of the model, having the highest Type 1 error rates. This may be because the positive effect of age used in the binomial part of the model will decrease the amount of zero inflation in the sample. As was seen in previous sections, lower zero inflation rates lead to higher Type 1 error rates for misspecified ZIG analyses.

Figure 4.27 shows Type 1 error rates for RM_{SS} based tests at settings where $RM_{SS} = 1$ as outlined in Table 4.8. When data are simulated as ZIG, $\nu^{-1} = 0.5$, and there is no correlation between age and group, all Type 1 error rates are within the acceptable range. When the correlation is high there are differences in Type 1 error across settings. Setting ‘G’, where there is an age effect in the binomial part of the model, has slightly elevated Type 1 error rates for analyses using RM_{SS} under both ZIG and ZILN distributional assumptions. For setting ’N’, where there are no nonzero age effects, the Type 1 error rate when assuming ZIG is slightly conservative. For data simulated as ZIG with $\nu^{-1} = 2$, Type 1 error rates assuming ZILN are higher than those assuming ZIG. With the high correlation between age and group when correctly assuming ZILN, Type 1 error rates for settings that include a positive effect of age in the binomial part of the model (‘G’ and ‘H’) are slightly elevated. Analyses incorrectly assuming ZIG have conservative Type 1 error rates. For analyses where data are simulated as ZILN and $\nu^{-1} = 0.5$, within setting Type 1 error rates for analyses assuming ZIG are higher than those for analyses assuming ZILN. For settings where the probability of a nonzero outcome increases with age (‘G’, and ‘H’), Type 1 error increases with the correlation between age and group. For ZILN data with $\nu^{-1} = 2$, Type 1 error rates for ZIG analyses are high. Setting ‘G’ again has the highest Type 1 error rates. For analyses correctly assuming ZILN Type 1 error rates are within range except for setting ‘G’ which is slightly high.

Type 1 error rates for RM_{MAR} based tests under the metric-based settings from Table 4.8 with $RM_{MAR} = 1$ are shown in Figure 4.28. When the data are simulated

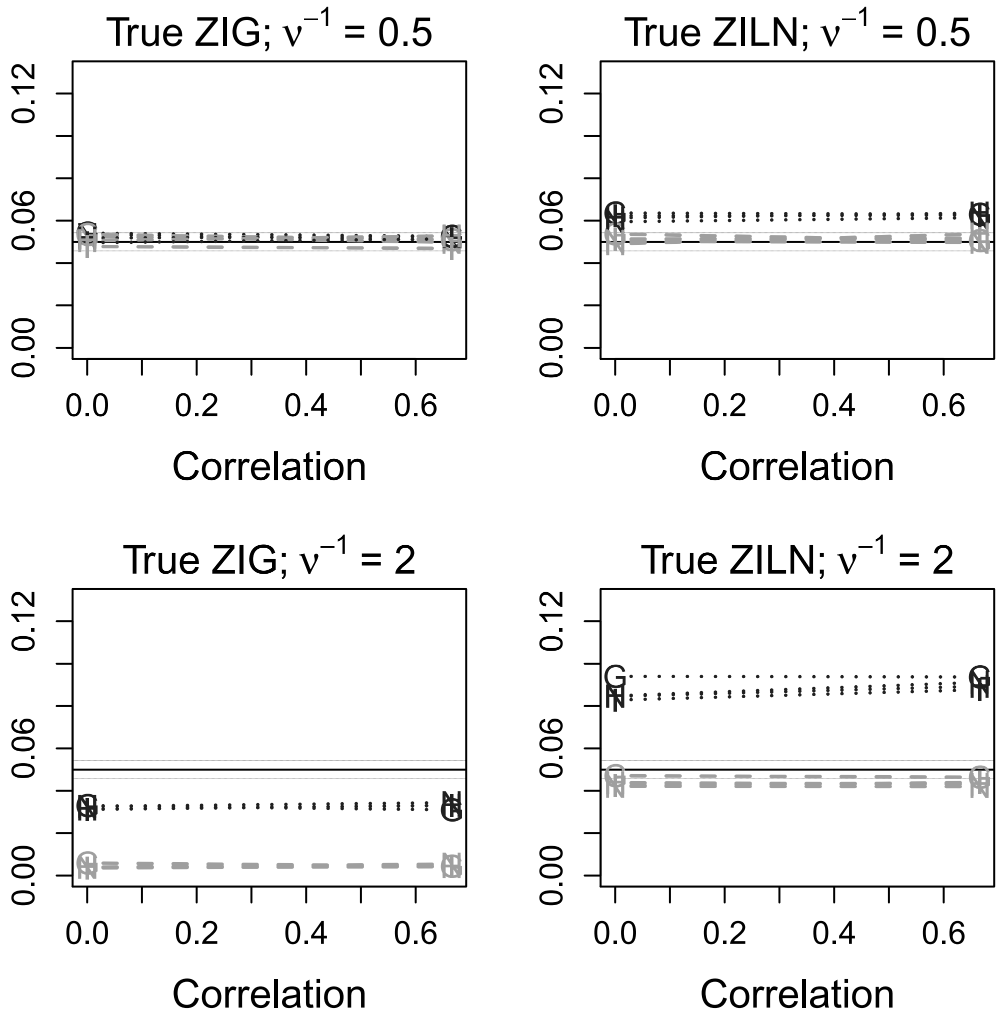

Figure 4.26: Type 1 error for DM at metric-based null adjusting for a continuous covariate. Labels refer to settings outlined in Table 4.8 and darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

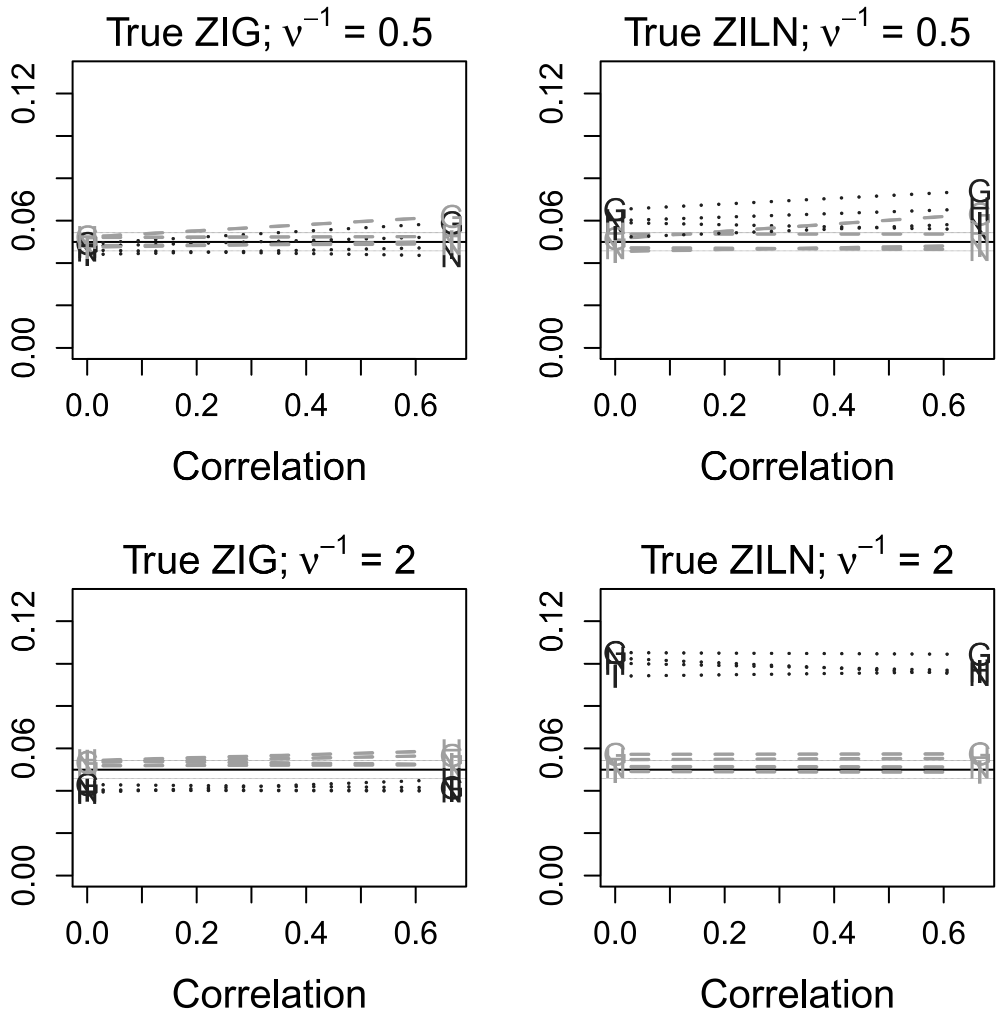

Figure 4.27: Type 1 error for RM_{SS} at metric-based null adjusting for a continuous covariate. Labels refer to settings outlined in Table 4.8 and darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

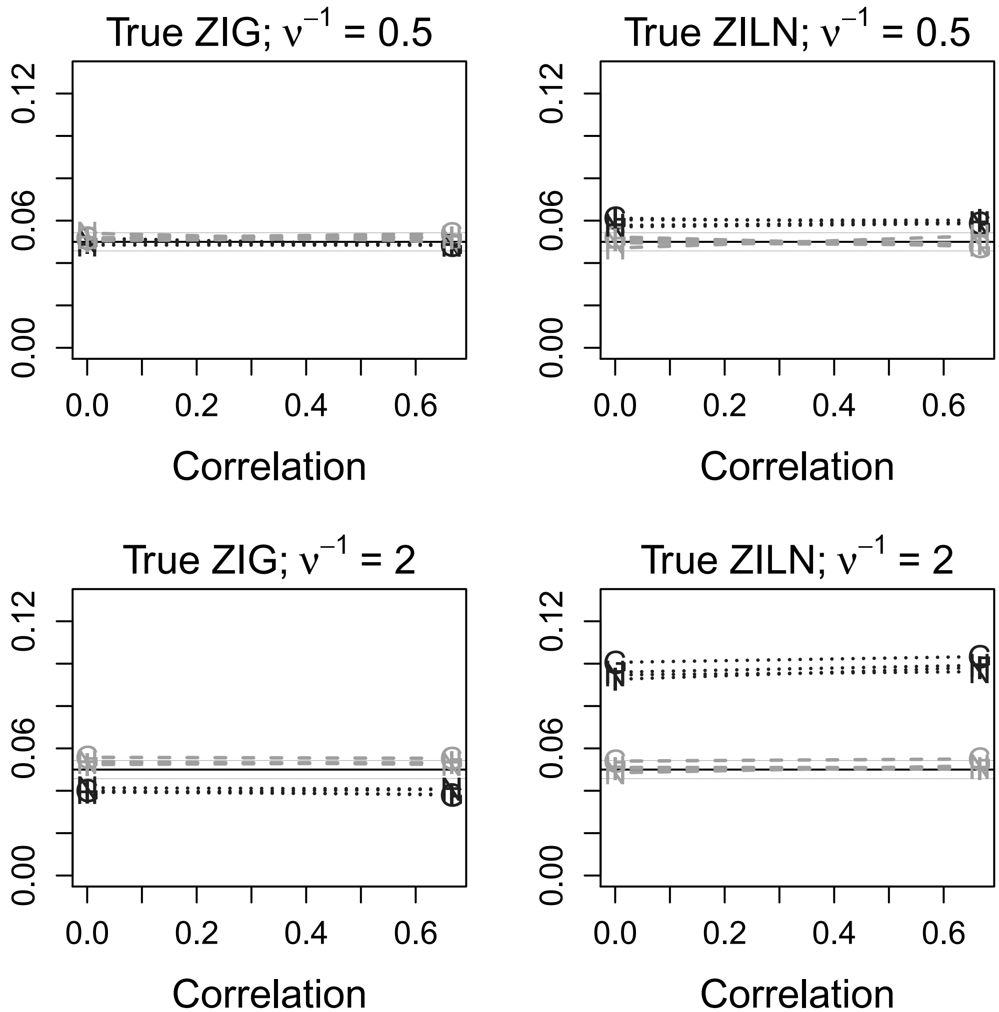

Figure 4.28: Type 1 error for RM_{MAR} at metric-based null adjusting for a continuous covariate. Labels refer to settings outlined in Table 4.8 and darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

from ZIG distributions and $\nu^{-1} = 0.5$, Type 1 error rates are within the appropriate range. For ZIG data with $\nu^{-1} = 2$, Type 1 error rates when ZILN is incorrectly assumed are slightly high for setting 'G' and within range for the other settings. When ZIG is correctly assumed, Type 1 error rates are slightly conservative. When the data are simulated from ZILN distributions with $\nu^{-1} = 0.5$, RM_{MAR} assuming ZILN has appropriate Type 1 error rates. Assuming ZIG results in slightly elevated Type 1 error rates. For ZILN data with $\nu^{-1} = 2$, Type 1 error for $RM_{MAR} = 1$ analyses assuming ZILN is slightly elevated for setting 'G' when the correlation between age and group is high and appropriate otherwise. When ZIG is incorrectly assumed, Type 1 error rates are elevated with the highest Type 1 error rates resulting from simulations using setting 'G'.

Settings designed to be at the metric-based null for RM_{SS}, or for DM and RM_{MAR}, do not necessarily result in null settings in terms of ranks. When one group has more nonzero values and the other has higher nonzero values, the marginal means may be the same, but the distributions of the ranks will be different. As such the results shown in Figure 4.29 show the power to find such a difference in rank distributions when the data are simulated from a marginal mean based null framework. This figure shows that the power to find such a difference is higher when the data come from a ZIG distribution than from a ZILN distribution. Also, RTR power increases as ν^{-1} increases. Setting 'N', where there are no nonzero effects of the adjusting covariate, has the lowest power levels. Settings 'G', 'H', and 'I' each have scenarios where they have slightly higher power levels than the other settings, with setting 'G' being slightly higher than the other scenarios at a greater frequency.

Figures 4.30 and 4.31 compare Type 1 error rates across metric and within setting. Because the relative values of Type 1 error rates for the various metrics have few differences across settings only two of the four settings are shown. For all settings when $\nu^{-1} = 2$ and the data are simulated as ZIG, the difference between

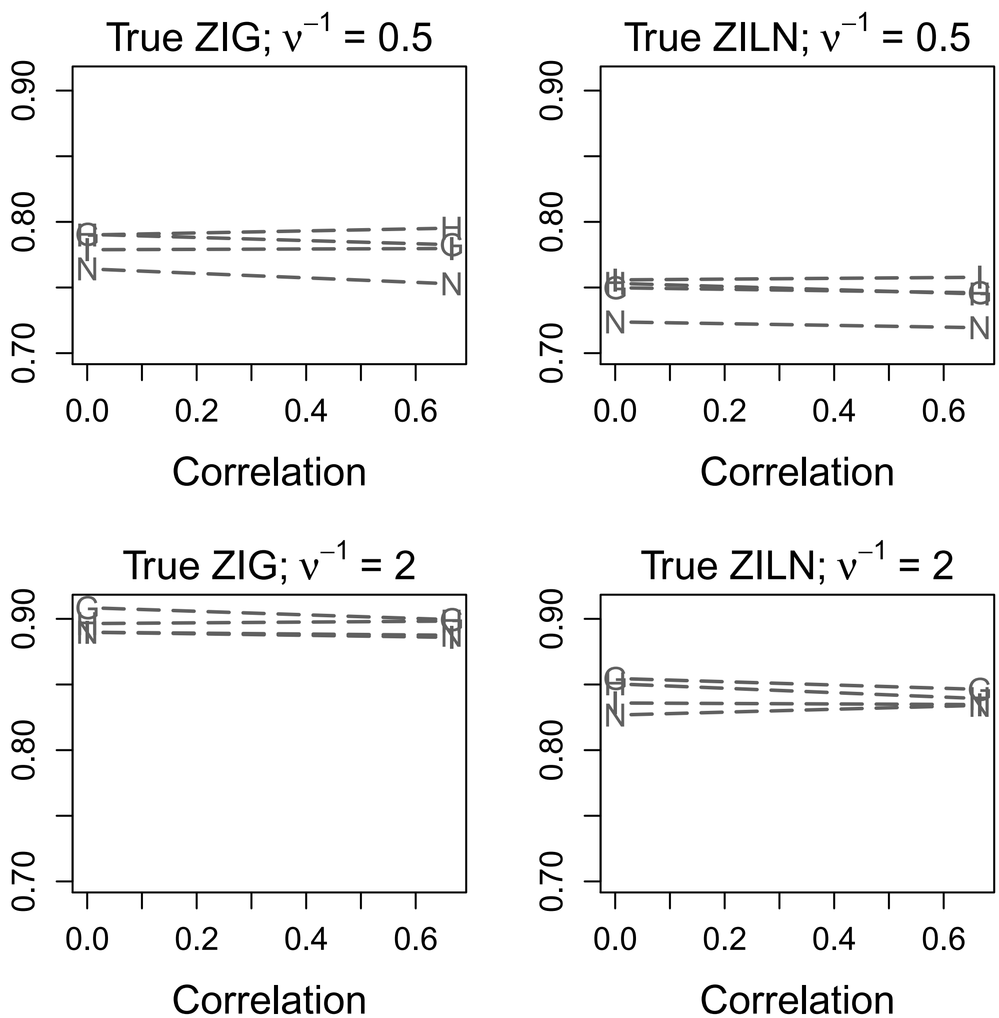

Figure 4.29: Power for Rank Transform Regression at marginal mean based null adjusting for a continuous covariate. Labels refer to settings outlined in Table 4.8.

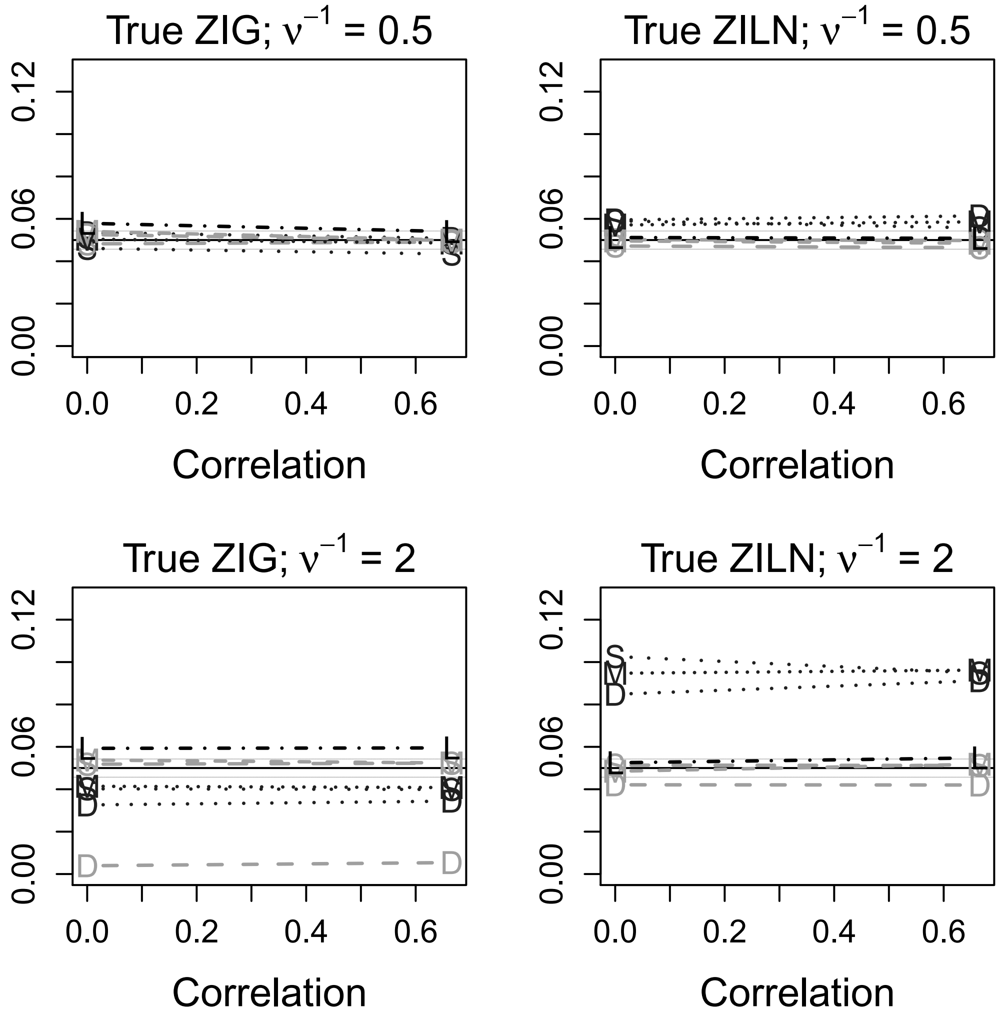

Figure 4.30: Type 1 error for all mean-based tests at metric-based null adjusting for a continuous covariate; setting 'N' from Table 4.8. 'D' represents the Type 1 error rates for tests based on DM, 'S' those based on RM_{SS}, and 'M' those based on RM_{MAR} with darker grey symbols with dotted lines for ZIG analyses and lighter grey symbols with dashed lines for ZILN analyses. SLR analysis results are shown with the symbol 'L' and black dash-dot lines.

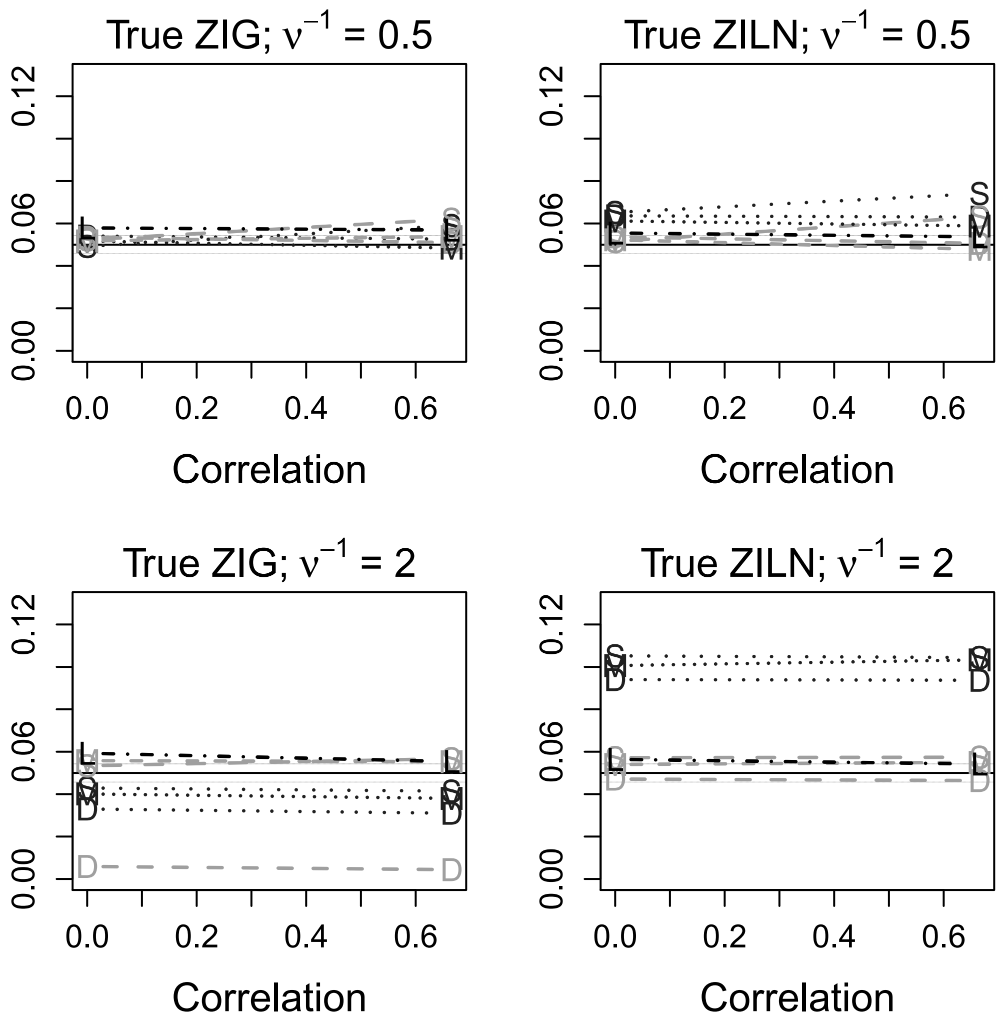

Figure 4.31: Type 1 error for all mean-based tests at metric-based null adjusting for a continuous covariate; setting 'I' from Table 4.8. 'D' represents the Type 1 error rates for tests based on DM, 'S' those based on RM_{SS}, and 'M' those based on RM_{MAR} with darker grey symbols with dotted lines for ZIG analyses and lighter grey symbols with dashed lines for ZILN analyses. SLR analysis results are shown with the symbol 'L' and black dash-dot lines.

Type 1 error rates among the metrics are slight. For most settings, SLR has slightly elevated Type 1 error rates. Misspecified ZILN analyses typically have higher Type 1 error rates than do correctly specified ZIG analyses, although both are typically close the the appropriate range. RM_{SS} varies the most with setting. When ZIG data are analyzed as ZIG, RM_{SS} tests yield lower Type 1 error rates than the other tests for settings 'N' and 'I'. RM_{SS} tests assuming both ZIG and ZILN have higher Type 1 error rates than the other tests for setting 'G' when the correlation between group and age is high. The largest distinction between tests occurs when data are simulated from ZIG distributions and $\nu^{-1} = 2$. Within these settings, the lowest Type 1 error rates come from DM tests assuming ZILN followed by DM tests assuming ZIG. RM_{SS} and RM_{MAR} tests when correctly assuming ZIG have the next lowest Type 1 error rates and these are still slightly conservative, followed by RM_{SS} and RM_{MAR} when misspecified as ZILN which results are within range to slightly elevated. SLR typically has slightly elevated Type 1 error rates for ZIG simulated data with $\nu^{-1} = 2$; in some settings these Type 1 error rates are the highest of all of the tests, in other settings the Type 1 error rates for RM_{SS} and RM_{MAR} assuming ZILN as similar to those seen for SLR.

When the data are truly ZILN and ZIG is assumed, Type 1 error rates are typically high, while when ZILN is assumed, Type 1 error rates are typically appropriate. Within distributional assumption, the differences in Type 1 error rates are smaller when ZILN is the true distribution. When the data are from a ZILN distribution and $\nu^{-1} = 0.5$, for settings with adjusting effects equal to zero, DM assuming ZIG has slightly higher Type 1 error rates than do the other tests (see Figure 4.30) while for setting 'I' (Figure 4.31) where there is an effect of age in the continuous part of the model, RM_{SS} assuming ZIG has the highest Type 1 error rates especially when the correlation between group and age is high. Under the scenario of ZILN data with setting 'I' and high correlation, RM_{SS} correctly assuming ZILN also has elevated Type 1 error rates. When the data are truly ZILN and $\nu^{-1} = 2$, the Type 1 error rates observed when a ZIG distribution is assumed are greatly elevated. Among analyses incorrectly assuming ZIG, DM yields the lowest Type 1 error rates for all settings. RM_{SS} assuming ZIG has higher Type 1 error rates than RM_{MAR} when there is no correlation between age and group, and similar Type 1 error rates when the correlation is high. ZILN analyses have near appropriate Type 1 error rates for most settings and tests, but DM typically has slightly conservative rates, RM_{MAR} Type 1 error rates are in the appropriate range, and RM_{SS} is either within range similar to RM_{MAR} or slightly elevated. SLR Type 1 error rates are typically slightly elevated or just within the acceptable range.

4.8.3 Power for Consonant Effects

Power results for the settings outlined in Table 4.9 are presented in Figures 4.32 - 4.41. Within each figure the plots on the left show the power levels for ZIG simulated data and those on the right show the power levels for ZILN simulated data. From the top plots to the bottom plots, the organization is same as set out in Table 4.9. This organization is related to the effects included when the data were simulated. All analyses include age in the model, regardless of whether there are true age effects or not. The top plots contain power results for settings where there are no age effects. The middle plots include power results for settings with weak age effects. And the bottom plots show power results for settings with strong age effects. Within each plot, settings labeled 'G' include nonzero effects in the binomial part of the model only, settings labeled 'H' include nonzero effects in both parts of the model, and settings labeled 'I' include effects in only the continuous part of the model. Dark gray symbols with dotted lines are used to show results for ZIG analyses and light gray symbols with dashed lines are used to show results for ZILN analyses.

Figure 4.32 shows the power levels for settings from Table 4.9 with $\nu^{-1} = 0.5$

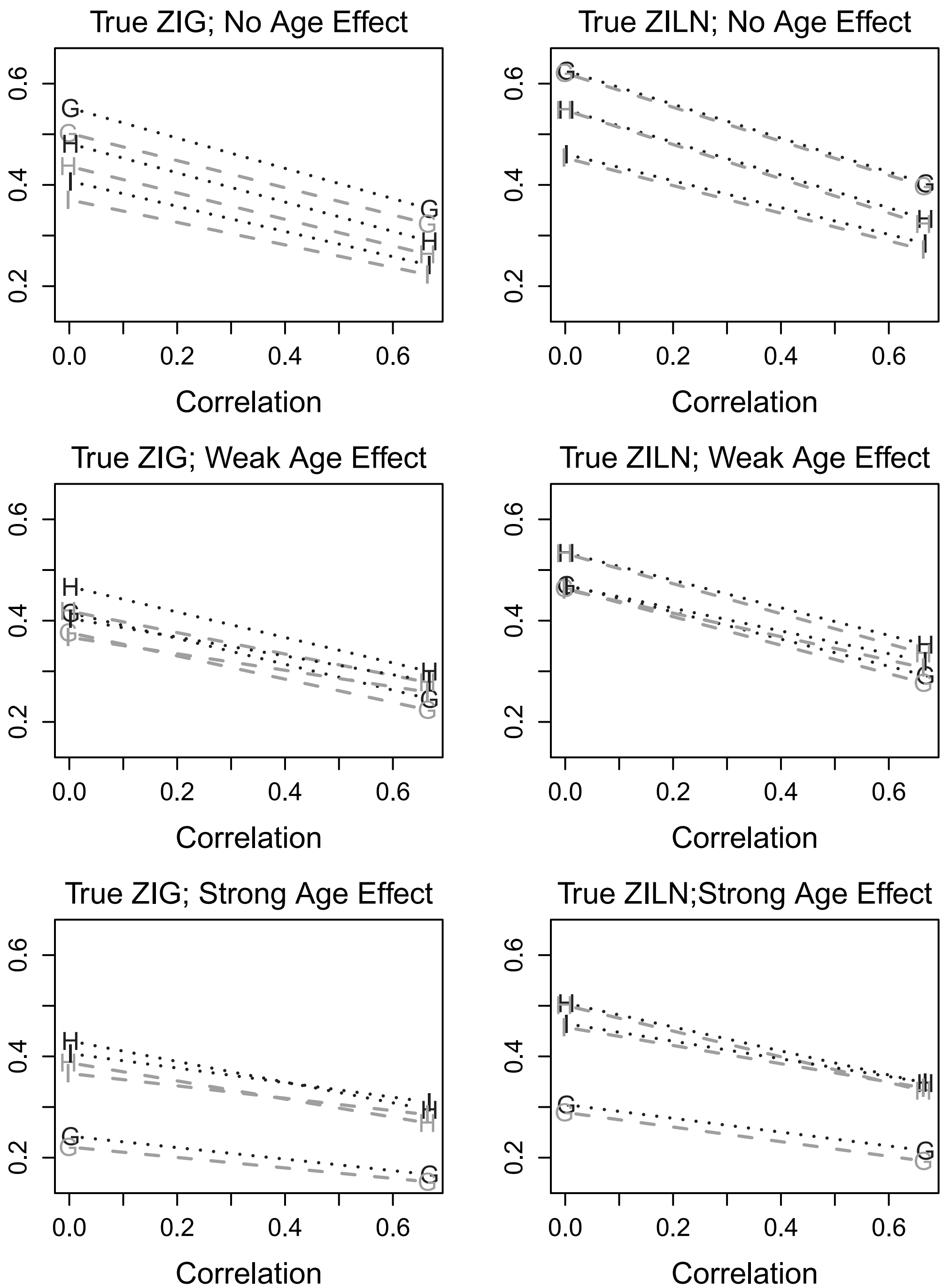

Figure 4.32: Power for DM with a continuous adjusting covariate and $\nu^{-1} = 0.5$. Graphs are labeled with the strength of the adjusting age effect. Within the graphs, setting labels correspond to those outlined in Table 4.9; 'G' represents effects only in the binomial part of the model, 'I' effect only in the continuous part of the model, and 'H' effects which are split between the two parts. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

for DM based tests. Power decreases with the correlation between age and gender. The strength of age effects influences the relative power levels among the three setting types. When all age effects equal zero (top plots), power increases with the effect of group in the binomial part of the model; setting 'G', followed by setting 'H', followed by setting 'I'. However, when age effects are present, setting 'H' which has effects in both the continuous and the binomial parts of the model has the highest power levels. When there are strong age effects used in the simulation, power levels for setting 'G' are considerably lower than those for 'H' or 'I'. For data that are simulated from ZIG distributions, analyses assuming ZIG have greater power levels than those assuming ZILN. When data are simulated from ZILN distributions, analyses assuming ZIG still have higher power levels, but the difference is slight.

Power levels for DM based tests when $\nu^{-1} = 2$ are shown in Figure 4.33. With this increase in ν^{-1} power levels decreased considerably for all settings and distributional assumptions. With no age effect and weak age effects, within setting power decreased with an increase in correlation between age and group. With a strong age effect, this is still the case when the data are simulated from ZILN distributions, but when data are simulated from ZIG distributions there is little change in power for different levels of correlation. For data simulated from ZIG distributions, when the analyses incorrectly assume ZILN, power levels are extremely low. For correctly specified ZIG analyses when there is no age effect, power increases as the effects in the binomial portion of the model increase and the effects in the continuous part of the model decrease such that setting 'G' has the highest power and setting 'I' has the lowest power. For weak age effects, the settings have similar levels of power. And for strong age effects setting 'G' has the lowest power.

When the data are simulated from ZILN distributions, power levels are higher than they are for data simulated from ZIG distributions. Power levels for incorrectly specified ZIG analyses are higher than those for correctly specified ZILN analyses.

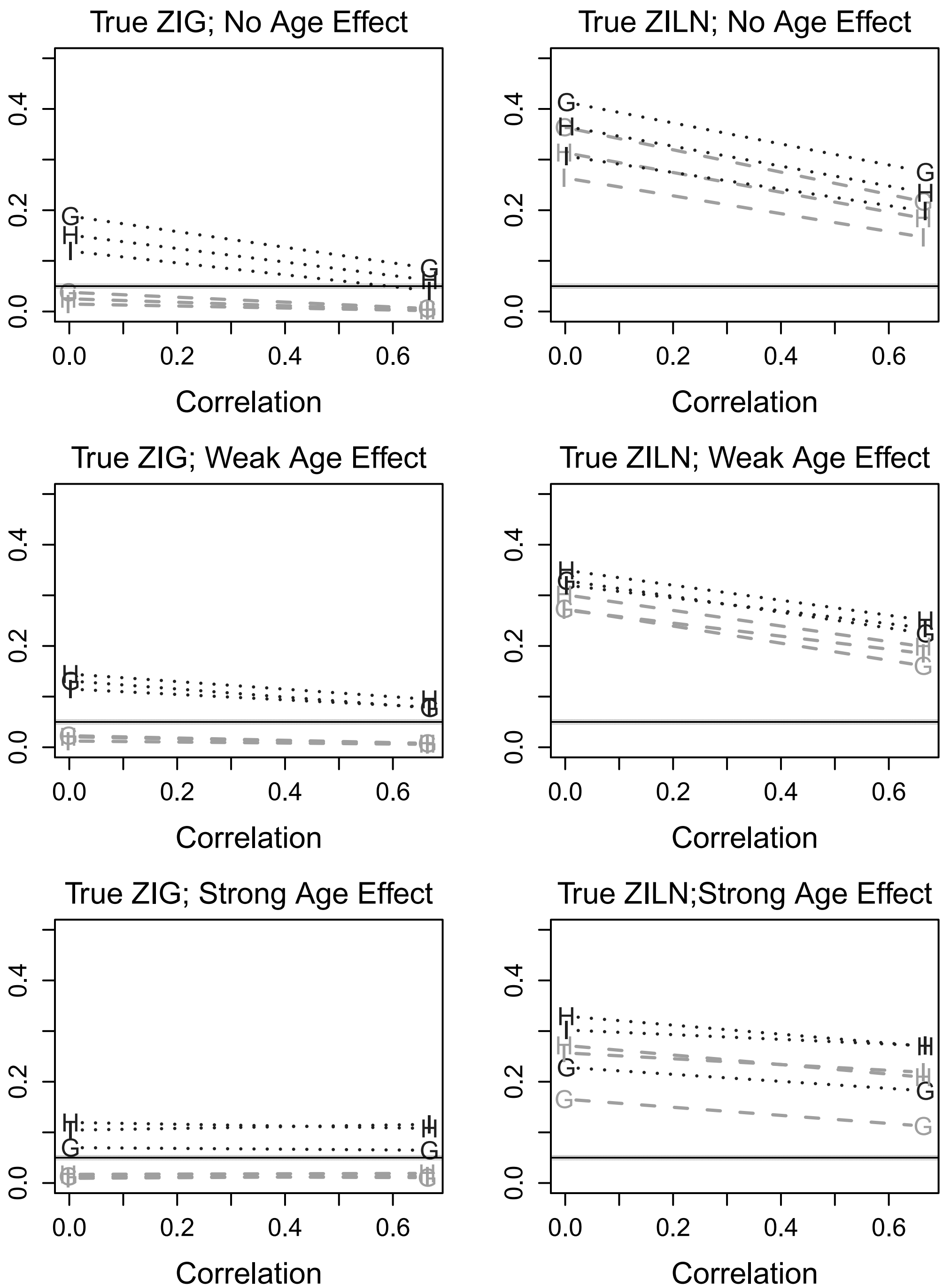

Figure 4.33: Power for DM with a continuous adjusting covariate and $\nu^{-1} = 2$. Graphs are labeled with the strength of the adjusting age effect. Within the graphs, setting labels correspond to those outlined in Table 4.9; 'G' represents effects only in the binomial part of the model, 'I' effect only in the continuous part of the model, and 'H' effects which are split between the two parts. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

However, when $\nu^{-1} = 2$, Type 1 error rates were so high that the incorrectly specified ZIG analyses should be discounted. As was the case for correctly specified ZIG analyses, the effect on power of the location of the group effects changes with the strength of the age effects. When there is no age effect (with age still included in the model), setting 'G' with group effects only in the binomial portion of the model has the highest power whereas when there are age effects, setting 'G' has the lowest power.

Figure 4.34 presents the power results for the simulations outlined in Table 4.9 when an RM_{SS} test is used and data are simulated with a ν^{-1} of 0.5. As is the case for all power results shown in this section, all settings reported have a RM_{SS} for group of 1.44 but differ as to how that RM_{SS} is attained. Figure 4.34 shows that when age effects that are truly zero are included in the analyses of ZIG and ZILN models using RM_{SS} the power for finding group effects decreases with the correlation between age and group. When there is a weak age effect, the power for finding an RM_{SS} of 1.44 increases with the correlation between age and group for setting 'G' but decreases for the other settings. For strong age effects, power is very low for setting 'G', power stays stable with respect to correlation for setting 'H', and power decreases with correlation for setting 'I' where the nonzero effects are solely in the continuous part of the model. As was the case with DM, analyses correctly assuming ZIG have higher power than analyses incorrectly assuming ZILN. Also, when the data are simulated from ZILN distributions, RM_{SS} tests assuming ZIG and ZILN have similar power levels.

RM_{SS} power results for simulations where $\nu^{-1} = 2$ are show in Figure 4.35. Power decreased relative to the results from Figure 4.34 where $\nu^{-1} = .5$. The decrease in power for analyses assuming ZILN, however, is not as extreme as it was for the DM results. For most settings, power levels for analyses assuming ZILN are lower than those for analyses assuming ZIG. These differences in power are not as large

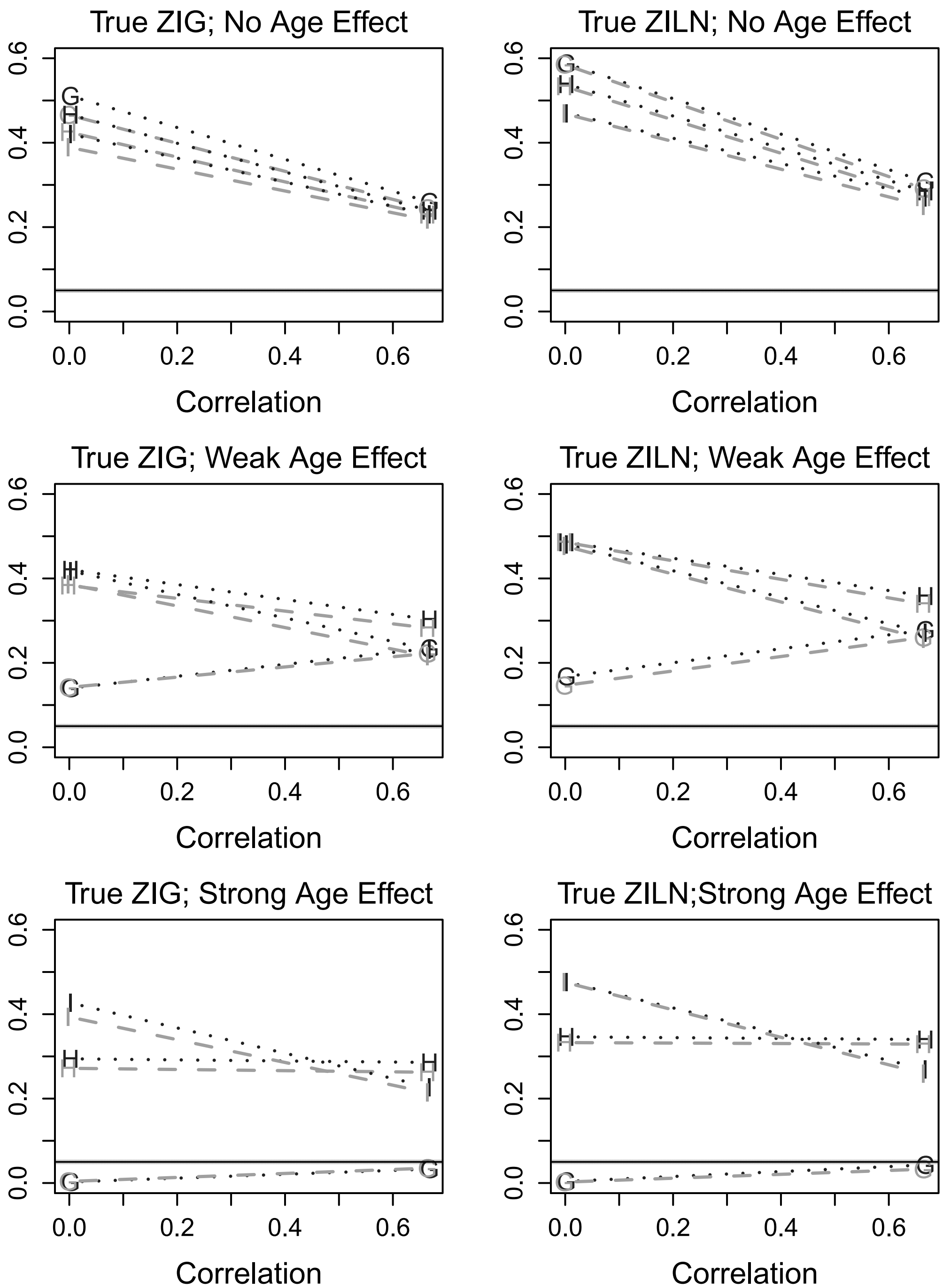

Figure 4.34: Power for RM_{SS} with a continuous adjusting covariate and $\nu^{-1} = 0.5$. Graphs are labeled with the strength of the adjusting age effect. Within the graphs, setting labels correspond to those outlined in Table 4.9; 'G' represents effects only in the binomial part of the model, 'I' effect only in the continuous part of the model, and 'H' effects which are split between the two parts. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

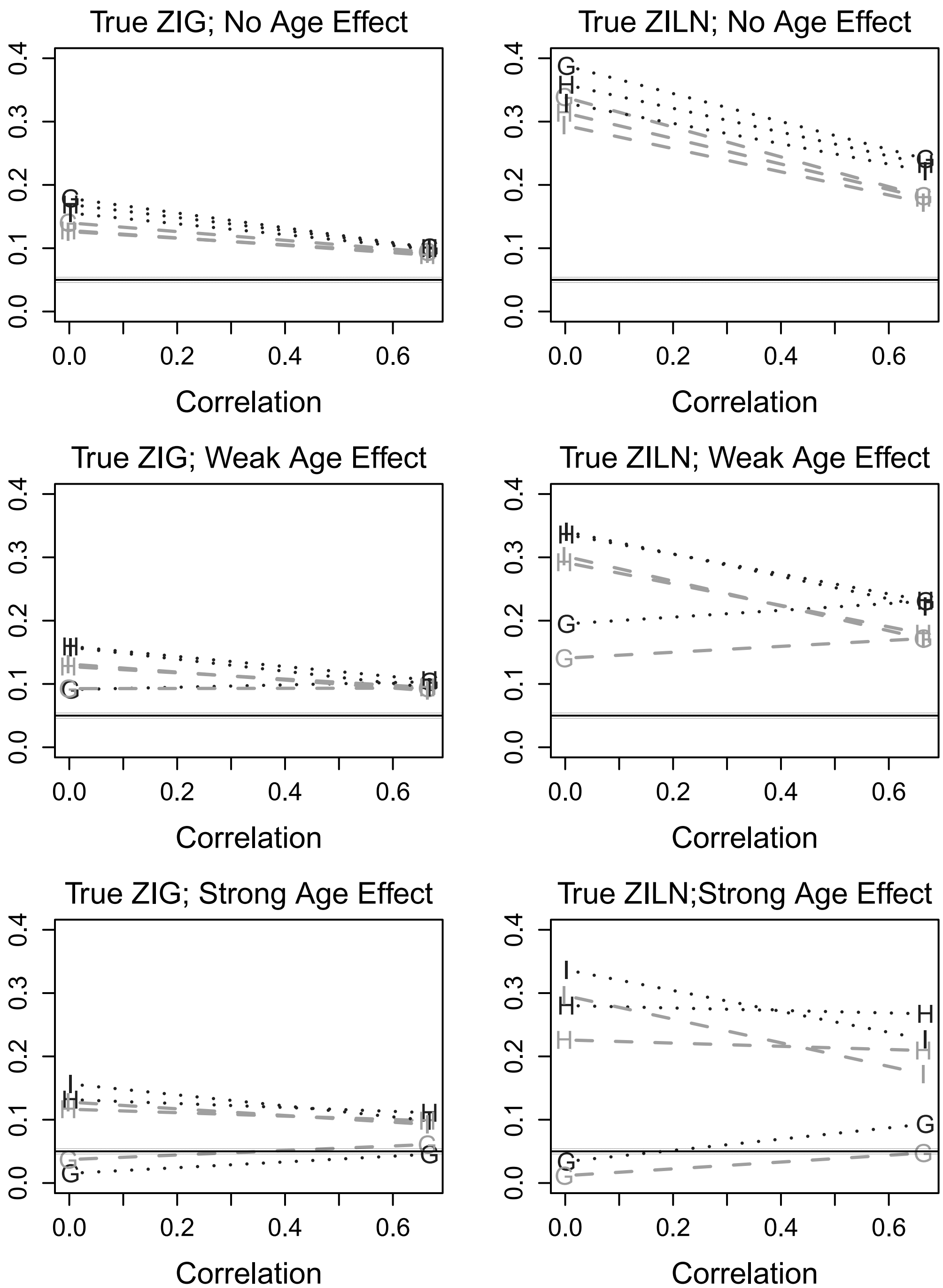

Figure 4.35: Power for RM_{SS} with a continuous adjusting covariate and $\nu^{-1} = 2$. Graphs are labeled with the strength of the adjusting age effect. Within the graphs, setting labels correspond to those outlined in Table 4.9; 'G' represents effects only in the binomial part of the model, 'I' effect only in the continuous part of the model, and 'H' effects which are split between the two parts. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

as those seen when using a DM based test. When the true age effects equal zero and the data are simulated from ZIG distributions, setting 'G' has the highest power when there is no correlation between group and age, power decreases with correlation between group and age for all settings, and at high correlation between group and age all settings have similar power levels. When the data are truly ZILN and age effects equal zero, setting 'G' has the highest power levels, followed by settings 'H' and 'I'. When there are weak adjusting effects, setting 'G' has the lowest power which increases with the age-by-group correlation while settings 'H' and 'I' have higher power levels which decrease with correlation. With strong age effects, setting 'G' has very low power levels which increase slightly with correlation, power levels for setting 'H' are fairly similar at the two correlation levels, and power for setting 'I' decreases with correlation. This leads to setting 'I' (where all effects are in the continuous part of the model) having the highest power levels when age and group are not correlated and setting 'H' (with effects in both parts of the model) having the highest power levels when there is a large correlation.

Figure 4.36 and Figure 4.37 show the power results for RM_{MAR} based tests for simulations where $\nu^{-1} = 0.5$ and $\nu^{-1} = 2$ respectively. Within both figures and within each setting, power decreases as the correlation between age and group increases. When the are no nonzero age effects, power is highest for setting 'G', followed by setting 'H', with setting 'I' having the lowest power levels. For weak age effects, setting 'G' has the lowest power levels and setting 'H' has the highest power levels. When there is a strong age effect, setting 'G' has the lowest power levels and settings 'H' and 'I' have similar power levels. For both $\nu^{-1} = 0.5$ and $\nu^{-1} = 2$ analyses assuming ZILN have lower power than analyses assuming ZIG. The differences in power between ZIG and ZILN analyses are larger when the true distribution is ZIG than when the true distribution is ZILN. Also, there is a greater difference in power between ZIG and ZILN analyses for simulations with $\nu^{-1} = 2$

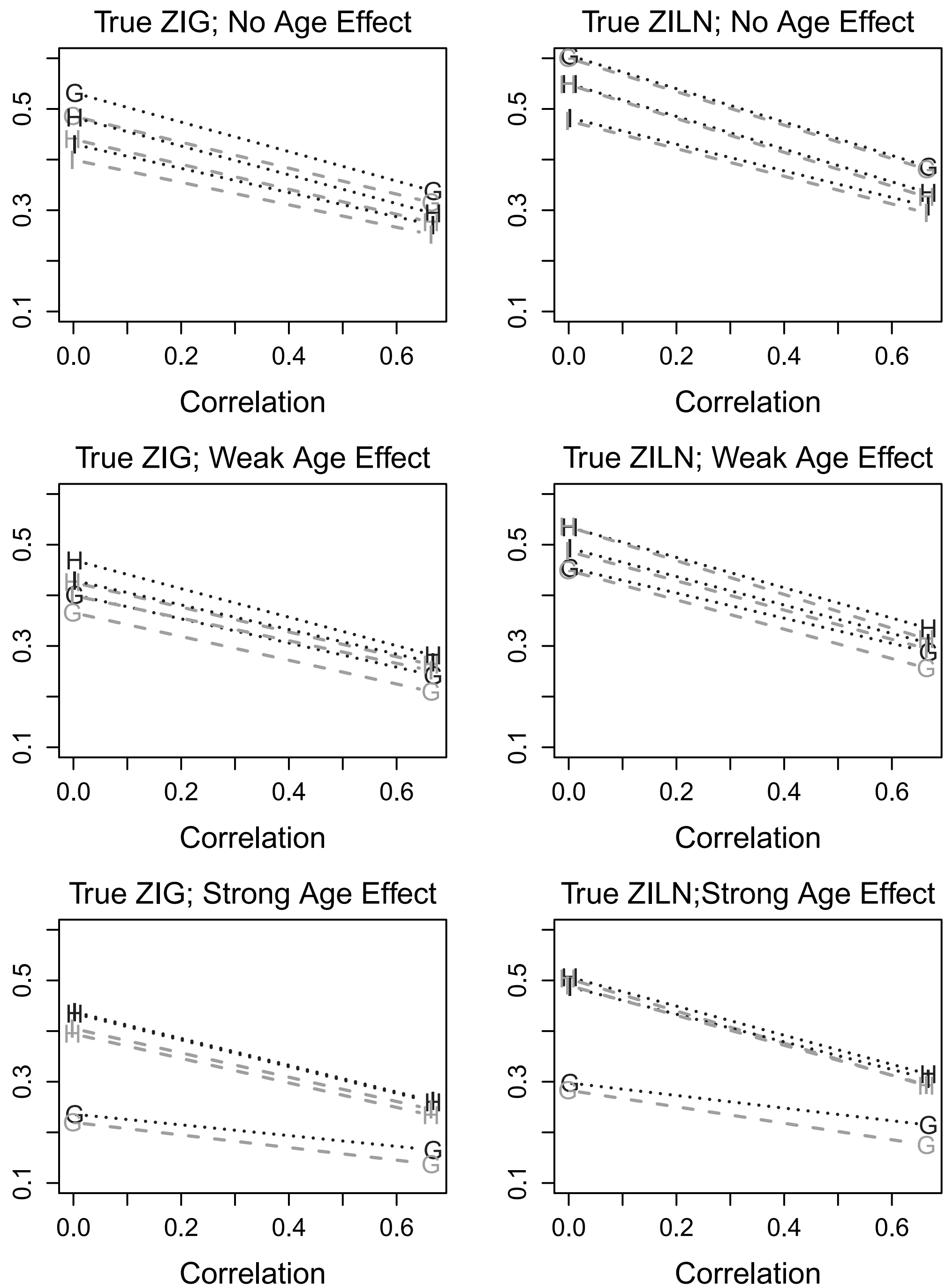

Figure 4.36: Power for RM_{MAR} with a continuous adjusting covariate and $\nu^{-1} = 0.5$. Graphs are labeled with the strength of the adjusting age effect. Within the graphs, setting labels correspond to those outlined in Table 4.9; 'G' represents effects only in the binomial part of the model, 'I' effect only in the continuous part of the model, and 'H' effects which are split between the two parts. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

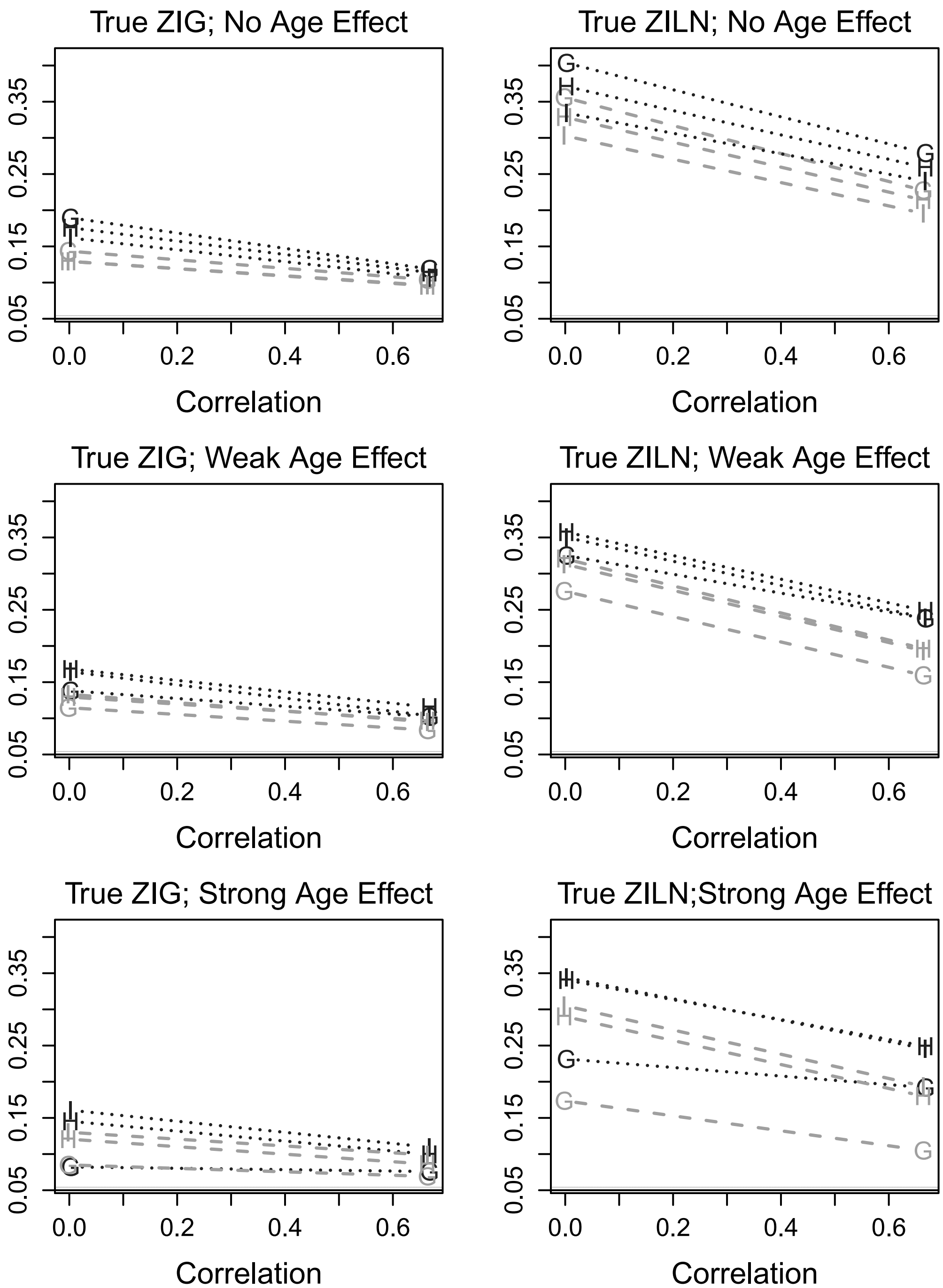

Figure 4.37: Power for RM_{MAR} with a continuous adjusting covariate and $\nu^{-1} = 2$. Graphs are labeled with the strength of the adjusting age effect. Within the graphs, setting labels correspond to those outlined in Table 4.9; 'G' represents effects only in the binomial part of the model, 'I' effect only in the continuous part of the model, and 'H' effects which are split between the two parts. Darker grey symbols with dotted lines refer to ZIG analyses and the lighter grey symbols with dashed lines to ZILN analyses.

than for simulations with $\nu^{-1} = 0.5$.

Figures 4.38 - 4.41 compare power levels among tests. They include power results for DM, RM_{SS}, RM_{MAR}, SLR, and RTR. These tests are represented by the symbols 'D', 'S', 'M', 'L', and 'R' respectively. Power levels for tests assuming ZIG are shown with dark gray symbols and dotted lines, those assuming ZILN are shown with light gray and dashed lines. Figure 4.38 and Figure 4.39 show power results for simulations where there are no age effects, but age is still included in the analyses. Figure 4.40 and Figure 4.41 contain power results for simulations with strong age effects with $\nu^{-1} = 0.5$ and $\nu^{-1} = 2$ respectively. Each of these figures consist of six plots. The plots on the left are for simulations where the true distribution is ZIG and the plots on the right are for simulations where the true distribution is ZILN. The top plots contain setting 'G' where the group $RM_{SS} = 1.44$ is attained purely through effects in the binomial part of the model. The middle plots show the results for setting 'I' where $RM_{SS} = 1.44$ is attained through an effect in the continuous part of the model. The bottom plots in each figure include the settings where the effect of group is split between the two parts of the model.

Figure 4.38 shows power results for the five proposed tests when adjusting for age in scenarios where data were simulated with no age effect and $\nu^{-1} = 0.5$. For all tests and under all effect settings, power decreases as the correlation between age and group increases. When the only nonzero effect of group is in the binomial portion of the model, RTR has the highest power for finding differences between groups. When the nonzero effect of group is in the binomial portion of the model and ZIG is the true data distribution, SLR has the next highest power, followed by DM assuming ZIG and then RM_{MAR} assuming ZIG. When there is no correlation between age and group these are followed by RM_{SS} assuming ZIG, then DM assuming ZILN, RM_{MAR} assuming ZILN, and finally RM_{SS} assuming ZILN. RM_{SS} exhibits a steeper decline in power with the increase in correlation than do the other tests. Because of this when

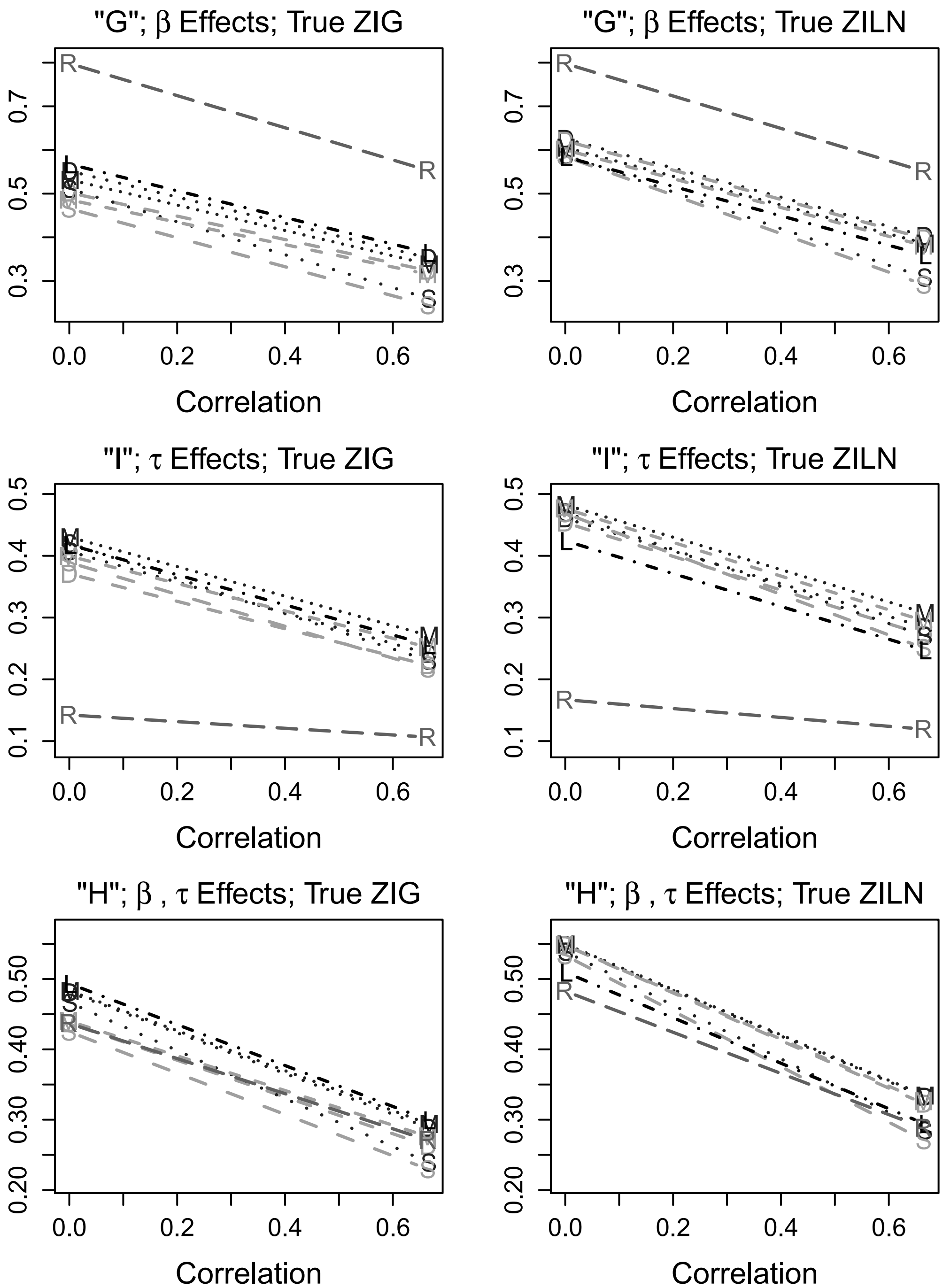

Figure 4.38: Comparison of power for mean-based tests adjusting for a non-existent age effect $\nu^{-1} = 0.5$. Graphs titles indicate settings as outlined in Table 4.9. 'D' represents the power for tests based on DM, 'S' those based on RM_{SS}, and 'M' those based on RM_{MAR} with darker grey symbols with dotted lines for ZIG analyses and lighter grey symbols with dashed lines for ZILN analyses. SLR analysis results are shown with the symbol 'L' and black dash-dot lines; RTR with the symbol 'R' and medium grey lines with long dashes.

the correlation is high, RM_{SS} assuming ZIG followed by RM_{SS} assuming ZILN has the lowest power levels. When the true distribution is ZILN, the differences in power between ZIG and ZILN analyses using the same metric are small. When there is no correlation between age and group, RTR has the highest power, followed by DM, then RM_{MAR}, then RM_{SS}, and finally SLR. When there is a correlation between age and group, RM_{SS} based tests have less power then SLR and the order otherwise stays the same.

The middle graphs show the power levels when the nonzero effects are solely in the continuous part of the model, setting 'I'. When this is the case, the power levels for RTR are considerably lower than the power levels for the other tests. When the simulated distribution is ZIG, for each test ZIG analyses have greater power levels than do ZILN analyses. RM_{MAR} tests have the highest power levels, DM has the lowest power when there is no correlation between age and gender, and DM and RM_{SS} have similar power levels when the correlation between age and gender is high. SLR has power levels similar to those for tests correctly assuming ZIG. Power levels are higher when the data are simulated from ZILN distributions than when they are simulated from ZIG distributions. When the true distribution is ZILN, differences in metric choice lead to larger differences in power than differences in distribution choice. SLR has slightly lower power than do the mean-based tests based on the two-part models. Power for RM_{MAR} tests are the highest followed by RM_{SS} and then DM when there is no correlation between age and group and followed by DM and then RM_{SS} when there is a correlation between age and group.

The bottom plots in Figure 4.38 show the power levels when the effects of group and age are included in simulating both the probability of a nonzero outcome and the distribution of the values of the nonzero outcome. As in the other scenarios, power levels are higher for ZILN data than they are for ZIG data. When the true data are ZIG, power for finding group differences using SLR is slightly higher than for other tests and within test type, and power is higher when ZIG is correctly assumed than when ZILN is assumed. For ZIG data, RM_{SS} power levels are the lowest within distributional assumption and decline most steeply as the correlation increases. When the data are truly ZILN and there is no correlation between group and age, DM, RM_{SS}, and RM_{MAR} all have similar levels of power while SLR and RTR have lower power levels. Power for RM_{SS} declines more steeply with correlation than do the other tests. Because of this RM_{SS} has the lowest power when there is a correlation between group and age with levels similar to those seen for SLR and RTR. DM and RM_{MAR} for analyses assuming both ZIG and ZILN have similar levels of power.

Figure 4.39 compares the power levels of the metrics proposed when there are no simulated age effects and $\nu^{-1} = 2$. In similar form to the previous figure, power levels decrease with increasing correlation between age and group, although the decreases in power are not quite as drastic as they were when $\nu^{-1} = 0.5$. This may be in part because the power levels in these $\nu^{-1} = 2$ scenarios are already lower than they were when $\nu^{-1} = 0.5$. The top plots compare the power levels of the several tests when there is a nonzero effect of group in the continuous part of the model only. For this setting, RTR has much higher power than does SLR and the ZIG and ZILN based tests. When the true data distribution is ZIG, DM assuming a ZILN distribution has the lowest power. SLR has higher power levels than do the ZIG and ZILN based tests. When there is no correlation between age and group, DM, $RMSS$, and RM_{MAR} tests assuming ZIG have higher power levels than do $RMSS$ and RM_{MAR} tests assuming ZILN. When there is a correlation between group and age, the ZIG and ZILN tests (excluding ZILN DM) have power levels somewhat closer to each other with ZIG RM_{MAR} having the highest power and ZIG DM the lowest. When the data are simulated from ZILN distributions, power levels for ZIG and ZILN based tests are higher than they were for data simulated from ZILN distributions. Power levels for ZIG based tests are slightly higher than those for ZILN based tests, but since the

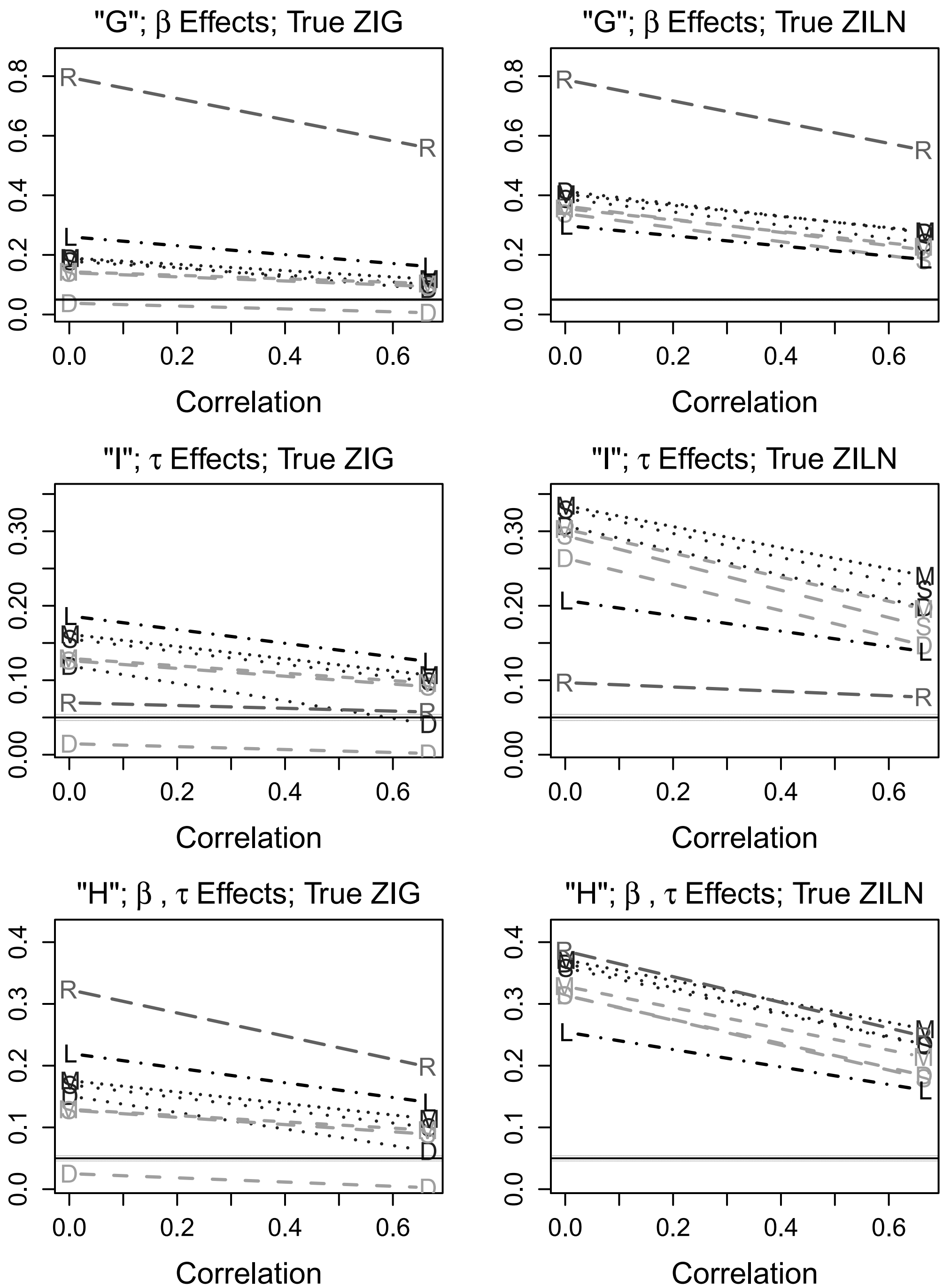

Figure 4.39: Comparison of power for mean-based tests adjusting for a non-existent age effect $\nu^{-1} = 2$. Graphs titles indicate settings as outlined in Table 4.9. 'D' represents the power for tests based on DM, 'S' those based on RM_{SS}, and 'M' those based on RM_{MAR} with darker grey symbols with dotted lines for ZIG analyses and lighter grey symbols with dashed lines for ZILN analyses. SLR analysis results are shown with the symbol 'L' and black dash-dot lines; RTR with the symbol 'R' and medium grey lines with long dashes.

misspecified ZIG tests have unacceptably elevated Type 1 error rates, these must be dismissed. Among the ZILN tests, DM and RM_{MAR} have slightly better power levels than RM_{SS}. SLR tests have lower power than the ZILN tests when there is no correlation between age and group and power similar to RM_{SS} assuming ZILN when there is a correlation between age and group.

The middle plots show setting 'I' where the effect of group is only nonzero in the continuous part of the model. For setting 'I' when simulated from a true ZIG distribution, SLR has the highest power levels followed by RM_{MAR} and RM_{SS} assuming ZIG and then RM_{MAR} and RM_{SS} assuming ZILN. Then when the correlation between group and age equals zero, DM assuming ZIG is the next lowest followed by RTR, but when there is a correlation between group and age, RTR has higher power than DM assuming ZIG. At both correlation levels, DM assuming ZILN has the lowest power. When the data are truly ZILN, tests assuming ZIG have the highest power levels, which again should be discounted due to high Type 1 error rates. Among tests which correctly specify ZILN, RM_{MAR} has the highest power, followed by RM_{SS}, and then DM. These have higher power than SLR analyses which have higher power than RTR analyses.

The bottom plots compare the power levels for setting 'H' where effects of group are present in both parts of the model. When the data are simulated from ZIG distributions, RTR has substantially higher power levels than the other tests. SLR has the next highest power levels. DM assuming ZILN consistently has the lowest power levels. DM assuming ZIG decreases in power with the increase in correlation to a greater extent than do the other tests causing the order of the power for the ZIG tests and RM_{MAR} and RM_{SS} assuming ZILN to differ depending on the correlation. When the correlation between group and age is nonexistent the relative order of the power levels are ZIG RM_{MAR}, followed by ZIG RM_{SS}, followed by ZIG DM, followed by RM_{MAR} and RM_{SS} assuming ZILN. At the setting with correlation between group and age, the order of the power is: ZIG RM_{MAR}, followed by ZIG RM_{SS}, followed by RM_{MAR} and RM_{SS} assuming ZILN, followed by ZIG DM. When data are simulated from a ZILN distribution, RTR has the highest power of the acceptable tests. Tests assuming ZILN have lower power with power for RM_{MAR} being slightly higher than power for RM_{SS} and DM. SLR has the lowest power levels for both correlation settings.

Figure 4.40 presents power results for settings where there are nonzero age effects involved in the simulations and where $\nu^{-1} = 0.5$. The power levels observed here with age effects present are lower than those seen in Figure 4.38 where no age effects were present. In Figure 4.40 the decline in power within most tests as the correlation with age increases is smaller than it was when there was no age effect. The top plots show that when the effects of age and of group are only in the binomial part of the model, RTR has the highest power levels and RM_{SS} tests have the lowest power levels. When the data are simulated from ZIG distributions, SLR has higher power levels than DM and RM_{MAR} assuming ZIG which have higher power levels than DM and RM_{MAR} assuming ZILN. When there is a correlation between age and group the differences between the power levels of SLR and the other tests are larger. When the data are simulated from ZILN distributions and there is no correlation between group and age, DM and RM_{MAR} assuming ZIG have higher power levels that DM and RM_{MAR} assuming ZILN which have higher power levels than SLR. When there is a correlation between group and age, SLR has higher power levels than DM and RM_{MAR} based tests.

Power results for the setting where nonzero effects of age and group are only in the continuous part of the model are shown in the middle plots. For these settings, SLR and RTR have lower power levels than the tests which assume ZIG or ZILN distributions. When there is no correlation between age and group and for both true ZIG and true ZILN data, SLR has considerably higher power than RTR. When

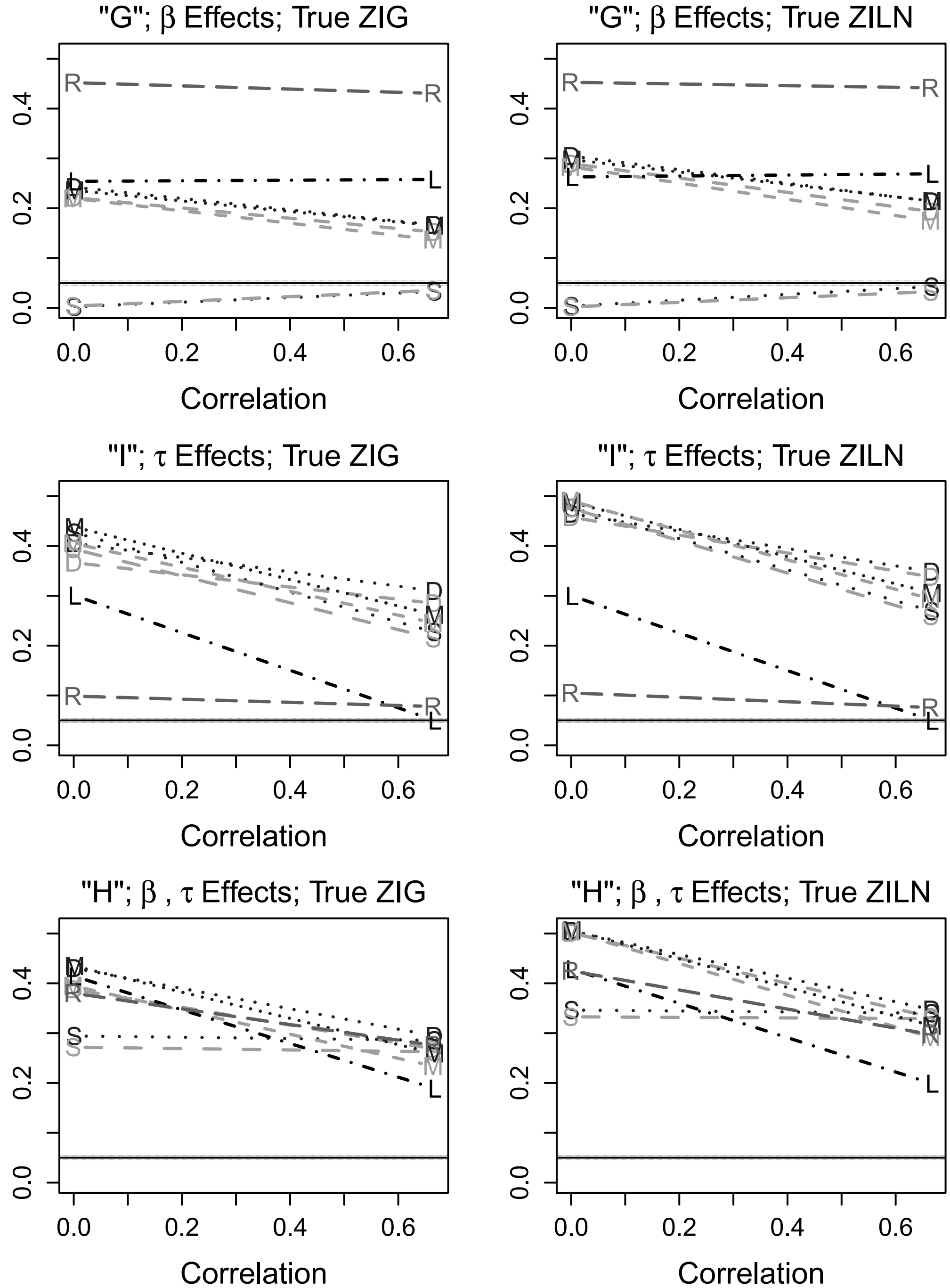

Figure 4.40: Comparison of power for mean-based tests adjusting for a strong age effect $\nu^{-1} = 0.5$. Graphs titles indicate settings as outlined in Table 4.9. 'D' represents the power for tests based on DM, 'S' those based on RM_{SS}, and 'M' those based on RM_{MAR} with darker grey symbols with dotted lines for ZIG analyses and lighter grey symbols with dashed lines for ZILN analyses. SLR analysis results are shown with the symbol 'L' and black dash-dot lines; RTR with the symbol 'R' and medium grey lines with long dashes.

there is a correlation between group and age, RTR has slightly higher power than SLR. The relative power levels of the ZIG and ZILN based tests change with the correlation of group and age. For data simulated as ZIG with no correlation between age and group, RM_{MAR} assuming ZIG has the highest power, followed by $RM_{S}S$ assuming ZIG, followed by DM assuming ZIG and RM_{MAR} assuming ZILN, then RM_{SS} assuming ZILN with DM assuming ZILN having the lowest power. When there is a correlation between group and age the order changes with DM assuming ZIG and followed by DM assuming ZILN having the highest power, then RM_{MAR} assuming ZIG and followed by RM_{MAR} assuming ZILN, and then RM_{SS} assuming ZIG followed by RM_{SS} assuming ZILN having the lowest power levels. When the true data are ZILN and there is no correlation between group and age the power levels for all ZIG and ZILN based tests are fairly close to each other with tests based on RM_{MAR} having the highest power and tests based on DM having the lowest. When there is a correlation between age and group, there are differences in power among the tests with DM tests having higher power than RM_{MAR} tests which have higher power than RM_{SS} tests.

Power results for simulations where the effects of age and group affect both the probability of nonzero outcomes and their values are shown in the bottom plots. For these settings, RTR and SLR power levels are within similar ranges as the other tests. For data simulated as ZIG with no correlation between age and group, DM and RM_{MAR} assuming ZIG have the highest power levels, followed by SLR, then DM and RM_{MAR} assuming ZILN and RTR, followed by RM_{SS} assuming ZIG and ZILN which have much lower power than the other tests. When there is a correlation between group and age, SLR has lower power than the other tests, while the other tests have similar power levels with the highest being DM assuming ZIG and the lowest being RM_{MAR} assuming ZILN.

Figure 4.41 shows the power results for the various tests using simulations where

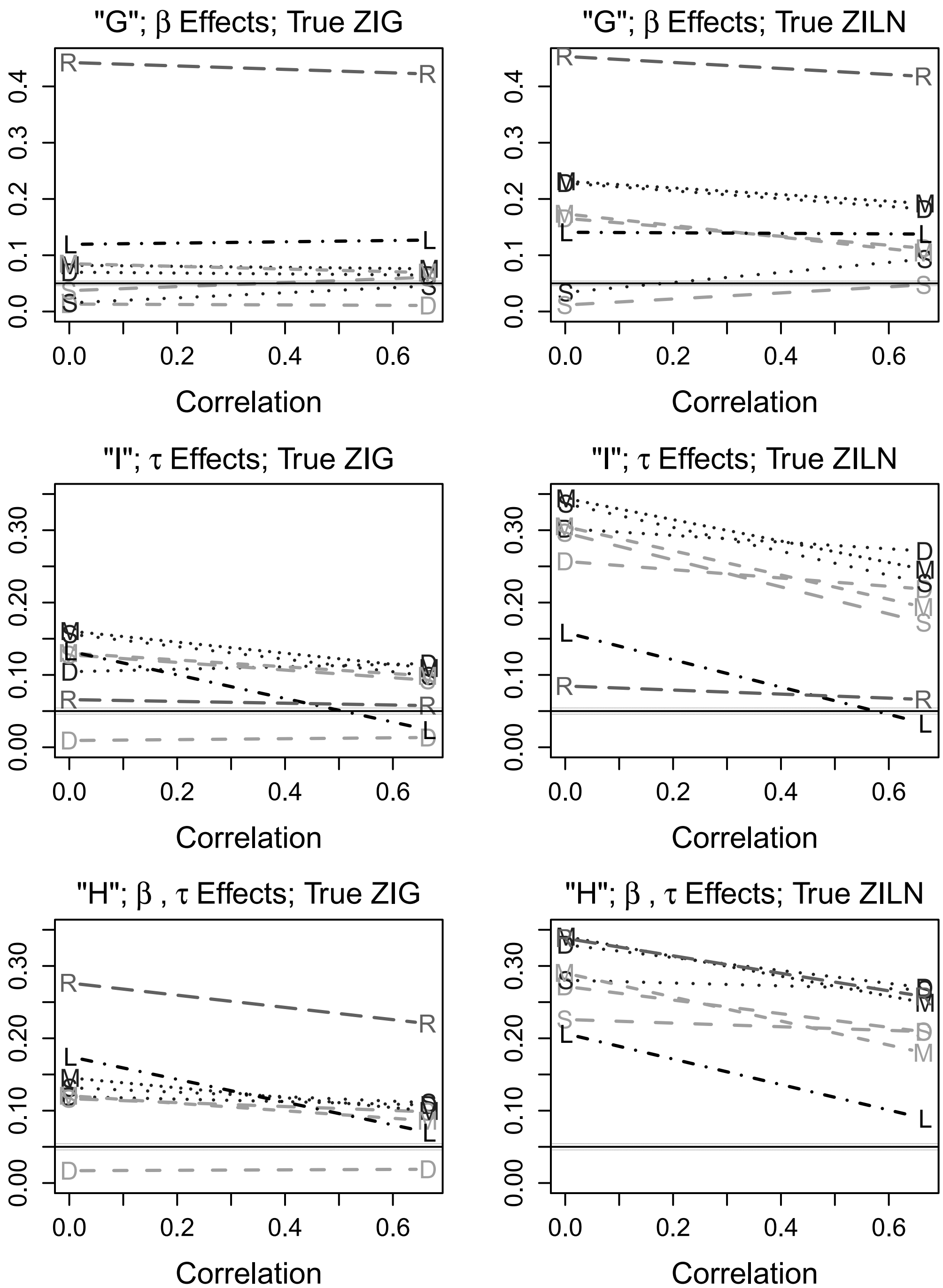

Figure 4.41: Comparison of power for mean-based tests adjusting for a strong age effect $\nu^{-1} = 2$. Graphs titles indicate settings as outlined in Table 4.9. 'D' represents the power for tests based on DM, 'S' those based on RM_{SS}, and 'M' those based on RM_{MAR} with darker grey symbols with dotted lines for ZIG analyses and lighter grey symbols with dashed lines for ZILN analyses. SLR analysis results are shown with the symbol 'L' and black dash-dot lines; RTR with the symbol 'R' and medium grey lines with long dashes.

there are strong age effects and $\nu^{-1} = 2$. RTR has much higher power levels when the effect of group is in the binomial portion of the model, setting 'G', than do the other tests. For setting 'G' when the data are truly ZIG, after RTR, SLR has the highest power followed by RM_{MAR} both when assuming ZIG and ZILN, followed by DM assuming ZIG, with RM_{SS} for both ZIG and ZILN analysis assumptions and DM assuming ZILN having the lowest power levels. When the data are simulated from ZILN distributions, after RTR, the highest power levels are for RM_{MAR} and RM_{SS} incorrectly assuming ZIG. However, misspecified ZIG analyses when $\nu^{-1} = 2$ must be discounted because of unacceptably high Type 1 error rates. Among the more acceptable tests, RM_{MAR} and RM_{SS} correctly assuming ZILN have higher power than SLR when there is no correlation between age and group and lower power than SLR when the correlation between group and age is high. RM_{SS} correctly assuming a ZILN distribution has very low power levels for finding such binomial group effects.

For setting 'I' the effect of group is in the continuous part of the model, influencing the data simulated by increasing the conditional mean ($E(Y|Y > 0)$). For this setting when the data are simulated from ZIG distributions, tests assuming a ZIG distribution again have higher power than the corresponding tests assuming a ZILN distribution. With no correlation between group and age, RM_SS and RM_{MAR} assuming ZIG have the highest power levels followed by SLR and RM_SS and RM_{MAR} assuming ZILN, then DM assuming ZIG, then RTR, with the lowest power levels occurring with misspecified ZILN when DM is used for the test. When there is a correlation between group and age, DM assuming ZIG has the highest power levels, followed by RM_{MAR} assuming ZIG, then RM_{SS} assuming ZIG and RM_SS and RM_{MAR} assuming ZILN, followed by RTR and then SLR and DM assuming ZILN. Power levels are higher when the data are truly ZILN. With ZILN as the true distribution for setting 'I' with $\nu^{-1} = 2$, power levels for the two-part model mean-based tests are higher than those for SLR and RTR. When there is no correlation between group and age, SLR has higher power than RTR and when there is a correlation between group and age, RTR has higher power than SLR. Among the tests where ZILN is correctly assumed and there is no correlation between age and group, RM_{MAR} and RM_{SS} have higher power levels than DM. When there is a correlation between group and age, DM has the highest power levels followed by RM_{MAR} and then RM_{SS}.

Setting 'H', pictured in the bottom plots, contains effects of group that influence both the probability of a nonzero value and the average of the nonzero values. When the setting 'H' data are simulated from a ZIG distribution, RTR has the highest power levels. When there is no correlation between group and age, SLR has the next highest power levels, followed by the ZIG and ZILN tests (excluding ZILN DM) with RM_{MAR} assuming ZIG having the highest power among them. When there is a correlation between age and group, SLR has lower power levels than tests assuming ZIG and ZILN (excluding ZILN DM). DM tests assuming ZILN have the lowest power at both correlation levels. When the data are simulated from ZILN distributions, RTR has power levels that are similar to the incorrectly specified (and inappropriate) ZIG analyses, which are higher than the ZILN analyses and SLR has the lowest power. Among the ZILN analyses when there is no correlation between group and age, RM_{MAR} yields the highest power levels, followed by DM, and then RM_{SS}. When there is a correlation between group and age, DM and RM_{SS} have the highest power levels for the tests assuming ZILN, with RM_{MAR} having the lowest.

4.9 Summary

This chapter included our most novel methods. We proposed a test based on the difference of means and two tests related to mean ratios (both a ratio of marginal means and an average of the subject specific ratio of means) which included covariate adjustment.

The tests proposed in this chapter could be used to adjust for any number of covariates simultaneously. However, in our simulation we focus only on adjusting for one dichotomous or continuous covariate at a time. Within each of these focuses a simulation study using three groups of settings was performed. The first examines Type 1 error rates when there are no group differences, and varied effects of the adjusting covariates. The second group of settings examines Type 1 error rates when there are dissonant effects, again with varying effects of the adjusting covariates. The final group of settings examines the power to observe group differences in the presence of adjusting covariates. In addition to examining settings with different effects of group and adjusting covariates, the data are constructed as to impose various levels of correlation between the group variable and the adjusting covariate.

As was the case in the non-adjusting scenario in Chapter 3, for all settings and adjusting covariates, Type 1 error rates for DM tests are extremely low when $\nu^{-1} = 2$, and more markedly so for DM tests assuming ZILN. Type 1 error rates slowly improve with sample size, this low Type 1 error rate may be indicative of slow convergence of the DM Wald test to a standard normal. Also, Type 1 error is elevated when ZILN data are analysed assuming ZIG, and more so when ν^{-1} is large. The Type 1 error rates of DM tests remain similar or decline only slightly as the correlation between group and the adjusting covariate increases. These results hold for adjusting for both dichotomous and continuous covariates and for both general null and dissonant effects leading to metric-based null settings.

The RM_{SS}-based test has Type 1 error rates more dependent on setting than is the case for the other two-part model tests. RM_{SS} Type 1 error rate changes with the location of the effect in the general null settings, the size of the dissonant effects in the mean-based null settings, and for some settings with the level of correlation. This results in RM_{SS} for some settings having Type 1 error rates that are higher than the other tests, for other settings Type 1 error rates that are at similar levels when compared to other tests, and occasionally Type 1 error rates that are lower than other settings.

The RM_{MAR} Type 1 error rates followed similar patterns as the DM results except for ZIG data with $\nu^{-1} = 2$ where RM_{MAR} had appropriate Type 1 error rates for analyses assuming ZILN (instead of excessively conservative Type 1 error rates) and only slightly conservative Type 1 error rates for analyses assuming ZIG. Another difference between RM_{MAR} and DM results is that when ZILN data area analysed assuming ZIG, RM_{MAR} has slightly higher Type 1 error rates than DM for some settings.

Setting choice greatly influenced the relative power of the examined tests. This is most pronounced when the power for RTR and, to a lessor extent, SLR are compared to the two-part model based tests. For example, RTR has much higher power levels than the two-part model based tests when one group has a higher probability of a non-zero outcome than another, but RTR has much lower power than the two-part model based tests when conditional group means are different. Another important power result is that power is very low (at or near 0) when ZIG data with $\nu^{-1} = 2$ are incorrectly analysed assuming ZILN. Typically, power levels are higher when the data are simulated from ZILN distributions than when the data are simulated from ZIG distributions. And typically, power decreases as correlation increases for most tests and settings. (However, RM_{SS} and SLR occasionally break this rule.)

CHAPTER 5
APPLICATIONS AND MODEL FITTING CONSIDERATIONS

5.1 Driving Simulator Scenario and Lane Departure Severity Score

Our data example comes from a study of driving in Parkinson's disease (PD). PD and control subjects drove in a driving simulator, SIREN (Simulator for Interdisciplinary Research in Ergonomics and Neuroscience). SIREN consists of a 1994 GM Saturn, embedded electronic sensors, videos cameras, a sound system, and 150 degree view screen to the front and a smaller, 50 degree, screen visible in the rear-view mirror. As the subject drives, data on location, steering, pedals etc. are recorded at 30 frames per second. These data are then used to calculate various driving safety metrics.[22, 23]

Although various driving metrics have been used to measure unsafe driving including standard deviation of lane position and steering variability, measures of lane departure correspond directly to the rules of the road and so could be a valuable addition to a battery of driving safety measures. Specifically, lane deviation count is another measure of unsafe driving that is used in driving research.[24] However, lane deviation count is a limited measure in that it only considers the number of times that the tire of a subject's car crosses the lane line and does not included the severity of the lane line crossing. In previous research we have used a metric which we will call 'Lane Departure Severity Score' (LDSS) which measures the extent to which a subject is outside of their lane. This metric had previously been used by Uc et al, but was termed the 'percent volume outside of the lane'.[9]

Figure 5.1 gives three examples of paths traveled 60 second portion of a simulator drive. This plot shows the subject's position on the road in relation to the lane lines on the x-axis with time in seconds on the y-axis. The position of subject's vehicle is shown with gray lines approximating the position of the left and right sides of the vehicle, with the vehicle facing and traveling towards the right side of the figure. If the entire vehicle is below the dashed lane line but above the lower solid black line, the subject is driving in appropriate driving lane; driving outside of those boundaries constitutes a lane deviation. The road is a rural 2-lane highway with the width of each lane of 3.6 meters. The subject's car is 1.556 meters wide. For each subject, LDSS was calculated by finding the width of the car that was outside of the lane at each frame, adding those widths up over the frames in the segment, then dividing by the length of the segment (typically 60 seconds * 30 frames per second) and the width of the car to create a unit free metric that measures the extent to which a subject is out of the driving lane. Referring to Figure 5.1, LDSS is the percentage of the area between the two gray lines (denoting the locations of the left and right sides of the vehicle at a particular time point) that is outside of the subject's designated lane lines. The top plot illustrates the common occurrence of LDSS=0, which occurs when a subject remains in the lane for the entire segment. The middle plot illustrates an LDSS of 0.0146 resulting from a slight lane deviation which lasted a few seconds. The bottom plot shows an example of an extremely high LDSS of .3949. In this example, the car is approximately 40% outside of the lane for most of the drive segment. Another way a subject could have an LDSS of around .4 would be if they were in the lane for 60% of the drive (about 36 seconds) and completely outside of the lane for 40 % of the drive segment (about 24 seconds).

LDSS is a semicontinuous variable with subjects that do not drive outside of the lane having an LDSS equal to zero and subjects that do drive outside of the lane having LDSS greater than zero. As subjects can be directly observed to not exit the lane for the duration of the segment, the zero values observed can be considered true zeros that are not subject to censoring. Since LDSS is a semicontinuous variable without censoring, the methods outlined in this dissertation may be used to compare the expected means of two groups. In this instance subjects with Parkinson's disease

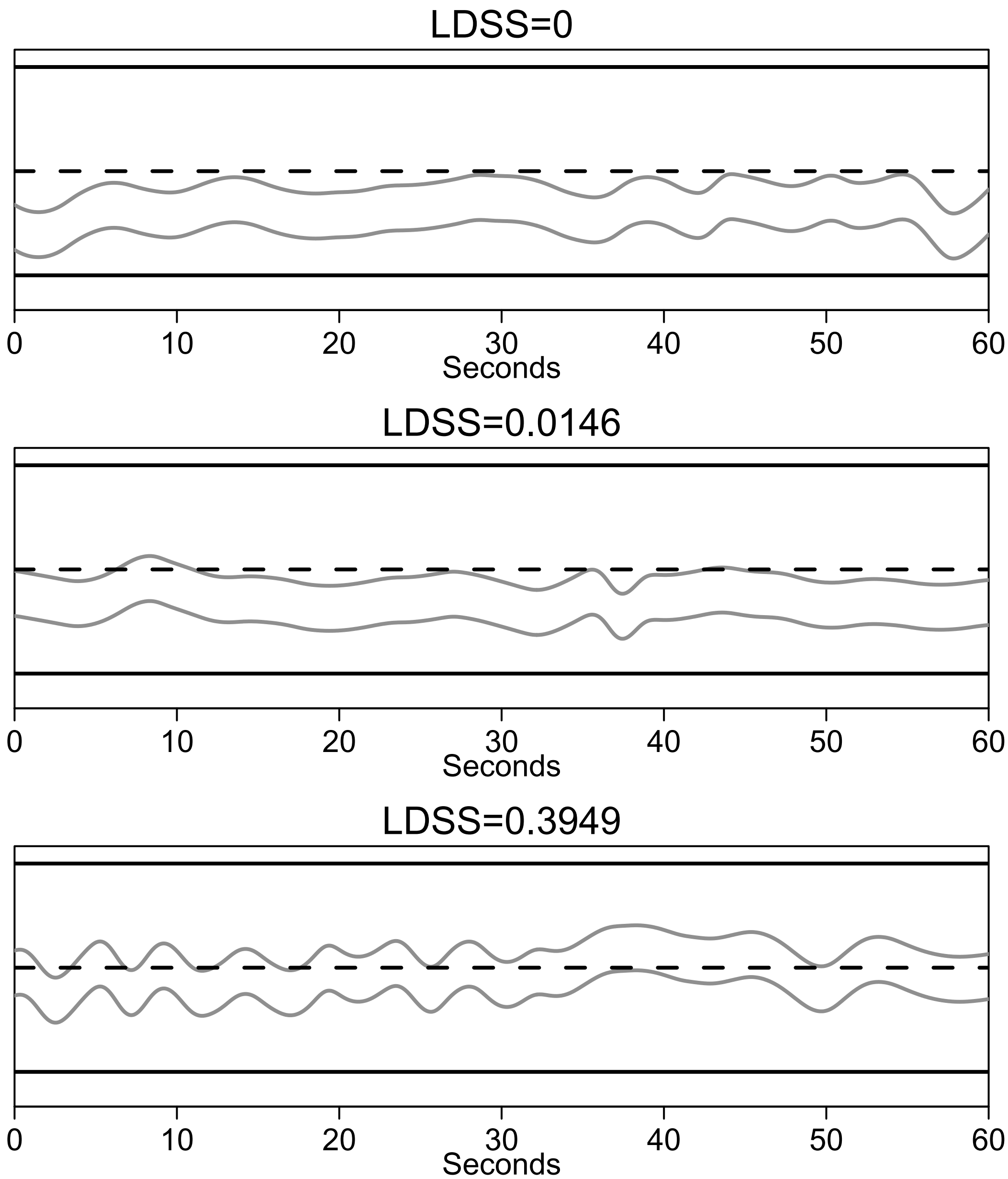

Figure 5.1: Illustration of Lane Departure Severity Score over a 60 second segment.

will be compared to control subjects. Sections 5.2.1, 5.2.2, and 5.2.3 present analyses results comparing PD and control subjects without any covariate adjustment, adjusting for gender, and the adjusting for age, respectively.

5.2 Lane Keeping Example

In our study, 76 PD subjects and 51 control subjects completed the drive segment that is used for these analyses. LDSS was calculated for a one minute segment in which the subjects were driving in the fog near the end of a longer drive with no other tasks or distractions present. A few subjects drove through the uneventful segment too quickly and so a slightly shorter drive length was used to calculate their LDSS.

With data recovered at 30 frames per second and precise lane position information even very small lane deviations are observable. The smallest observable lane deviation would be for a vehicle outside of the lane frame one frame (1/30th of a second) by 0.001 m. The resultant smallest calculable positive LDSS is about 3.57e-7. The next smallest LDSS would come from either being outside of the lane for one frame by 0.002 m or by being outside of the lane for 2 frames by 0.001 m each frame. This would double the LDSS to 7.15e-7. For the segment used in this dissertation, all subjects with observed lane deviations had LDSS values greater than 0.00002, which is about 77 times the minimum possible positive value.

5.2.1 Two Group Comparison Without Covariate Adjustment

The results of PD and Control comparisons without any covariate adjustments are presented in Table 5.1. The tests included are, ZIG and ZILN mean-based tests, Wilcoxon test, and 2-part 2 d.f. Wald tests. For the Welch's t-test and ZIG and ZILN DM results estimates and 95% confidence intervals of the difference of the

	Estimate [95% CI]	Test Statistic	p-value
t-test[1]	0.0034 [0.0008, 0.0059]	T = 2.58	0.0111
Wilcoxon	- - -	$\chi^2_1 = 12.78$	0.0003
ZIG DM	0.0034 [0.0006, 0.0062]	Z = 5.50	0.0190
ZIG RM	5.5691 [1.1185, 27.729]	Z = 4.40	0.0360
ZIG 2 d.f. Wald	- - -	$\chi^2_2 = 10.84$	0.0044
ZILN DM	0.0081 [-0.0009, 0.0172]	Z = 3.12	0.0775
ZILN RM	18.525 [3.007, 114.13]	Z = 9.90	0.0017
ZILN 2 d.f. Wald	- - -	$\chi^2_2 = 14.54$	0.0007
2 d.f. Wald-Wilcoxon	- - -	$\chi^2_2 = 13.36$	0.0013

Table 5.1: Estimates, 95% confidence intervals (CI), test statistics, and p-values for tests comparing PD and Control subjects without covariate adjustment.

[1] t-test is a Welch's two-sample t-test with d.f. = 112.711.

overall expected means, the expected LDSS given PD minus the expected LDSS given Controls, as well as test statistics and p-values are presented. For ZIG and ZILN RM the ratio of the expected group means is presented with the expected LDSS given PD divided by the expected LDSS given Controls and 95% confidence intervals for the ratio as well as test statistics and p-value included. For the Wilcoxon, ZIG and ZILN 2 d.f. Wald, and the 2 d.f. Wald-Wilcoxon tests, we report only test statistics and p-values.

For the mean-based tests, when assuming ZILN, the confidence intervals for both DM and RM are wider than when assuming ZIG; more than double the width for DM and more than four times the width for RM. This may be in part because the estimate for coefficient of variation (CoV) is much larger under the ZILN model ($CoV = 5.5708$) than under the ZIG model ($CoV = 1.5867$). ZIG and ZILN models also give slightly different mean estimates. The group means found assuming ZIG were estimate as $M_0 = 0.0007$ and $M_1 = 0.0041$; when assuming ZILN, the group means were $M_0 = 0.0005$ and $M_1 = 0.0086$. Assuming ZIG, $DM = 0.0034$ with 95% C.I. [0.0006, 0.0062]; this means that PD subjects had LDSS values that were 0.0034 higher on average than the values of LDSS for control subjects. When assuming ZILN, the estimate for the difference between the two group means is $DM = 0.0081$ 95% C.I. [-0.0009, 0.0172]. The t-test estimate of DM is very close to that of the ZIG DM results with $DM = 0.0034$ and a narrower 95% C.I. confidence interval of [0.0008, 0.0059]. Both the t-test and ZIG DM find estimates that are significantly different from 0 at the 0.05 level, ZILN DM does not result in a DM significantly different from 0 at the 0.05 level. For the RM metric the interpretations are that, assuming ZIG, subjects with Parkinson's disease have an LDSS that is on average 5.5691 times the LDSS of control subjects (95% C.I.: [1.1185, 27.729]) and that assuming ZILN, subjects with PD have and LDSS that is on average 18.525 times that of the control subjects (95% C.I.: [3.007, 114.13]). RM results showed a difference in the two groups regardless of distribution assumed with RM significantly different that 1 at the 0.05 significance level for both ZIG and ZILN model assumptions. The Wilcoxon rank sum test compares the two groups in a directional manner using all of the data for a 1 d.f. test including both zero and non-zero values. The Wilcoxon rank sum test finds that the two groups, PD and control, differ in terms of land departure severity score (p-value = 0.0003).

The three 2 d.f. tests included in Table 5.1 are the two-part tests which were defined in Chapter 2. As a review, for each of these three tests, two χ_1^2 tests are added together to create a global test of any difference between the two groups in terms of either probability of a non-zero outcome or the value of a non-zero outcome given it is non-zero (e.g. the hypotheses noted in Equation 1.2 from Chapter 1). For Wald-Wilcoxon, ZIG 2-part Wald, and ZILN 2-part Wald, the first part of the two-part test tests the difference between the two groups in terms of the probability

		Test Statistic	p-value
$P(Y > 0)$ post-test	$W(\beta_1)$	Z = 3.214	0.0013
$Y\|Y > 0$ post-tests	ZIG $W(\tau_1)$	Z = 0.714	0.4752
	ZILN $W(\tau_1)$	Z = 2.052	0.0402
	Wilcoxon	Z = 1.740	0.0818

Table 5.2: Post-tests as follow-up to the 2-part tests; comparing PD and Control subjects without covariate adjustment.

of a non-zero outcome using logistic regression. The second part differs depending on assumptions of ZIG, ZILN, or use of a Wilcoxon, with each test comparing the mean of the two group LDSS given LDSS > 0. All three two-part tests show that there is a statistically significant difference between PD and Controls in terms of either the probability of going outside of the lane and/or the extent outside of the lane given a lane deviation (all p-values< 0.005).

To complete the interpretation of significant two-part tests, post-tests can be performed using each part separately to determine whether the difference between the two groups observed in the two-part tests is because of differing probabilities of non-zero outcomes or differences in the values of those non-zero outcomes. For this non-adjusting situation, post-tests are show in Table 5.2. For all three 2 d.f. tests, the portion modeling the probability of a non-zero outcome is identical; where the tests differ in is the method used to analyze the distribution of the non-zero outcomes. From $W(\beta_1)$, we see that PD cases have a higher probability of exiting the lane than do controls. However, the extent outside of the lane does not differ significantly according to ZIG, but according to ZILN, cases the extent outside of the lane given a lane deviation is higher for cases than for controls. The AIC of two models indicate almost identical fit: ZIG AIC = -126.98, ZILN AIC = -128.94.

5.2.2 Two Group Comparison Adjusted for Gender

Our sample of PD and control subjects does not include equal numbers of each gender in each group. For PD subjects our study had 65 male and 11 female subjects. Among controls the breakdown of gender was 26 male and 25 female.The resultant correlation between group and gender is 0.3758 (p-value=0.00001). Because gender may also influence driving habits, in this section we compare PD and control subjects while adjusting for gender. We also examine the effect of gender adjusted for group.

Table 5.3 contains estimates, 95% confidence intervals, test statistics, and p-values for SLR, DM, RM_{SS}, and RM_{MAR} for the effect of group adjusted for gender and test statistics and p-values for rank transformed regression and 2 d.f. Wald tests. DM, RM_{SS}, RM_{MAR}, and 2 d.f. Wald tests are examined in both ZIG and ZILN frameworks. For these models, AIC for the ZIG regression was lower than for the ZILN regression framework; assuming ZIG, AIC = -130.64 while assuming ZILN, AIC = -125.9.

Some of the metrics find significant differences between groups adjusting for gender while others do not. When assuming the ZIG, the DM, RM_{SS}, and RM_{MAR} tests all do not show statistically significant differences at the 0.05 significance level between PD and controls in terms of LDSS. However, DM is close to significant with p-value=0.0513, with PD subjects having an LDSS 0.00276 higher on average than the average LDSS for controls, 95% C.I. [-0.00002, 0.00553]. Assuming ZILN, on the other hand, DM and RM_{SS} are still not significantly different from 0 and 1 respectively, but the RM_{MAR} results indicate that on average PD subjects have an LDSS that is significantly higher than that of control subjects; $RM_{MAR} = 15.5278$ 95% C.I. [2.27843, 105.82]. As was the case in Section 5.2.1, the 95 % C.I. are much wider for ZILN analyses. Simple linear regression did not find significant effects for

	Estimate (95% CI)	Test Statistic[1]	p-value
SLR	0.00260 [-0.00055, 0.00576]	T = 1.63	0.1047
RTR	- - -	T = 3.48	0.0007
ZIG DM	0.00276 [-0.00002, 0.00553]	Z = 1.95	0.0513
ZIG RM_{SS}	4.18644 [0.11829, 148.16]	Z =0.79	0.4313
ZIG RM_{MAR}	4.23732 [0.83624, 21.471]	Z =1.74	0.0812
ZIG 2 d.f. Wald	- - -	$\chi^2_2 = 9.76$	0.0076
ZILN DM	0.00784 [-0.00112, 0.0168]	Z = 1.71	0.0864
ZILN RM_{SS}	15.4423 [0.37562, 634.86]	Z = 1.44	0.1489
ZILN RM_{MAR}	15.5278 [2.27843, 105.82]	Z = 2.80	0.0051
ZILN 2 d.f. Wald	- - -	$\chi^2_2 = 12.53$	0.0019

Table 5.3: Estimates, 95% confidence intervals (CI), test statistics, and p-values for tests comparing PD and Control subjects, adjusting for gender.

[1]SLR and RTR test statistics are distributed as a t-test with d.f.=124.

group when adjusted for gender. RTR found a significant effect of group such that the ranks of LDSS tended to be higher for PD subjects than for controls. The two-part Wald tests assuming ZIG and assuming ZILN were both significant for group effects adjusted for gender.

Table 5.4 presents the test statistics, p-values, and appropriate estimates and 95% confidence intervals for the effect of gender adjusted for group. For ZIG regression, DM, RM_{SS}, and RM_{MAR} all concur that there is a significant effect of gender on the overall gender-specific means of LDSS. Specifically, males have an LDSS that is 0.00263 higher than the average LDSS of females when adjusting for group, 95% C.I. [0.00030, 0.00496]. Recalling Figure 5.1, this means that the percent of the area between the gray lines representing the left and right ties that is outside of the subject's lane is on average higher in males by an addition of 0.263% compared to the

	Estimate (95% CI)	Test Statistic[1]	p-value
SLR	0.00216 [-0.00127, 0.00559]	T = 1.25	0.2142
RTR	- - -	T = -0.03	0.9778
ZIG DM	0.00263 [0.00030, 0.00496]	Z = 2.21	0.0270
ZIG RM_{SS}	5.18195 [1.12975, 23.769]	Z = 2.12	0.0343
ZIG RM_{MAR}	5.25988 [1.33612, 20.706]	Z = 2.37	0.0176
ZIG 2 d.f. Wald	- - -	$\chi^2_2 = 8.39$	0.0151
ZILN DM	0.00268 [-0.00468, 0.01005]	Z = 0.71	0.4751
ZILN RM_{SS}	1.77440 [0.29684, 10.607]	Z = 0.63	0.5296
ZILN RM_{MAR}	1.81140 [0.34807, 9.427]	Z = 0.71	0.4802
ZILN 2 d.f. Wald	- - -	$\chi^2_2 = 0.9720$	0.6151

Table 5.4: Estimates, 95% confidence intervals (CI), test statistics, and p-values for tests comparing male and female subjects adjusting for group.

[1] SLR and RTR test statistics are distributed as a t-test with d.f.=124; DM, RM_{SS}, and RM_{MAR} tests are standard normal tests; and ZIG and ZILN 2 d.f. Wald tests are χ^2_2 tests.

percent of the area outside of the lane for females. RM_{SS} for gender assuming ZIG is 5.18, 95% C.I. [1.13, 23.77]. RM_{SS} reflects a weighted average across PD and control groups of the average ratio of LDSS for males given group over the average LDSS for females given group; the weighting was done in terms of the number of subjects per group. RM_{MAR} for gender equals 5.26; this means that the average LDSS for males is 5.26 times as high as the average LDSS for females. When ZILN regression is assumed, no significant gender effects are found using any of the mean comparison metrics. Simple linear regression and rank transformed regression did not find significant effects for gender adjusted for group. The two-part Wald tests assuming ZIG reports a significant difference in probability of non-zero outcome or value of

			Test Statistic	p-value
Group	$P(Y > 0)$ post-test	$W(\beta_1)$	Z = 3.114	0.0009
	$Y \mid Y > 0$ post-tests	ZIG $W(\tau_1)$	Z = 0.238	0.4057
		ZILN $W(\tau_1)$	Z = 1.681	0.0464
Gender	$P(Y > 0)$ post-test	$W(\beta_1)$	Z = -0.378	0.3524
	$Y \mid Y > 0$ post-tests	ZIG $W(\tau_1)$	Z = 2.872	0.0020
		ZILN $W(\tau_1)$	Z = 0.910	0.1814

Table 5.5: Post-tests as follow-up to the 2-part tests for group and gender adjusting for each other.

the non-zero outcome between genders whereas ZILN did not find any significant difference.

Since Tables 5.3 and 5.4 found some significant group effects in the 2 d.f. tests, the post-tests for them were performed and are presented in Table 5.5. From this we can see that, the significant effect for group found in the 2 d.f. tests is due to a difference in the binomial part of the model which shows that PD subjects are more likely to go outside of the lane than are control subjects (p-value=0.0009). In the ZIG model, there is no significant effect of group in the continuous part of the model (p=0.4057), but if we assume the ZILN model the PD subjects who go outside of there lane are also likely to do so to a greater extent than the control subjects who go outside of the lane (p=0.0464). For gender on the other hand, the differences between male and females when found was found in the continuous part of the model. Males and females have similar probabilities of going outside of the lanes (p-value=0.3524), but when ZIG is assumed, male subjects who go outside of the lane do so to a greater extent than female subjects who go outside of the lanes (p-value=0.0020).

5.2.3 Two Group Comparison Adjusted for Age

As was done in Section 5.2.2, the following section also utilizes the methods outlined in Chapter 4 for adjusting for covariates, but in this case we will be adjusting for the continuous covariate of age. The correlation between group and age in our sample is slight, spearman's ρ=0.1562 (p-value=0.0795). As in Section 5.2.2 the observed distribution of the covariate, in this case age, is used for the marginalization of DM, RM_{SS}, and RM_{MAR}. DM measures the average difference between group 1 and group 0 averaged over the ages observed in our sample. RM_{MAR} is the ratio of the expected group means given the observed age distribution. RM_{SS} is the average of the subject specific ratio of means calculated at each age observed. Figure 5.2 is a histogram showing the distribution of age combined between the two groups. This is the distribution of age that was used to calculate the marginal means for both groups and the resultant DM and RM_{MAR} values. This distribution was also used in calculating RM_{SS}. Ages found in our study range from 43 to 84 with most of the subjects in their late 50's, 60's, or 70's (first quartile of age is 59.7, third quartile 72.3).

From Table 5.6 it is seen that in terms of the mean-based tests for $DM = 0$, $RM_{SS} = 1$, and $RM_{MAR} = 1$ when a ZIG model adjusting for age is assumed the group effects observed are not significant. When ZILN is assumed, the ratio of the marginalized means RM_{MAR} is significantly greater than 1, specifically, the overall mean for LDSS averaged over the observed ages is 16.29 times greater for PD subjects than it is for control subjects (95% C.I.: [2.55, 103.95]). The effect of group is significant for SLR, RTR, and ZIG and ZILN 2-part Wald tests. According to SLR, when adjusting for age, PD subjects have higher LDSS scores on average than do control subjects. RTR shows that PD subjects have higher ranks of LDSS when

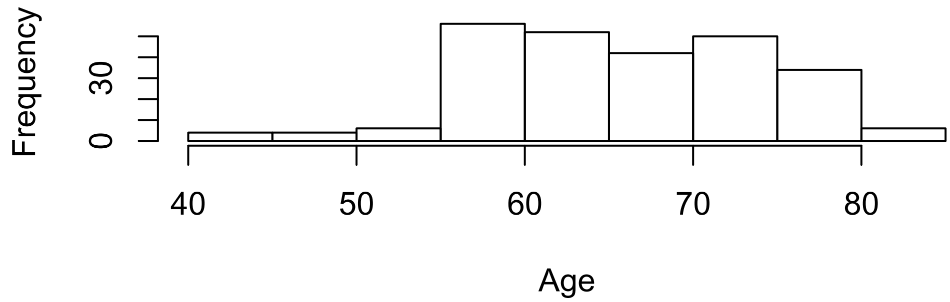

Figure 5.2: Overall age distribution

	Estimate (95% CI)	Test Statistic	p-value
SLR	0.00313 [0.00019, 0.00607]	T = 2.11	0.0370
RTR	- - -	T = 3.61	0.0004
ZIG DM	0.00287 [-0.01045, 0.01619]	Z = 0.42	0.67300
ZIG RM_{SS}	4.06334 [0.16092, 102.61]	Z = 0.85	0.39474
ZIG RM_{MAR}	4.01053 [0.78557, 20.475]	Z = 1.67	0.09495
ZIG 2-part Wald	- - -	$\chi^2_2 = 9.97$	0.0068
ZILN DM	0.00817 [-0.00113, 0.01748]	Z = 1.72	0.08508
ZILN RM_{SS}	16.4879 [0.58054, 468.27]	Z = 1.64	0.10069
ZILN RM_{MAR}	16.2935 [2.55389, 103.95]	Z = 2.95	0.00316
ZILN 2-part Wald	- - -	$\chi^2_2 = 13.48$	0.0012

Table 5.6: Estimates, 95% confidence intervals (CI), test statistics, and p-values for tests comparing PD and Control subjects adjusting for age.

[1] SLR and RTR test statistics are distributed as a t-test with d.f.=124; DM, RM_{SS}, and RM_{MAR} tests are standard normal tests; and ZIG and ZILN 2 d.f. Wald tests are χ^2_2 tests.

adjusting for age than do control subjects. And according to ZIG and ZILN two-part Wald tests adjusting for age, PD and control subjects differ in terms of probability of non-zero outcome and/or the average on the non-zero outcomes. None of the tests used find a significant effect of age on LDSS when adjusting for group.

Following up on the significant 2 d.f. tests, Table 5.7 contains results of post-tests for group adjsuted for age. From Table 5.7 there is a clear effect of group in the binomial portion of the models, indicating that the probability of going outside of the lane in the 60 second fog segment is higher for PD subjects than it is for control subjects. For the post-test for a group effect on the extent outside of the lane given subject goes outside of the lane, different conclusions are reach depending on the model assumed. When ZIG regression is assumed, the effect of group on the value of the positive LDSS scores is not significant. On the other hand, when ZILN regression is assumed the conclusion is that in addition being more likely to go outside of the lane, PD subjects also have a higher LDSS given LDSS>0 than do controls. These adjusting for age regression models have higher AIC values than those adjusting for gender. Also, there is little difference between AIC for ZIG and ZILN, although ZILN is slightly preferred; AIC given ZIG regression equals -123.97 and AIC given ZILN regression is -125.83. Although mean-based tests for age have not yet been developed, 2 d.f. tests can be used straightforwardly. The 2 d.f. tests for age adjusted for group found no significant effect of age.

		Test Statistic	p-value
$P(Y>0)$ post-test	$W(\beta_1)$	Z = 3.143	0.0008
$Y\|Y>0$ post-tests	ZIG $W(\tau_1)$	Z = 0.295	0.3841
	ZILN $W(\tau_1)$	Z = 1.897	0.0289

Table 5.7: Post-tests for group adjusted for age as follow-up to the 2-part tests.

5.2.4 Lane Departure Severity Score Compared to Other Metrics

In addition to driving in a simulator, PD and control subjects performed a series of visual, cognitive, and motor tasks. This section compares LDSS to other driving measures and relates LDSS to performance on these above tasks using only the PD subjects.

The battery of tasks are explained in more detail by Uc et al.[25] The subjects completed basic visual tasks such as near and far visual acuity (Near VA and Far VA), and contrast sensitivity (CS); visual perception tasks such as motion perception measured by structure from motion (SFM), attention measured by useful field of view (UFOV), and spatial perception measured by judgement of line orientation (JLO); and visual cognition tasks such as blocks, complex figure test (CFT) copy and recall, and Benton visual retention test. Executive function was measured using the trail making test (TMT) and controlled oral word association (COWA). Verbal memory was tested using auditory verbal learning tests (AVLT) and general cognition measured using the mini mental status examination (MMSE). Depression was assessed using the geriatric depression scale (GDS). Sleepiness was measured before the drive using the Epsworth sleepiness scale (ESS) and during the drive using the Stanford sleepiness scale (SSS). Motor function was measured through finger tapping (Tap), a 7 meter walk (Walk), and a functional reach task (Reach). Additionally, for the PD subjects, we noted the disease duration, Hoehn-Yahr stage, and their Levodopa equivalent medication. The disease status was also rated using the Schwab-England score and the unified Parkinson's disease rating scale (UPDRS) for both motor and activities of daily living (ADL).

Table 5.8 compares LDSS to other vehicle control measures using only the subjects with Parkinson's disease. Pearson correlations between LDSS and the other

	Correlation	p-value
Speed	0.1295	0.2649
Speed variability	0.0549	0.6374
Steering variability	0.6147	< 0.0001
Standard deviation of lane position (SDLP)	0.7197	< 0.0001
Number of lane deviations	0.8609	< 0.0001

Table 5.8: Pearson correlation for LDSS with other vehicle control measures using PD subjects.

		LDSS		SDLP		Lane deviations	
	n	Corr.	p-value	Corr.	p-value	Corr.	p-value
Disease Duration	68	0.1074	0.3835	0.2371	0.0515	0.1837	0.1337
Hoehn-Yahr stage	75	0.1329	0.2558	0.1048	0.3710	0.1435	0.2195
UPDRS-ADL	75	0.1006	0.3906	0.2165	0.0621	0.0664	0.5715
UPDRS-motor	75	0.1040	0.3748	0.2385	0.0394	0.1732	0.1373
UPDRS-total	75	0.1776	0.1275	0.2901	0.0116	0.2069	0.0749
Schwab-England	73	-0.3415	0.0031	-0.1845	0.1182	-0.2857	0.0143
Levodopa equivalent	68	0.0762	0.5371	0.1159	0.3466	0.0702	0.5697

Table 5.9: Pearson correlations between lane control measures and PD specific measures, using only PD subjects.

measures are presented. From this table, it is clear that LDSS is highly correlated to other lane keeping variables, specifically steering variability, standard deviation of lane position (SDLP), and lane deviation count. The greatest correlation is between LDSS and lane deviation count which makes sense as LDSS was devised as an extension of lane deviation count.

	n	LDSS Corr.	LDSS p-value	SDLP Corr.	SDLP p-value	Lane deviations Corr.	Lane deviations p-value
Age	75	0.2140	0.0653	0.0493	0.6747	0.1219	0.2975
Education	74	0.2046	0.0804	0.1738	0.1386	0.2843	0.0141
Near VA	71	-0.1134	0.3462	-0.1284	0.2859	-0.0499	0.6797
Far VA	72	-0.0267	0.8236	0.0888	0.4581	0.0509	0.6714
CS	75	-0.1799	0.1224	-0.2489	0.0313	-0.2533	0.0284
SFM	65	0.1763	0.1601	0.4625	0.0001	0.4333	0.0003
UFOV	74	0.4202	0.0002	0.3091	0.0074	0.4204	0.0002
JLO	75	-0.0109	0.9260	-0.1473	0.2072	-0.0671	0.5671
Blocks	73	-0.3815	0.0009	-0.3425	0.0030	-0.3659	0.0015
CFT-Copy	73	-0.3983	0.0005	-0.3540	0.0021	-0.4260	0.0002
CFT-Recall	73	-0.2215	0.0596	-0.2545	0.0298	-0.2206	0.0607
BVRT-Error	74	0.3417	0.0029	0.2883	0.0127	0.3128	0.0067
TMT(B-A)	74	0.3583	0.0017	0.3325	0.0038	0.4021	0.0004
COWA	75	-0.1083	0.3549	-0.1415	0.2259	-0.1411	0.2273
AVLT-Recall	75	-0.2286	0.0486	-0.1664	0.1535	-0.1883	0.1058
MMSE	75	-0.3172	0.0056	-0.2825	0.0141	-0.2692	0.0195
GDS	73	0.1178	0.3209	0.2537	0.0303	0.1567	0.1855
ESS	72	0.0024	0.9842	0.0348	0.7718	-0.0204	0.8649
Tap	73	-0.0842	0.4791	-0.2057	0.0808	-0.1160	0.3286
Walk	68	0.0312	0.8008	-0.0406	0.7422	0.0462	0.7085
Reach	75	-0.1513	0.1950	-0.3060	0.0076	-0.2649	0.0216

Table 5.10: Pearson correlations between lane control measures and various visual, cognitive, and motor tasks using only PD subjects.

The correlations between lane keeping variables (LDSS, SDLP, and lane deviations) and Parkinson's disease related measures are presented in Table 5.9. Many of these PD related metrics are not significantly correlated with LDSS. The Schwab-England score, however, is significantly correlated to LDSS with a correlation of -0.3215 (p=0.0031).

Finally, Table 5.10 presents the Pearson correlations and related p-values for the remaining visual, cognitive, and motor tasks presented to our PD subjects. Several of these tests are significantly correlated to LDSS. Specifically, visual perception measured by UFOV; visual cognition measured by blocks, CFT-copy, and BVRT-error; executive function measured by trail making test; verbal memory measured by AVLT-recall; and general cognition measured by MMSE. Motor tests, as well as other visual tests, were not correlated to LDSS.

5.3 Model Fitting

We now consider visual methods for the model fit of the continuous part of the two-part models. Since we are using for log-normal regression simple linear regression on the log of the outcome variable, standard model fitting checks such as Q-Q plots and residual plots for can be straightforwardly used. We will focus on Q-Q plots and the patterns seen which indicate when assuming a ZILN distribution may be appropriate or when a ZIG distribution might, alternatively, be considered.

5.3.1 Quantile-Quantile Plots Comparing Gamma and Log-Normal Models

Q-Q plots can be constructed using the non-zero responses to evaluate the fit of the continuous part of two-part models. Specifically, the conditional residuals for the non-zero outcomes modeled as log-normal can be used to examine assuming ZILN is appropriate, and if a ZIG distribution could be considered. To create the example

Q-Q plots shown in Figure 5.3, four data sets were simulated. The plots on the left show Q-Q plots for data simulated from ZIG distributions and those on the right are for data simulated from ZILN distributions; the top plots were simulated with $\nu^{-1} = 0.5$ and the bottom with $\nu^{-1} = 2$. For these simulations, setting 'I' from Chapter 4 Table 4.6 was used. This setting included both group and gender effects in only the continuous part of the model. Outcomes were simulated for 200 subjects per data set; the level of zero inflation was 0.5 leading to an expected value of 100 subjects with non-zero outcomes in each data set. Each data set was then analyzed assuming ZILN. Then for subjects with non-zero outcomes, conditional residuals were

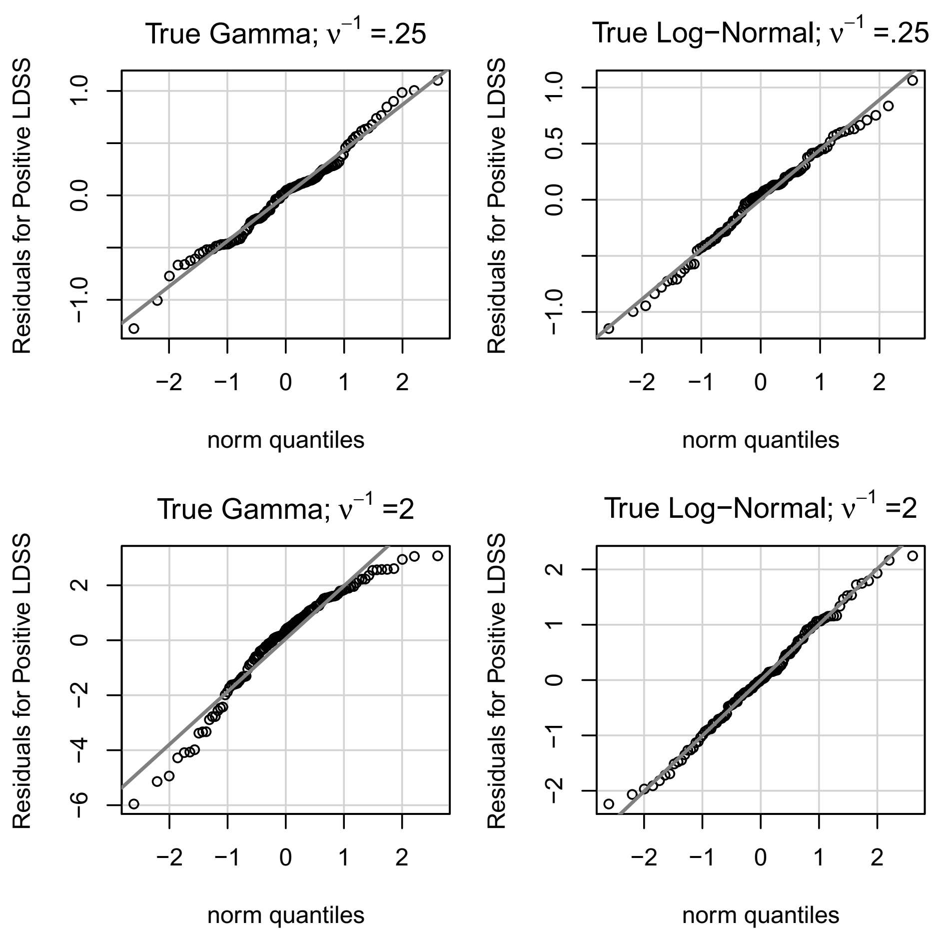

Figure 5.3: Q-Q plots for gamma and log-normal data assuming log-normal distributions.

calculated as the observed outcome minus the expected outcome given the outcome was non-zero, $\log(Y) - E(\log(Y)|Y > 0, X_1 = x_1, Z_1 = z_1)$. This residual was then used to create the Q-Q plots seen in Figure 5.3. These plots were created in R using the qqPlot function in R.[26]

When the continuous data are truly from a gamma distribution, the Q-Q plots of from the log-normal analyses will have a concave curve. This results in part from gamma distribution having more extremely low values compared to the log-normal. The curvature is greater and more obvious when ν^{-1} is large. At smaller values of ν^{-1},gamma and log-normal distributions are less distinguishable.

Figure 5.4 shows Q-Q plots for our data example of LDSS in a one minute fog scenario. These plots compared conditional residuals for log(LDSS) to normal quantiles. The plot on the left shows the residuals for the data with non-zero LDSS values including for an extremely low value. Further follow-up showed that this extremely low value for LDSS was obtained because of machine zero issues for a subject who drove on the lane line briefly, but did not cross it. After discovery, this

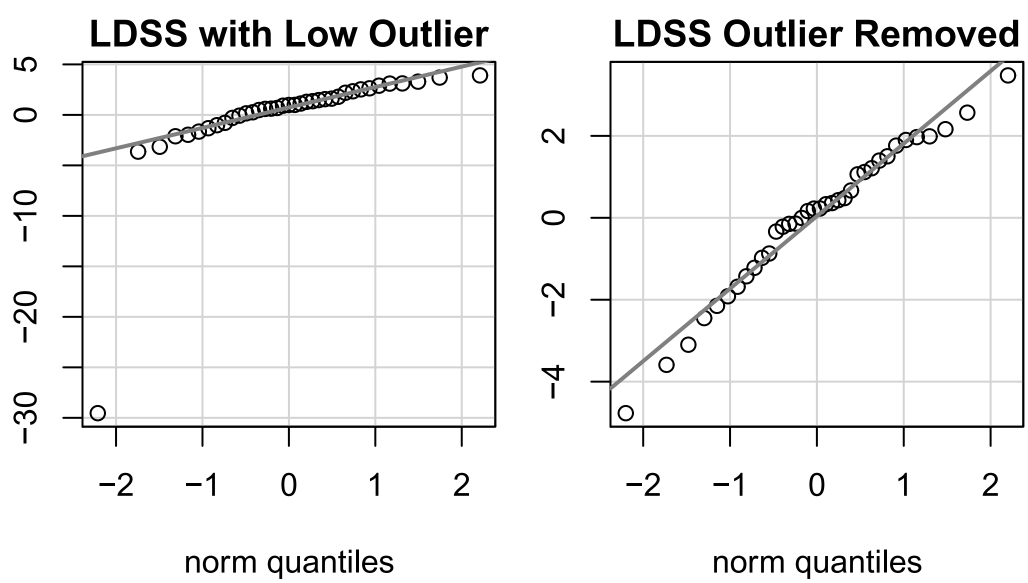

Figure 5.4: Q-Q plots for residuals for non-zero LDSS assuming a ZILN distribution with covariates group and gender, with and without extreme outlier.

value was changed to zero for all analyses. The plot on the right contains a Q-Q plot using for the subjects who went outside of the lanes (LDSS>0). There were 37 subjects with non-zero Lane Departure Severity Scores included in this plot on the right. With so few subjects, it is not obvious whether gamma or log-normal distributions fit this data better. However, it seems that there is a slight concave down curvature to the points which would indicate that gamma might be a better fit than log-normal. Other Q-Q plot examples will be explored in Section 5.3.2.

5.3.2 Extension of Lane Keeping Example to Include a Safety Window

Up to this point we have been considering a subject who does not deviate from the driving lane to be within the law; this is one of the justifications for considering the zeros denoting a subject never leaving the lane as true zeros. However, if one were to shift gears and think of it as a measure solely of driver safety, it may be of interest to explore similar metrics to examine to what extent a subject stays within a slightly narrower area than the lane lines which we will call being within a safety window. Figure 5.5 shows the path of three example drives. In addition to the regular lane lines, smaller artificial lane lines are shown to denote the narrower lane that the subject would need to be driving in to have a certain safety window. The actual lane was 3.6 meters wide, the narrower lane superimposed on this plot is instead 2.8 meters wide allowing for a safety window of .4 meters before the lane lines on both sides of the vehicle. The departure from this narrower lane will be called LDSS28 to denote the 2.8 meter lane width being assumed. In examining the sensitivity of LDSS with a safety window to the choice of safety window, we analyzed our data with lane widths ranging from 2.2 meters (LDSS22) to 3.4 meters (LDSS34) in addition to the true lane width of 3.6 meters (LDSS).

Some of the subjects who did not exit the lanes for the duration of the segment

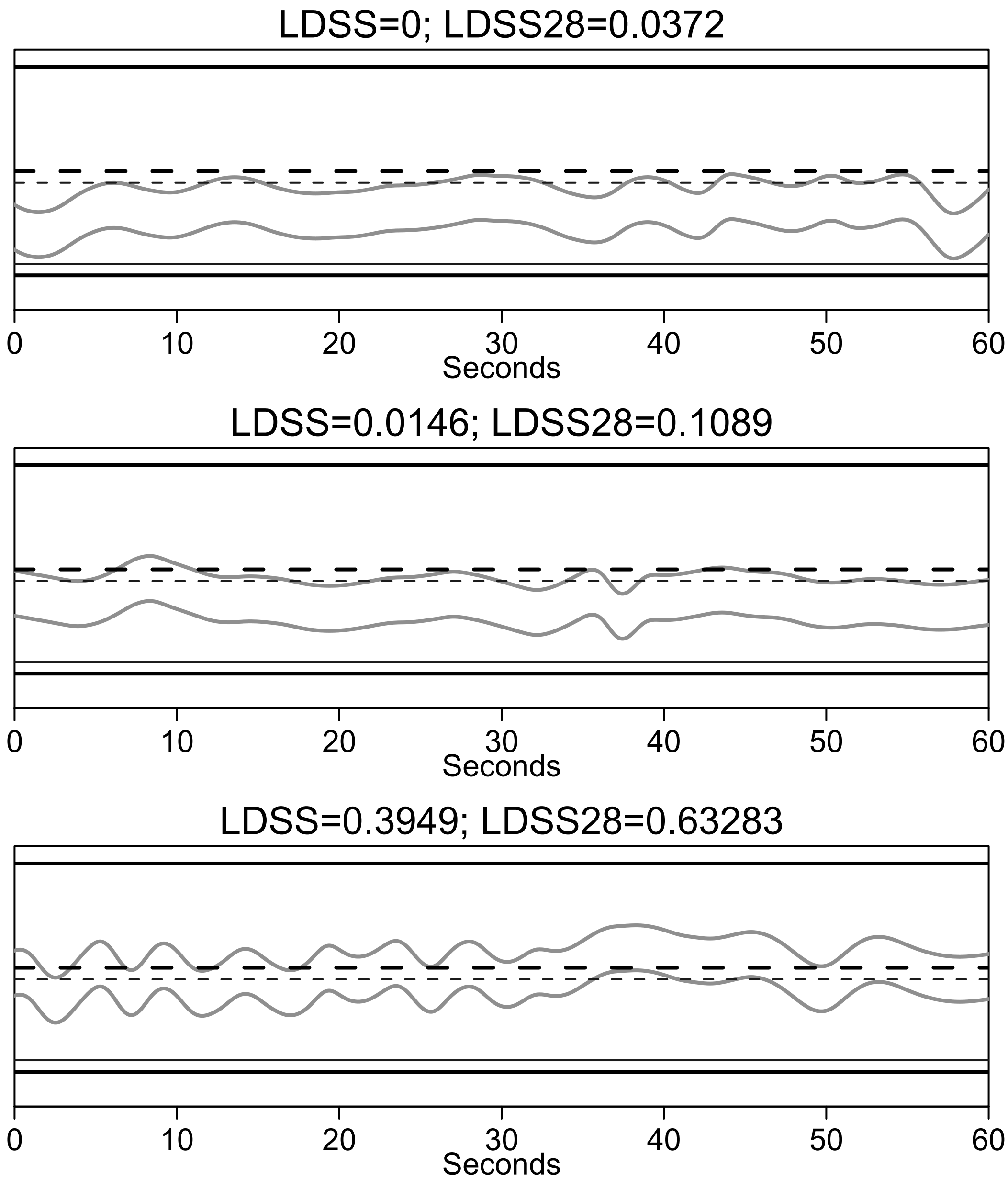

Figure 5.5: Illustration of LDSS and LDSS28 over a 60 second segment

(LDSS=0) drove close enough to the lane lines that that they will have gone outside of a safety window. This can be seen by in the first plot in Figure 5.5 where the gray lines denoting the wheels of the vehicle do not cross the larger lines that indicate the true lanes, but do cross the smaller dashed lane which indicates the edge of the safety window, leading to an LDSS28 of 0.0372. As the required safety window increases, the number of subjects with non-zero values will also increase, potentially changing the distribution of the non-zero outcomes as many of these new non-zero values will be small. For subjects who were already outside of the true driving lane, the portion outside of the new smaller lanes will be higher leading to higher LDSS scores as safety window gets larger and the 'lane width' gets smaller. This can be seen the bottom two plots in Figure 5.5 where LDSS28 is greater than the original LDSS.

In Section 5.4, we noted that the model for LDSS using group and gender as covariates may have been the most appropriate model given our data. One reason for this selection was because using these covariates and ZIG regression yielded the lowest AIC values of all of the proposed models. Another was that a gender variable was significant in at least some cases whereas age had not been significant using any of the applicable metrics proposed. For this section, we will present results from analyses of LDSS22 to LDSS34 and the original LDSS when using the covariates group and gender.

Figure 5.6 shows test statistics for the effect of group adjusted for gender using the metrics we have proposed in this dissertation. The top set of plots show the test statistics for the ZIG and ZILN mean-based tests; DM test statistics are symbolized by 'D', RM_{MAR} test statistics are denoted by 'M', and RM_{SS} test statistics by 'S'. The plot on the top left shows the ZIG mean-based test statistic results and the plot on the top right shows the ZILN mean-based test statistic results. In these two plots, all test statistics are distributed as a standard normal; p-value cut-offs are show at $\alpha = 0.05$, $\alpha = 0.01$, $\alpha = 0.001$, and $\alpha = 0.0001$ by gray dotted lines. The

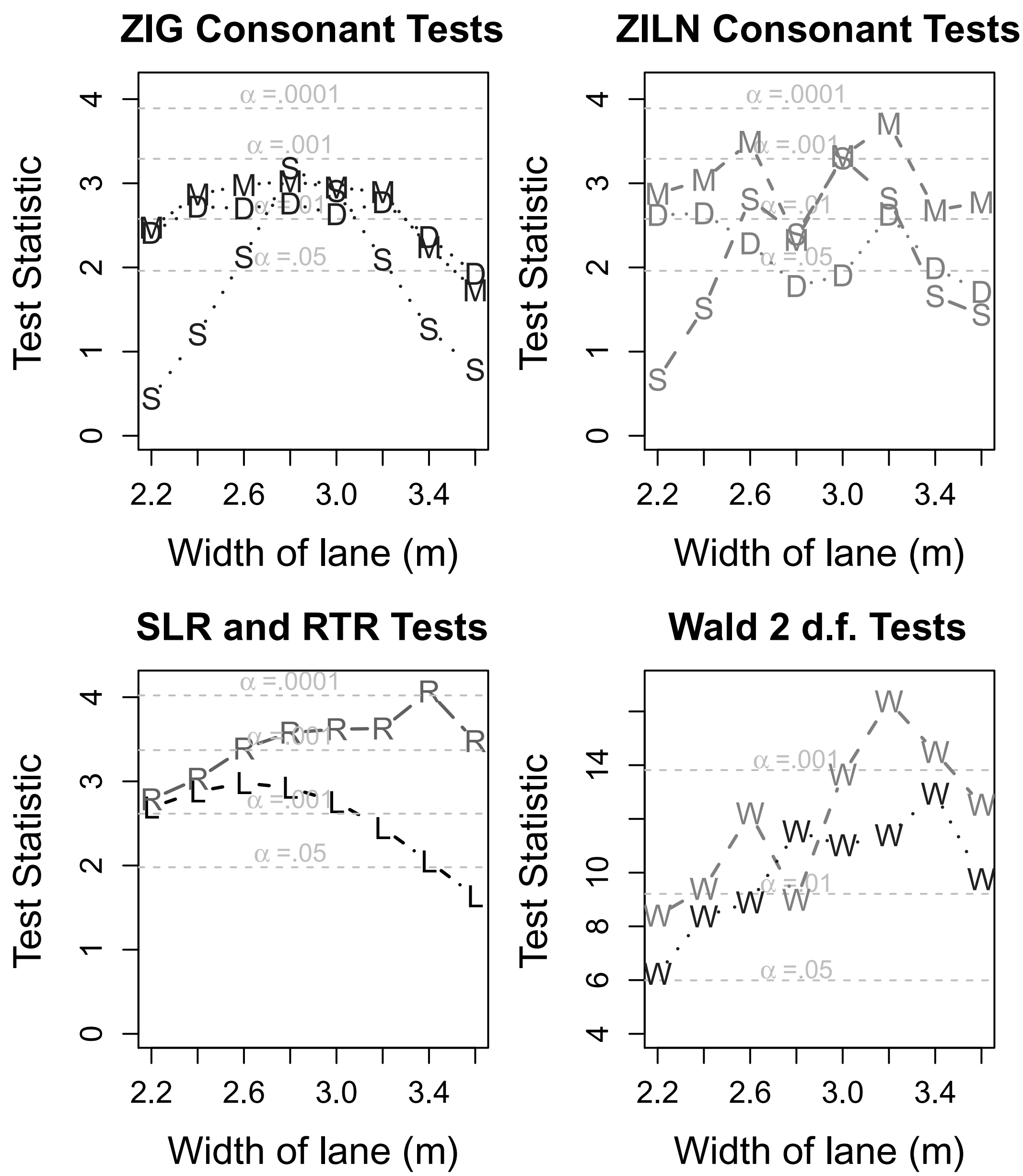

Figure 5.6: Test statistics for group effect on LDSS using various lane widths.

bottom left plot shows test statistics for simple linear regression, denoted by 'L', and rank transformed regression, denoted by 'R'. These test statistics are distributed as Student's t with d.f.=124. Finally, the plot on the bottom right contains the test statistics for the 2 d.f. Wald tests resulting from assuming ZIG distribution in dark gray, and assuming a ZILN distribution in the lighter gray.

For ZIG analyses, the test statistic for RM_{SS} increases steeply until a lane width of 2.8 m and then decreases steeply. The changes in test statistic for DM and RM_{MAR} occur much more gradually. For all lane widths smaller than the true lane width of 3.6 m, DM and RM_{MAR} based tests yielded significant group effects with p-values less than 0.05 and for lane widths from 2.4 m to 3.2 m DM and RM_{MAR} tests resulted in p-values less than 0.01. For ZILN, on the other hand, test statistic values were much more sporadic. Although some are higher than those for ZIG regression, others were lower. From the bottom left plot, the test statistics for RTR tend to increase with lane width but show a significant group effect for all lane widths. Test statistics for SLR increase slightly as lane width increases until about 2.6 m decreasing after that; all SLR tests found a significant group effect except for the true lane width. Both ZIG and ZILN 2 d.f. tests were significant at all lane widths. Both ZIG and ZILN tests generally increase in significance as lane width increases except for in a couple of sporadic places and at the larger lane widths when the significance decreases again.

Figure 5.7 shows test statistics for the post-test for the 2 d.f. tests at the various lane widths. This post-test plot shows the significance separately of the two parts of the model. This elucidates which part of the model lead to the overall significance of the two-part tests. 'B' indicates the 1 d.f. test for the binomial part of the model, 'G' indicates the 1 d.f. test for the continuous part of the model assuming ZIG, 'N' indicates the 1 d.f. test for the continuous part of the model assuming ZILN. As lane width increases, the number of subjects with non-zero values decreases, and the

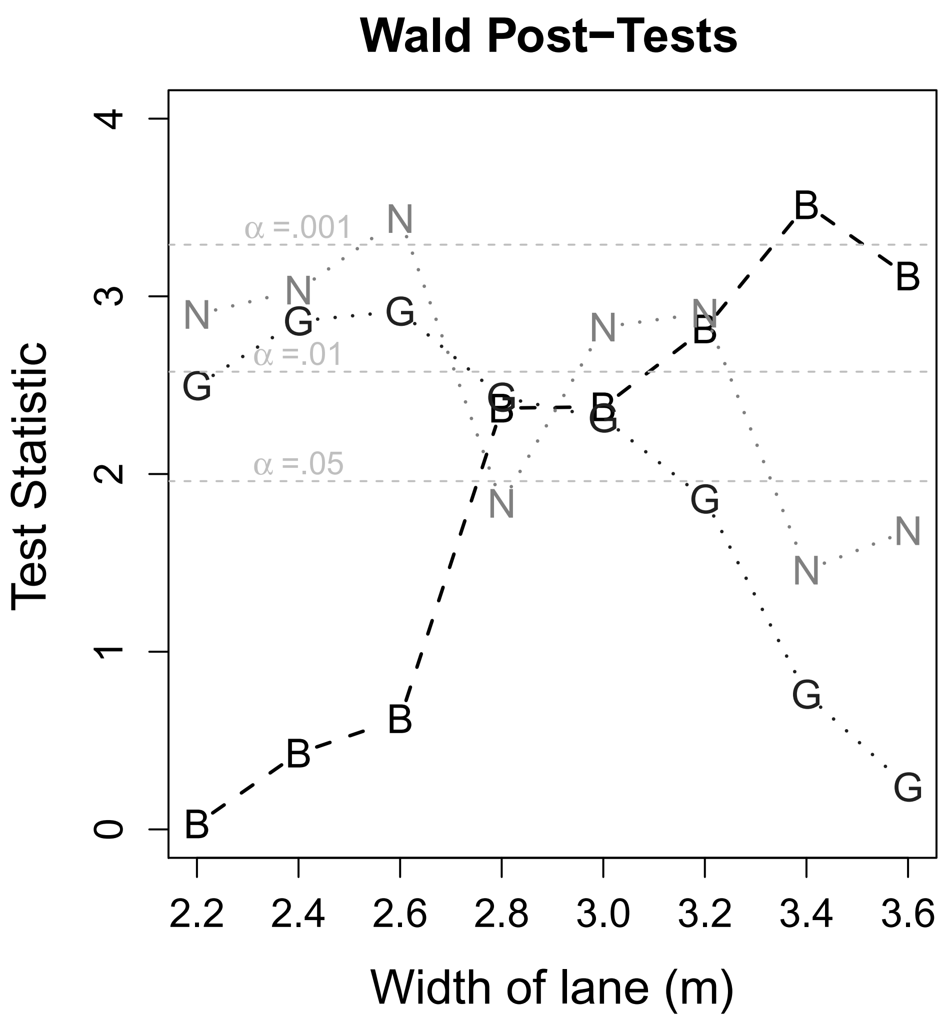

Figure 5.7: Post-tests for group effect on LDSS using various lane widths

significance of the binomial part of the model increases until the change from 3.4 m to 3.6 m where the test statistic decreases slightly. When the new lane width is 2.6 m or smaller (safety window of .4 m or more on either side of the vehicle) there is no significant difference observed in the probability of a non-zero outcome. At these points, most subjects exit these imaginary smaller lanes at some point during the segment leaving little room for distinguishing between the probabilities of those that do so at those lane widths. The significance of the continuous part of the model follows essentially an opposite trend with a general decrease in significance as the lane width widens towards the true lane lane width. At the narrower lane widths most subjects have non-zero values leading to a greater power for finding such differences.

Figure 5.8 shows Q-Q plots for various levels of safety windows. As larger safety windows (smaller lane widths) are used, the non-zero sample size increases and the lack of fit for log-normal is clearer. Note that the distribution of the non-zero values is also changing; specifically, it is shifting to the right while adding more small values. At LDSS28 and LDSS24 there is a clear concave pattern to the Q-Q plots indicating that ZIG might be a better model than ZILN. The slight concavity of the Q-Q plot for LDSS or the low outliers seen in LDSS32 could be used to argue that ZIG is a better fit than ZILN, but is not as clear in those cases as it is for LDSS 28 and LDSS24. (It is possible that other distributions could lead to this concavity of the Q-Q plots, but that is beyond the scope of this present dissertation.) This result is also consistent with AIC results which find that ZIG models have lower AIC for all levels of safety window, but the distinction between ZIG and ZILN becomes stronger at larger safety windows, particularly for lane widths of 3.0 m or smaller.

5.4 Summary

In this chapter, driving outcome LDSS was defined in further detail. Then the methods of this dissertation were applied to a sample of driving data in order

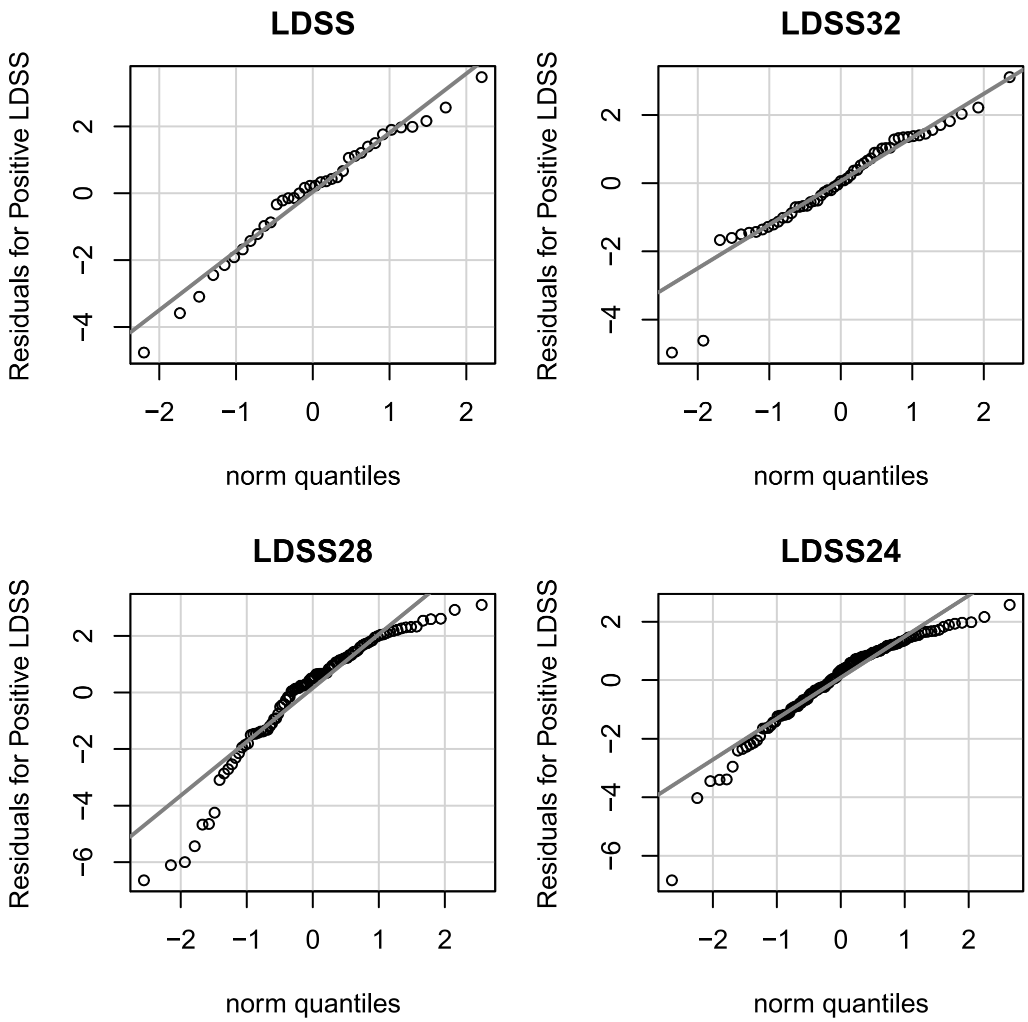

Figure 5.8: Q-Q plots for LDSS at selected lane widths

to compare the driving safety of PD and control subjects. We performed analyses without covariate adjustment as well as analyses that adjusted for gender and analyses that adjusted for age. Finally, the fit of the log-normal model for our data set was examined and LDSS related driving metrics that look at the extent to which a driver allows for a safety window were briefly explored.

In analysing this data set we found that adjusting for age may be unnecessary in this study as the correlation between age and group is small and there are no significant effects of age on LDSS. Gender should be included as an adjusting covariate for group comparisons of LDSS. This is evident as gender and group are correlated, and, for at least some of the tests explored, gender had a significant effect on LDSS. Also, a ZIG model with gender and group had the lowest AIC of all of the models. That model found a marginally significant difference between group means, but found significant differences in gender in terms of mean differences and both ratio of means metrics. The two-part tests for this model, however, found significant effects for both group and gender. For the group effect, the significant post-test was in the binomial part of the model, PD subjects had a higher probability of going outside of the lane (higher probability of LDSS>0). The effect of gender, on the other hand, was found in the continuous part of the model, males did not exit the lane more frequently than females, but when they exited they did so to a greater extent (resulting in a higher LDSS given LDSS>0). This difference of the location of the effects may be in part the reason that gender effects were found in the mean-based tests but group effects were not (even though the estimate for DM of group was higher than the estimate for DM of gender).

A small simulation found that when data are truly ZILN, a Q-Q plot created from the conditional residuals for the subjects with non-zero outcomes using ZILN regression follow the appropriate linear form. We also found that when the data are simulated from a ZIG distribution but a ZILN model is run such Q-Q plots have a concave downward form. The Q-Q plots from our data for a model adjusting for gender do not show a conclusive pattern, but seem to suggest that ZILN may not be the best model and that ZIG may be more appropriate. This agrees with the AIC from the model adjusting for gender which was lower when ZIG was assumed.

In relaxing the lane line definitions to create variants of LDSS that require a safety window, we found that the results of the tests assuming ZIG (except for RM_{SS}) only changed gradually as lane width changed. On the other hand, when ZILN was assumed, the results rose and fell unpredictably as the lane width changed. This sporadic nature of the test results when assuming ZILN may be due in part to ZILN models not fitting these data well. When the coefficient of variation is high, the differences between ZIG and ZILN distributions are such that ZILN distributions do not fit the data well. Coefficient of variation estimates for ZILN range greatly as lane width is changed which is likely responsible for the instability of the log-normal regression results when coefficient of variation is high and the model fit is poor. Given the results of Q-Q plots, AIC, and the greater stability of the ZIG results, it seems that ZIG could be preferred for these data.

CHAPTER 6
CONCLUSIONS

In this dissertation we have explored analysis methods for non-negative data containing zero values, called semicontinuous data. These semicontinuous data can be modeled through a two-step process which involves modeling the probability of a nonzero outcome and modeling the mean of the nonzero outcomes. This type of model for semicontinuous data has been used frequently in economic data, but has also been used in monitoring weather, in microbiology, and in medical applications. In this dissertation, we have addressed methods to combine the two parts of these models under both ZIG and ZILN frameworks.

This dissertation has combined the parts of the two-part models in ways that address one of two hypotheses. First, in Chapter 2, methods were introduced which look for a difference between the two groups explicitly in terms of the probability of a nonzero outcome and the conditional means of the two groups given a nonzero outcome. Global 2 d.f. tests were used to look for group differences. When the 2 d.f. tests are significant, post-tests can be used (as shown in Chapter 5) to determine whether the groups differed in terms of the probability of a nonzero outcome or in terms of the conditional group means of the nonzero outcomes, or both. These two-part tests were extensions of the two-part tests proposed by Lachenbruch [14], who proposed adding together two χ^2_1 tests to create a global χ^2_2 test in order to test for the group differences in probability and conditional mean simultaneously. In Chapter 2 we compared two-part tests based on ZIG and ZILN models obtained by score, LRT, and Wald methods to Lachenbruch's proposed Wald-Wilcoxon two-part test.

Secondly, the two-part model can be combined to create estimates and tests based on the overall mean of the semicontinuous outcome given group. While the two-part tests were looking for any differences in probability or conditional mean between the two groups, mean-based tests compare the overall means of the semicontinuous outcome for the two groups. In some instances the effect of group on the probability of a non-zero outcome and the effect of group on the value of a nonzero outcome may have opposite effects on the overall mean. We called these effects dissonant effects. When the null hypothesis is in terms of the mean, these can at times cancel out so that there is no difference in the overall mean of the semicontinuous outcome for the two groups. When such effects may exist, it is particularly important to determine which hypothesis is of interest. In these situations, mean-based tests could correctly find no difference in the overall means of two groups while two-part tests could correctly find significant differences between the two groups. In practice, one must determine if one is interested in a mean-based comparison (e.g. for decision making purposes) or in finding a differences in the distribution of the outcome, in terms of the probability of a nonzero outcome and the expected value of the nonzero outcomes.

References for the mean-based tests that we have used in this dissertation are outlined in Table 6.1. In Section 3.3.1 we outlined a DM test based on a ZIG. This test was the same as that proposed by McLerran [12]. Tu and Zhou [1] proposed a Wald test for the comparison of the means of multiple groups allowing for CoV to vary with group membership. The DM test we proposed in Section 3.3.2 could be found using Tu and Zhou's methods assuming only two groups and constraining CoV to be equal for the two groups. In Section 3.4.2 we adapted a RM test assuming ZILN proposed in Zhou and Tu [11]; again they allowed CoV to vary whereas we simplified they test by constraining CoV to be equal for the two groups. The final test used in the non-adjusting framework was a RM test assuming ZIG (Section 3.4.1) which had not been developed elsewhere.

Very little previous work has been done on mean-based tests for semicontinuous data when adjusting for covariates. The only test which we are aware of was based on DM assuming a ZILN distribution. This was proposed by Dominici and

	Section	References
ZIG DM no covariate adjusting	3.3.1	McLerran [12]
ZILN DM no covariate adjusting	3.3.2	Tu and Zhou [1]
ZIG RM no covariate adjusting	3.4.1	NA
ZILN RM no covariate adjusting	3.4.2	Zhou and Tu [11]
ZIG DM adjusting for covariates	4.3.1	NA
ZILN DM adjusting for covariates	4.3.2	Dominici and Zeger [20]
ZIG RM_{SS} adjusting for covariates	4.4.1	NA
ZILN RM_{SS} adjusting for covariates	4.4.1	Tooze et al[1] [13]
ZIG RM_{MAR} adjusting for covariates	4.4.2	NA
ZILN RM_{MAR} adjusting for covariates	4.4.2	NA

Table 6.1: Section and external references and lack thereof for mean-based tests.

[1] Tooze et al created no overall adjusted test but used RM at certain values of adjusting covariates for interpretation purposes only.

Zeger.[20] They proposed and compared several methods of estimating DM for a semicontinuous data including a proposed method based on marginalizing DM over the observed values of the covariates. We used the DM metric they proposed for a ZILN distribution in Section 4.3.2, but differ from them in that they estimated the standard error of DM using a bootstrapping technique whereas we found an approximate standard error for DM using the delta method. In Section 4.3.1 we introduced a DM assuming ZIG which was created by a similar marginalization of DM given the covariate values. In our literature review we found no formal tests for ratio of means adjusting for covariates. However, Tooze et al [13] include a ratio of means at various specified levels of the covariates in their interpretation of their 2-part model (ZILN) results. In Section 4.4 we proposed two methods of adjusting for covariates to create a ratio of means test for both ZIG and ZILN distributions; RM_{SS} in which subject specific ratios were calculated and then marginalized, and RM_{MAR} calculated as the ratio of the marginalized group means. Both of these metrics are completely novel to this dissertation.

In our simulation studies presented in Chapters 2, 3, and 4, we have compared the Type 1 error rates and Power for these tests under ZIG and ZILN frameworks, both when misspecified and when correctly specified. One of the most consistent effects that we found was that when the data were truly ZILN but ZIG was assumed, Type 1 error was elevated regardless of the metric used for the comparison. These elevated Type 1 errors for ZILN data misspecified as ZIG increase with ν^{-1} and also increase with sample size. It must be noted, however, as ν^{-1} increases and as sample size increases it becomes easier to distinguish log-normal regression from gamma regression through Q-Q plots and through AIC.[27] We observed this in our data example in Chapter 5 when the larger safety windows gave us more subjects with nonzero outcomes it was easier to see in the curvature in the Q-Q plot indicating the gamma model as the appropriate fit. It was also clear in the Q-Q plots from the simulated data that the difference between gamma and log-normal was more easily seen when ν^{-1} was large.

Another major simulation result is that when ZIG was the true data distribution but ZILN was assumed, Type 1 errors were appropriate for most metrics but power was typically lower than when ZIG was appropriately assumed. When 2 d.f. tests are used (Chapter 2) the power for ZIG data analyzed assuming ZILN is lower than when that same data are analyzed assuming ZIG when the nonzero effect of group is in the continuous part of the model. This discrepancy between ZIG and ZILN power for detecting effects in the continuous part of the model of ZIG data increases as ν^{-1} increases. This discrepancy exists as the power for detecting such effects decreases as ν^{-1} increases at a faster rate for ZILN than for ZIG analyses. When mean-based tests (DM, RM, RM_{SS}, and RM_{MAR}) are used, power is lower when ZILN is falsely assumed for ZIG data than when ZIG is correctly assumed. This difference between ZIG and ZILN power for ZIG data is the most pronounced for DM than for the other metrics.

We also found that when data simulated from ZIG distributions are correctly analyzed assuming ZIG distributions, there is some conservativeness regardless of metric being tested. For RM, RM_{SS}, and RM_{MAR} this conservativeness is slight. However, for DM, this over conservativeness is much more substantial. For tests based on DM, as sample size increases Type 1 error rates improve; this improvement happens slowly. In follow-up simulations, we found that at very large sample sizes as the sample size increases the Type 1 error rate increases (towards nominal) and that the distribution of p-values seems to be getting closer to uniform showing that at least part of the issue here is slow asymptotics for the DM test.

In Chapter 5 we examined the relationship of LDSS with group, both without covariate adjustment and with adjusting for sex or age. Both mean-based tests and two-part tests were used for the analyses. The best fitting model based on AIC, included group and sex. Also, based on Q-Q plots and AIC, ZIG regression was a better fit for this model than was ZILN regression. From this ZIG model, the two-part test showed a significant effect of group, but the mean-based tests showed only marginal effects. The post-tests for the two-part tests show that the effect of group is in the probability of exiting the lane where PD subjects had a greater probability of exiting the lane than did control subjects. The sex effect found in this model, on the other hand, was significant when using two-part tests and when using mean-based tests. This effect found that males had higher LDSS scores than did females. The post-tests for the two-part tests show that while males and females exited the lane at similar rates, when males exited the lane they did so to a greater extent than did females. Later in Chapter 5 smaller lane widths than the true lane width were used to develop LDSS related values requiring a safety window. Q-Q plots showed that these LDSS scores calculated within a safety window rather than by the true lane also more clearly fit ZIG than ZILN models. More subjects exited these artificial lanes than exited the true lane leading to larger nonzero sample sizes. This made it easier to find group difference. Under the ZIG model, DM and RM_{MAR} based tests found significant group differences at all artificial lane widths with PD subjects having higher LDSS values than control subjects. RM_{SS} was more impacted by lane width than the other metric, but found significant at some lane widths.

Two-part models (ZIG and ZILN) have limitations when the probabilities of nonzero outcome in one group are either particularly high or particularly low. Specifically, in some instances such a two-part model cannot be fit due to the distributions of the outcome. For example, when comparing two groups if the outcome for one group includes some zero values but the other group has contains only nonzero outcomes the binomial part of the model will not be able to be fit as the design matrix for that part of the model will not be full rank. For example, in our lane narrowing LDSS scores, the smallest lane width that we defined was 2.2 m wide. This was because that was the smallest lane width at which at least one subject from each group left the narrower lane. For lane widths smaller than 2.2 m the two-part model was not solvable. As the definition of the lane width was increased, more subjects did not go outside of the lane leading to more zero values. At some point, if a large enough lane were defined all of the subjects in at least one of the groups would have stayed inside of it leading to no nonzero values in that group. Then, neither part of the two-part model could be fit. In our simulations these limitation were also observed. For some settings, not all data sets simulated were solvable in many cases lack of zeros in one of the groups or too few non-zeros in one of the groups was the reason for this.

Two-part models inherit, to some extent, the properties of the individual parts. They are parametric models and are not particularly robust to certain types of outliers and other departures from the assumed distribution. Because of this, future researchers may be interested in using robust methods to estimate the parts of the two-part model. One such robust method that has been proposed is the minimal Hellinger distance method.[28] This method proposes replacing maximum likelihood estimators (MLE) in parametric models with estimators that minimize the Hellinger distance between the model density and a non-parametric estimator of the data density. This type of estimator will be asymptotically equivalent to the MLE that it replaces, and will be more robust to model misspecification.[28, 29]

For our simulations and analyses, we assumed a constant coefficient of variation across groups. This assumption could be relaxed but adds extra complexity to the development of the mean-based tests. Our simulations only included a selection of the possible effects and sample sizes. Particularly, we only used equal sample sizes and all adjusting effects were in the same direction as the group effect. Further exploration of mean-based tests for two-part models could included such settings. Further more complicated extensions of these methods could be into the realm of allow the y_i's to be correlated. Methods have been developed for running two-part models in such cases, but no tests combining the two part of the models have been developed.[13, 10, 30] Mean-based tests for continuous covariates could also be developed. Mean-based metrics for continuous covariates were introduced in Chapter 4, Section 4.7.1. Tests based on these metrics can be produced by methods similar to those used for the other mean-based metrics in this dissertation. However, the interpretations of these metrics are somewhat awkward and may suffer from lack of generalizability. Further methods could be developed to supplement and clarify the meaning and scope of such tests.

In this dissertation we have found that ZIG and ZILN models differ most when CoV is high. At high CoV values misspecification has the greatest effect, but also at these values the two models are more easily distinguishable. Because of this, we recommend first running a ZILN and looking at the Q-Q plots of the conditional residuals for the nonzero outcomes for obvious departures from the log-normal model to determine if a ZIG model is necessary. As found in Chapter 5 when the nonzero subset of the data are truly distributed as gamma but is analyzed as log-normal, the Q-Q plot will show a concave pattern. If such a pattern is clear, ZIG analyses would then be recommended as the final analysis. When such a pattern is not clear, using the ZILN will be the most appropriate. Our simulations and model fitting examples found a few reasons for this approach. First, when the differences between ZIG and ZILN are slight (low CoV), even if the true distribution is ZIG, ZILN models have appropriate Type 1 error rates and not much loss of power relative to ZIG models. Secondly, when the true model is ZILN, even to some extent at low values of CoV, falsely assumed ZIG models have elevated Type 1 error rates and should be avoided. Finally, when the two models differ greatly and the true model is ZIG, ZILN analyses will have very low power, but the patterns in the Q-Q plots and the comparative AIC values should be able to identify ZIG as the true model in these instances.

Following the model choice, the decision whether to use 2 d.f. tests or mean-based tests should depend on the hypothesis of interest. Two-part tests can be used when a researcher is interested in the differences in either occurrence or extent given occurrence. Mean-based tests can be used when a researcher is interested in an overall directional difference between two groups. We would recommend against RM_{SS}, since it is greatly is influenced by the correlation between the adjusting covariates and the group variable. In situations similar to the simulations presented in this dissertation and when the coefficient of variation is high, we would recommend RM or RM_{MAR} over DM as they have higher power.

APPENDIX A
DELTA METHOD EXAMPLE

A.1 Zero-Inflated Gamma Difference of Means and Ratio of Means Test Derivations

In this appendix we will outline the delta method results for the unadjusted DM and RM tests. Section A.1.1 outlines the asymptotic distribution of the parameters β_0, β_1,τ_0, and τ_1 when a ZIG distribution is assumed. Section A.1.2 and Section A.1.3 derive the asymptotic distributions of $DM(\beta, \tau)$ and $\log(RM(\beta, \tau))$ respectively. The score and information functions derived in Section A.1.1 were also used in the body of the dissertation for the estimation of β_0, β_1,τ_0, and τ_1 and for the development of the tests seen in Chapter 2.

A.1.1 ZIG Likelihood Set-up

ZIG regression models semicontinuous data in two parts. The first part models the probability of a non-zero outcome using logistic regression and the second part models the value of the outcome given the outcome is greater than zero using gamma regression. Let y_i be the value of the semicontinuous outcome where $i = 1, 2, \ldots$ n, X_i be the values of the covariates for subject i, and $p_i = p(X_i, \beta)$ be the probability that y_i is greater than 0 given X_i. Define $\mathbf{I}(y_i > 0)$ as the indicator function where $\mathbf{I}(y_i > 0) = 1$ if $y_i > 0$ and $\mathbf{I}(y_i > 0) = 0$ otherwise. For ZIG regression, logistic regression is used to model the probability of a non-zero outcome ($p(X_i, \beta)$ where $\text{logit}(p(X_i, \beta)) = X_i\beta$) and gamma regression with a log link is used to model the value of the non-zero outcomes such that

$$f(Y_i|Y_i > 0; X_i, \tau, \nu^{-1}) = \frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu_i}\right)^{\nu} Y_i^{\nu-1} exp\left(-\frac{\nu Y_i}{\mu_i}\right) \tag{A.1}$$

where μ_i modeled as $\log(\mu_i) = X_i\tau$.[17] For this appendix we are working with a two group comparison where x_i is a group indicator. Under this scenario, the log-likelihood for ZIG regression is

$$\begin{aligned} l(\beta,\tau) = & \sum_{i=1}^{n} -\log(1+e^{\beta_0+\beta_1 x_i}) + \mathbf{I}(y_i > 0)\,(\beta_0+\beta_1 x_i) \\ & + \sum_{i=1}^{n} \mathbf{I}(y_i > 0)(\nu-1)\log(y_i) - \mathbf{I}(y_i > 0)log(\Gamma(\nu)) + \mathbf{I}(y_i > 0)\nu\log(\nu) \\ & - \nu \sum_{i=1}^{n} \mathbf{I}(y_i > 0)\left(\tau_0 + \tau_1 x_i + \frac{y_i}{e^{\tau_0+\tau_1}}\right). \end{aligned} \tag{A.2}$$

Starting with the $l(\beta,\tau)$ defined in Equation A.2, the score vector is defined by taking the derivative of $l(\beta,\tau)$ in terms of the β and τ parameters. The score vector is defined in Equations A.3 - A.7. The score function, $U(\beta,\tau)$, is

$$U(\beta,\tau)' = \begin{pmatrix} \frac{\delta l(\beta,\tau)}{\delta\beta_0} & \frac{\delta l(\beta,\tau)}{\delta\beta_1} & \frac{\delta l(\beta,\tau)}{\delta\tau_0} & \frac{\delta l(\beta,\tau)}{\delta\tau_1} \end{pmatrix} \tag{A.3}$$

with the elements of $U(\beta,\tau)$ defined as

$$\frac{\delta l(\beta,\tau)}{\delta\beta_0} = \sum_{i=1}^{n} \mathbf{I}(y_i > 0) - \frac{exp^{\beta_0+\beta_1 x_i}}{1+exp^{\beta_0+\beta_1 x_i}}, \tag{A.4}$$

$$\frac{\delta l(\beta,\tau)}{\delta\beta_1} = \sum_{i=1}^{n} \left(\mathbf{I}(y_i > 0) - \frac{exp^{\beta_0+\beta_1 x_i}}{1+exp^{\beta_0+\beta_1 x_i}}\right) x_i, \tag{A.5}$$

$$\frac{\delta l(\beta,\tau)}{\delta\tau_0} = \nu \sum_{i=1}^{n} \mathbf{I}(y_i > 0)\left(\frac{y_i}{e^{\tau_0+\tau_1 x_i}} - 1\right), \tag{A.6}$$

and

$$\frac{\delta l(\beta,\tau)}{\delta\tau_1} = \nu \sum_{i=1}^{n} \mathbf{I}(y_i > 0)x_i\left(\frac{y_i}{e^{\tau_0+\tau_1 x_i}} - 1\right). \tag{A.7}$$

These equations above can be condensed into matrix notation to equate to the score functions used in Chapter 2.Namely,

$$U(\beta,\tau) = \begin{pmatrix} U_b(\beta) \\ U_c(\tau) \end{pmatrix} = \begin{pmatrix} X'(Y_b - P) \\ X' D_{Y_b}\left(\frac{Y-\mu}{\mu}\right) \end{pmatrix} \tag{A.8}$$

where Y_b is an n $\times$ 1 vector defined such that the element $Y_{bi} = \mathbf{I}(y_i > 0)$, D_{Y_b} is an n $\times$ n diagonal matrix with Y_b on the diagonals and 0 elsewhere, P is an n $\times$ 1 vector containing for each row i, $P(Y_i > 0) = \frac{exp^{\beta_0+\beta_1 x_i}}{1+exp^{\beta_0+\beta_1 x_i}}$, and $\frac{Y-\mu}{\mu}$ is an n $\times$ 1 vector with elements $\frac{y_i - e^{\tau_0+\tau_1 x_i}}{e^{\tau_0+\tau_1 x_i}}$.

To find the information matrix $I(\beta_0, \beta_1, \tau_0, \tau_1)$ we will find the second derivative of $l(\beta,\tau)$. The information matrix is defined as follows:

$$I(\beta,\tau) = -\begin{pmatrix} \frac{\delta^2 l(\beta,\tau)}{\delta\beta_0\delta\beta_0} & \frac{\delta^2 l(\beta,\tau)}{\delta\beta_0\delta\beta_1} & \frac{\delta^2 l(\beta,\tau)}{\delta\beta_0\delta\tau_0} & \frac{\delta^2 l(\beta,\tau)}{\delta\beta_0\delta\tau_1} \\ \frac{\delta^2 l(\beta,\tau)}{\delta\beta_1\delta\beta_0} & \frac{\delta^2 l(\beta,\tau)}{\delta\beta_1\delta\beta_1} & \frac{\delta^2 l(\beta,\tau)}{\delta\beta_1\delta\tau_0} & \frac{\delta^2 l(\beta,\tau)}{\delta\beta_1\delta\tau_1} \\ \frac{\delta^2 l(\beta,\tau)}{\delta\tau_0\delta\beta_0} & \frac{\delta^2 l(\beta,\tau)}{\delta\tau_0\delta\beta_1} & \frac{\delta^2 l(\beta,\tau)}{\delta\tau_0\delta\tau_0} & \frac{\delta^2 l(\beta,\tau)}{\delta\tau_0\delta\tau_1} \\ \frac{\delta^2 l(\beta,\tau)}{\delta\tau_1\delta\beta_0} & \frac{\delta^2 l(\beta,\tau)}{\delta\tau_1\delta\beta_1} & \frac{\delta^2 l(\beta,\tau)}{\delta\tau_1\delta\tau_0} & \frac{\delta^2 l(\beta,\tau)}{\delta\tau_1\delta\tau_1} \end{pmatrix}. \tag{A.9}$$

These second derivatives can be found by taking derivatives of the first derivatives found for the score function in Equations A.4 to A.7. Equations A.10 - A.19 give the second derivatives found in lower triangle including as below the diagonal of the information matrix outlined in Equation A.9. Due to the symmetry of $I(\beta,\tau)$ one only needs the lower triangle of $I(\beta,\tau)$, since $\frac{\delta^2 l(\beta,\tau)}{\delta\theta_k\delta\theta_l} = \frac{\delta^2 l(\beta,\tau)}{\delta\theta_l\delta\theta_k}$. Equations A.10 to A.13 show the derivatives of Equation A.4 with respect to β_0, β_1, τ_0, and τ_1 respectively. These derivatives make up the first column of $I(\beta,\tau)$. Note that Equation A.4 does not contain τ_0 or τ_1 and so the derivatives of Equation A.4 with respect to these parameters is zero.

$$\begin{aligned} \frac{\delta^2 l(\beta,\tau)}{\delta\beta_0\delta\beta_0} &= \sum_{i=1}^{n} -\frac{exp^{\beta_0+\beta_1 x_i}}{(1+exp^{\beta_0+\beta_1 x_i})^2} = \sum_{i=1}^{n} -p_i(1-p_i) \\ &= -n_0 p_0(1-p_0) - n_1 p_1(1-p_1) \end{aligned} \tag{A.10}$$

$$\frac{\delta^2 l(\beta,\tau)}{\delta\beta_0\delta\beta_1} = \sum_{i=1}^{n} -x_i \frac{exp^{\beta_0+\beta_1 x_i}}{(1+exp^{\beta_0+\beta_1 x_i})^2} = \sum_{i=1}^{n} -x_i p_i(1-p_i) = -n_1 p_1(1-p_1) \quad \text{(A.11)}$$

$$\frac{\delta^2 l(\beta,\tau)}{\delta\beta_0\delta\tau_0} = 0 \quad \text{(A.12)}$$

$$\frac{\delta^2 l(\beta,\tau)}{\delta\beta_0\delta\tau_1} = 0 \quad \text{(A.13)}$$

Equations A.14 to A.16 show the derivatives of Equation A.5 with respect to β_1, τ_0, and τ_1 respectively. These derivatives fill in the the second column of $I(\beta,\tau)$ at and below the diagonal. Note that Equation A.5 does not contain τ_0 or τ_1 and so the derivatives of Equation A.5 with respect to these parameters is zero.

$$\frac{\delta^2 l(\beta,\tau)}{\delta\beta_1\delta\beta_1} = \sum_{i=1}^{n} -x_{1i}^2 \frac{exp^{\beta_0+\beta_1 x_i}}{(1+exp^{\beta_0+\beta_1 x_i})^2} = \sum_{i=1}^{n} -x_{1i}^2 p_i(1-p_i) = -n_1 p_1(1-p_1) \quad \text{(A.14)}$$

$$\frac{\delta^2 l(\beta,\tau)}{\delta\beta_1\delta\tau_0} = 0 \quad \text{(A.15)}$$

$$\frac{\delta^2 l(\beta,\tau)}{\delta\beta_1\delta\tau_1} = 0 \quad \text{(A.16)}$$

Equations A.17 and A.18 show the derivatives of Equation A.6 with respect to τ_0, and τ_1 respectively. These derivatives fill in the the third column of $I(\beta,\tau)$ at and below the diagonal.

$$\frac{\delta^2 l(\beta,\tau)}{\delta\tau_0\delta\tau_0} = -\nu \sum_{i=1}^{n} \mathbf{I}(y_i > 0)\left(\frac{y_i}{e^{\tau_0+\tau_1 x_i}}\right) \quad \text{(A.17)}$$

$$\frac{\delta^2 l(\beta,\tau)}{\delta\tau_0\delta\tau_1} = -\nu\sum_{i=1}^{n}\mathbf{I}(y_i>0)\left(\frac{x_i y_i}{e^{\tau_0+\tau_1 x_i}}\right) \tag{A.18}$$

And finally, Equation A.19 shows the derivative of Equation A.6 with respect to τ_1. This is the second derivative of $l(\beta,\tau)$ with respect to τ_1 and fills in the fourth diagonal position of $I(\beta,\tau)$.

$$\frac{\delta^2 l(\beta,\tau)}{\delta\tau_1\delta\tau_1} = -\nu\sum_{i=1}^{n}\mathbf{I}(y_i>0)\left(\frac{x_i^2 y_i}{e^{\tau_0+\tau_1 x_i}}\right) \tag{A.19}$$

Plugging these second derivatives into $I(\beta,\tau)$ from Equation A.9, the Fisher information is

$$I(\beta,\tau) = \begin{pmatrix} I_b(\beta) & 0 \\ 0 & I_c(\tau) \end{pmatrix}. \tag{A.20}$$

where

$$I_b(\beta) = \begin{pmatrix} n_0p_0(1-p_0)+n_1p_1(1-p_1) & n_1p_1(1-p_1) \\ n_1p_1(1-p_1) & n_1p_1(1-p_1) \end{pmatrix}. \tag{A.21}$$

and

$$I_c(\tau) = \begin{pmatrix} \nu\sum_{i=1}^{n}\mathbf{I}(y_i>0)\left(\frac{y_i}{e^{\tau_0+\tau_1 x_i}}\right) & \nu\sum_{i=1}^{n}\mathbf{I}(y_i>0)\left(\frac{x_i y_i}{e^{\tau_0+\tau_1 x_i}}\right) \\ \nu\sum_{i=1}^{n}\mathbf{I}(y_i>0)\left(\frac{x_i y_i}{e^{\tau_0+\tau_1 x_i}}\right) & \nu\sum_{i=1}^{n}\mathbf{I}(y_i>0)\left(\frac{x_i^2 y_i}{e^{\tau_0+\tau_1 x_i}}\right) \end{pmatrix}. \tag{A.22}$$

Taking the expectation of the Fisher information matrix $I(\beta,\tau)$, the expected Fisher information matrix, $\mathcal{I}(\beta,\tau)$, can be found. Since, $I_b(\beta)$ does not contain y_i, the expected Fisher information for the binomial part of the model, $\mathcal{I}_b(\beta)$, is equivalent to the Fisher information.

For $I_c(\tau)$, on the other hand, the expected information does not equal the Fisher information, as y_i is involved in $I_c(\tau)$. A couple of useful notes for the calculation of $\mathcal{I}_c(\tau) = E(I_c(\tau))$; First, $E\left(\frac{\delta l(\theta)}{\delta\theta_k}\right) = 0$ for any θ_k. That is, the expected value of each of the first derivatives calculated in Equations A.4 - A.7 are equal to zero. Secondly,

$E(\mathbf{I}(y_i > 0)) = P(y_i > 0) = p_i$. Equations to find the expected value of the elements of $I_c(\tau)$ follow:

$$\begin{aligned}
E\left(\frac{\delta^2 l(\beta,\tau)}{\delta\tau_0\delta\tau_0}\right) &= E\left(\nu\sum_{i=1}^{n}\mathbf{I}(y_i > 0)\left(\frac{y_i}{e^{\tau_0+\tau_1 x_i}}\right)\right) \\
&= E\left(I\nu\sum_{i=1}^{n}(y_i > 0)\left(\frac{y_i}{e^{\tau_0+\tau_1 x_i}} - 1 + 1\right)\right) \\
&= E\left(\nu\sum_{i=1}^{n}\mathbf{I}(y_i > 0)\left(\frac{y_i}{e^{\tau_0+\tau_1 x_i}} - 1\right)\right) + E\left(\nu\sum_{i=1}^{n}\mathbf{I}(y_i > 0)\right) \\
&= 0 + \nu\sum_{i=1}^{n}E(\mathbf{I}(y_i > 0)) = \nu\sum_{i=1}^{n}p_i = n_0p_0 + n_1p_1
\end{aligned} \tag{A.23}$$

because $\nu\sum_{i=1}^{n}E\left(\mathbf{I}(y_i > 0)\left(\frac{y_i}{e^{\tau_0+\tau_1 x_i}} - 1\right)\right) = 0$ since $E\left(\frac{\delta l(\beta,\tau)}{\delta\tau_0}\right) = 0$ implies that $E\left(\nu\sum_{i=1}^{n}\mathbf{I}(y_i > 0)\left(\frac{y_i}{e^{\tau_0+\tau_1 x_i}} - 1\right)\right) = 0$.

The off-diagonal element of $\mathcal{I}_c(\tau) = E(I_c(\tau))$ is

$$\begin{aligned}
E\left(\frac{\delta^2 l(\beta,\tau)}{\delta\tau_1\delta\tau_0}\right) &= E\left(\nu\sum_{i=1}^{n}\mathbf{I}(y_i > 0)x_i\left(\frac{y_i}{e^{\tau_0+\tau_1 x_i}}\right)\right) \\
&= E\left(\nu\sum_{i=1}^{n}\mathbf{I}(y_i > 0)x_i\left(\frac{y_i}{e^{\tau_0+\tau_1 x_i}} - 1 + 1\right)\right) \\
&= E\left(\nu\sum_{i=1}^{n}\mathbf{I}(y_i > 0)x_i\left(\frac{y_i}{e^{\tau_0+\tau_1 x_i}} - 1\right)\right) + E\left(\nu\sum_{i=1}^{n}\mathbf{I}(y_i > 0)x_i\right) \\
&= 0 + \nu\sum_{i=1}^{n}x_iE\left(\mathbf{I}(y_i > 0)\right) = \nu\sum_{i=1}^{n}x_ip_i = \nu n_1p_1
\end{aligned} \tag{A.24}$$

Finally, the bottom right element of $\mathcal{I}_c(\tau)$ is

$$\begin{aligned}
E\left(\frac{\delta^2 l(\beta,\tau)}{\delta\tau_1\delta\tau_1}\right) &= E\left(\nu\sum_{i=1}^{n}\mathbf{I}(y_i>0)x_i^2\left(\frac{y_i}{e^{\tau_0+\tau_1 x_i}}\right)\right)\\
&= E\left(\nu\sum_{i=1}^{n}\mathbf{I}(y_i>0)x_i^2\left(\frac{y_i}{e^{\tau_0+\tau_1 x_i}}-1+1\right)\right)\\
&= E\left(\nu\sum_{i=1}^{n}\mathbf{I}(y_i>0)x_i^2\left(\frac{y_i}{e^{\tau_0+\tau_1 x_i}}-1\right)\right)+E\left(\nu\sum_{i=1}^{n}\mathbf{I}(y_i>0)x_i^2\right)\\
&= \nu\sum_{i=1}^{n}x_i^2 E\left(\mathbf{I}(y_i>0)\left(\frac{y_i}{e^{\tau_0+\tau_1 x_i}}-1\right)\right)+E\left(\nu\sum_{i=1}^{n}\mathbf{I}(y_i>0)x_i^2\right)\\
&= \nu\sum_{i=1}^{n}x_i^2\times 0+\nu\sum_{i=1}^{n}x_i^2 E\left(\mathbf{I}(y_i>0)\right)\\
&= 0+\nu\sum_{i=1}^{n}x_i^2 E\left(\mathbf{I}(y_i>0)\right)=\nu\sum_{i=1}^{n}x_i^2 p_i=\nu n_1 p_1
\end{aligned} \tag{A.25}$$

Putting these together, the expected information $\mathcal{I}_c(\tau)$ is

$$\mathcal{I}_c(\tau)=\begin{pmatrix}\nu\sum_{i=1}^{n}p_i & \nu\sum_{i=1}^{n}x_i p_i\\ \nu\sum_{i=1}^{n}x_i p_i & \nu\sum_{i=1}^{n}x_i^2 p_i\end{pmatrix}=\begin{pmatrix}n_0p_0+n_1p_1 & n_1p_1\\ n_1p_1 & n_1p_1\end{pmatrix}. \tag{A.26}$$

In matrix form the expected information, $\mathcal{I}(\beta,\tau)$ is

$$\mathcal{I}(\beta,\tau)=\begin{pmatrix}X'D_{p(1-p)}X & 0\\ 0 & \nu X'D_pX\end{pmatrix} \tag{A.27}$$

where D_p is an n $\times$ n diagonal matrix with $p_i(1-p_i)$ on the diagonals and 0 elsewhere and D_p is also a diagonal n $\times$ n matrix but with p_i on the diagonals.

Using these results, the parameters from a ZIG model have the following joint asymptotic distributional properties:

$$\sqrt{n}\left(\begin{pmatrix}\hat\beta_0\\ \hat\beta_1\\ \hat\tau_0\\ \hat\tau_1\end{pmatrix}-\begin{pmatrix}\beta_0\\ \beta_1\\ \tau_0\\ \tau_1\end{pmatrix}\right)\xrightarrow{d}\mathrm{N}(0,\Sigma) \tag{A.28}$$

where $\Sigma = \mathrm{n}\mathcal{I}(\beta,\tau)^{-1}$ with n being the overall sample size. Σ is:

$$\Sigma = \begin{pmatrix} \frac{n}{n_0p_0(1-p_0)} & -\frac{n}{n_0p_0(1-p_0)} & 0 & 0 \\ -\frac{n}{n_0p_0(1-p_0)} & \frac{n}{n_0p_0(1-p_0)} + \frac{n}{n_1p_1(1-p_1)} & 0 & 0 \\ 0 & 0 & \nu^{-1}\frac{n}{n_0p_0} & -\nu^{-1}\frac{n}{n_0p_0} \\ 0 & 0 & -\nu^{-1}\frac{n}{n_0p_0} & \nu^{-1}\left(\frac{n}{n_0p_0} + \frac{n}{n_1p_1}\right) \end{pmatrix} \tag{A.29}$$

This can be written to be free of sample size by writing Σ in terms of weights where $w_0 = \frac{n_0}{n}$ is the proportion of the sample that is in group 0 and $w_1 = \frac{n_1}{n}$ is the proportion of the sample that is in group 1.

$$\Sigma = \begin{pmatrix} \frac{1}{w_0p_0(1-p_0)} & -\frac{1}{w_0p_0(1-p_0)} & 0 & 0 \\ -\frac{1}{w_0p_0(1-p_0)} & \frac{1}{w_0p_0(1-p_0)} + \frac{1}{w_1p_1(1-p_1)} & 0 & 0 \\ 0 & 0 & \nu^{-1}\frac{1}{w_0p_0} & -\nu^{-1}\frac{1}{w_0p_0} \\ 0 & 0 & -\nu^{-1}\frac{1}{w_0p_0} & \nu^{-1}\left(\frac{1}{w_0p_0} + \frac{1}{w_1p_1}\right) \end{pmatrix} \tag{A.30}$$

A.1.2 Approximate Distribution for Difference of Means

The delta method states that given $\sqrt{n}\left(\hat{\theta}-\theta\right) \xrightarrow{d} N\left(0,\Sigma\right)$ and a function $\mathrm{g}(\theta)$ with first derivatives $\Delta = \left(\frac{\delta \mathrm{g}(\theta)}{\delta\theta_i}\right)$, if $\frac{\delta \mathrm{g}(\theta)}{\delta\theta_i}$ is continuous for all i, and $\frac{\delta \mathrm{g}(\theta)}{\delta\theta_i} \neq 0$ for some i then[31, 32],

$$\sqrt{n}\left(\mathrm{g}(\hat{\theta}) - \mathrm{g}(\theta)\right) \xrightarrow{d} \mathrm{N}(0, \Delta\Sigma\Delta^T). \tag{A.31}$$

For the ZIG model with no adjusting covariates $\theta = (\beta_0, \beta_1, \tau_0, \tau_1)$. Then delta method can be use for any $\mathrm{g}(\theta) = \mathrm{g}(\beta_0, \beta_1, \tau_0, \tau_1)$ that fits the conditions outlined above. In this section, we use the delta method to find the asymptotic distribution for $DM(\beta,\tau)$ and in the next section we use it to find the asymptotic distribution for $RM(\beta,\tau)$. The difference of means between two groups or the ratio of means of two groups can be written as function of β_0, β_1, τ_0, and τ_1. Then using Equations A.27 and A.28 then the delta method can be applied to find the asymptotic distribution of $DM(\hat{\beta}, \hat{\tau})$.

The difference of means between two groups without any covariate adjusting under the ZIG model expressed in terms of β_0, β_1, τ_0, and τ_1 is as follows:

$$\begin{aligned} \mathrm{g}(\beta,\tau) &= DM = E(Y_i|X=1) - E(Y_i|X=0) \\ &= E(Y_i|Y_i>0, X=1)P(Y_i>0|X=1) \\ &- E(Y_i|Y_i>0, X=0)P(Y_i>0|X=0) \\ &= e^{\tau_0+\tau_1}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} - e^{\tau_0}\frac{e^{\beta_0}}{1+e^{\beta_0}}. \end{aligned} \tag{A.32}$$

To find the asymptotic distribution of $DM(\hat{\beta}, \hat{\tau})$ and to check whether use of the delta method is appropriate, we first find the vector of derivatives of $DM(\beta, \tau)$. This is

$$\Delta_{DM}^T = \begin{pmatrix} \frac{\delta DM}{\delta \beta_0} \\ \frac{\delta DM}{\delta \beta_1} \\ \frac{\delta DM}{\delta \tau_0} \\ \frac{\delta DM}{\delta \tau_1} \end{pmatrix} = \begin{pmatrix} e^{\tau_0+\tau_1}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} - e^{\tau_0}\frac{e^{\beta_0}}{(1+e^{\beta_0})^2} \\ e^{\tau_0+\tau_1}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} \\ e^{\tau_0+\tau_1}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} - e^{\tau_0}\frac{e^{\beta_0}}{1+e^{\beta_0}} \\ e^{\tau_0+\tau_1}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} \end{pmatrix} \tag{A.33}$$

which can be simplified as follows where p_j is the probability of a non-zero outcome for group j and μ_j is the mean of the non-zero values for $j = 0, 1$:

$$\Delta_{DM}^T = \begin{pmatrix} p_1(1-p_1)\mu_1 - p_0(1-p_0)\mu_0 \\ p_1(1-p_1)\mu_1 \\ p_1\mu_1 - p_0\mu_0 \\ p_1\mu_1 \end{pmatrix}. \tag{A.34}$$

Then before proceeding with the delta method, we use this result to check the conditions under which the delta method can be used. To satisfy one condition, it can be shown that the derivatives seen in Equation A.33 are all continuous for all values of β_0, β_1, τ_0, and τ_1. Specifically, since the exponential function is a continuous function, a continuous function divided by a continuous function is continuous except where the denominator equal zero, since $(1 + e^{\theta})$ is always greater than zero then $\frac{e^{\theta}}{(1+e^{\theta})^2}$ is continuous, and finally since a continuous function times a continuous function, and a continuous function plus a continuous function both yield continuous functions, all of the derivatives of $DM(\beta, \tau)$ as seen in Equation A.33 are continuous.

A second condition states that $\frac{\delta \mathrm{g}(\theta)}{\delta \theta_i} \neq 0$ for some θ_i. This condition holds since $\frac{\delta DM}{\delta \tau_1} = e^{\tau_0+\tau_1} \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$ is greater than zero for all values of β_0, β_1, τ_0, and τ_1. This can be easily seen as $e^{\tau_0+\tau_1}$ is greater than zero for all τ_0 and τ_1 values and $\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$ is greater than zero for all values of β_0 and β_1.

Since the conditions necessary for use of the delta method are met, we can utilize the delta method to find an asymptotic variance for $DM(\beta, \tau)$. Using the delta method as explained in Equation A.79, Σ the asymptotic variance of β_0, β_1, τ_0, and τ_1 from Equation A.30, and the vector of derivatives Δ_{DM} from Equation A.34, the asymptotic distribution of $DM(\beta, \tau)$ can be established as

$$\sqrt{n}\left(DM(\hat{\beta}, \hat{\tau}) - DM(\beta, \tau)\right) \xrightarrow{d} \mathrm{N}(0, \Delta_{DM}\Sigma\Delta_{DM}^T) \tag{A.35}$$

where $\Delta_{DM}\Sigma\Delta_{DM}^T$ is found below in Equation A.37.

For the sake of space we begin by defining $\Delta_{DM}\Sigma$ in Equation A.36 as

$$\begin{aligned}
&\Delta_{DM}\Sigma = \\
&\begin{pmatrix} p_1(1-p_1)\mu_1 - p_0(1-p_0)\mu_0 & p_1(1-p_1)\mu_1 & p_1\mu_1 - p_0\mu_0 & p_1\mu_1 \end{pmatrix} \\
&\times \begin{pmatrix} \frac{1}{w_0p_0(1-p_0)} & -\frac{1}{w_0p_0(1-p_0)} & 0 & 0 \\ -\frac{1}{w_0p_0(1-p_0)} & \frac{1}{w_0p_0(1-p_0)} + \frac{1}{w_1p_1(1-p_1)} & 0 & 0 \\ 0 & 0 & \nu^{-1}\frac{1}{w_0p_0} & -\nu^{-1}\frac{1}{w_0p_0} \\ 0 & 0 & -\nu^{-1}\frac{1}{w_0p_0} & \nu^{-1}\left(\frac{1}{w_0p_0} + \frac{1}{w_1p_1}\right) \end{pmatrix} \\
&= \begin{pmatrix} -\frac{\mu_0}{w_0} & \frac{\mu_0}{w_0} + \frac{\mu_1}{w_1} & -\nu^{-1}\frac{\mu_0}{w_0} & \nu^{-1}\left(\frac{\mu_0}{w_0} + \frac{\mu_1}{w_1}\right) \end{pmatrix}.
\end{aligned} \tag{A.36}$$

Then in Equation A.37 we define the final $\Delta_{DM}\Sigma\Delta_{DM}^T$ as

$$\begin{aligned}
&\Delta_{DM}\Sigma\Delta_{DM}^T = \\
&\begin{pmatrix} -\frac{\mu_0}{w_0} & \frac{\mu_0}{w_0} + \frac{\mu_1}{w_1} & -\nu^{-1}\frac{\mu_0}{w_0} & \nu^{-1}\left(\frac{\mu_0}{w_0} + \frac{\mu_1}{w_1}\right) \end{pmatrix} \\
&\times \begin{pmatrix} p_1(1-p_1)\mu_1 - p_0(1-p_0)\mu_0 \\ p_1(1-p_1)\mu_1 \\ p_1\mu_1 - p_0\mu_0 \\ p_1\mu_1 \end{pmatrix} \\
&= \frac{\mu_0^2p_0(1-p_0)}{w_0} + \frac{\mu_1^2p_1(1-p_1)}{w_1} + \nu^{-1}\frac{\mu_0^2p_0}{w_0} + \nu^{-1}\frac{\mu_1^2p_1}{w_1}.
\end{aligned} \tag{A.37}$$

From this result, a variance of $DM(\hat{\beta}, \hat{\tau})$ can be defined for finite sample sizes. Specifically, since $w_0 = \frac{n_0}{n}$ and $w_1 = \frac{n_1}{n}$,

$$\begin{aligned}
&n\Delta_{DM}\Sigma\Delta_{DM}^T = \\
&Var(DM(\hat{\beta}, \hat{\tau})) = \frac{\mu_0^2p_0(1-p_0)}{n_0} + \frac{\mu_1^2p_1(1-p_1)}{n_1} + \nu^{-1}\frac{\mu_0^2p_0}{n_0} + \nu^{-1}\frac{\mu_1^2p_1}{n_1}.
\end{aligned} \tag{A.38}$$

This variance can then be used to create Wald confidence intervals or Wald-type hypothesis tests as was done in Chapter 3.

A.1.3 Approximate Distribution for Ratio of Means

The Delta Method approach can also be used to find asymptotic variance of the log of the ratio of the means $\log(RM(\beta,\tau))$, where

$$\begin{aligned} \mathrm{g}(\beta,\tau) = \log(RM(\beta,\tau)) &= \beta_1 + \tau_1 + \log(1+e^{\beta_0}) - \log(1+e^{\beta_0+\beta_1}) \\ &= \beta_1 + \tau_1 - \log(1-p_0) + \log(1-p_1). \end{aligned} \tag{A.39}$$

First, a vector of derivatives, $\Delta_{\log(RM(\beta,\tau))}$, could be constructed such that

$$\Delta^T_{\log(RM(\beta,\tau))} = \begin{pmatrix} \frac{\delta \log(RM(\beta,\tau))}{\delta\beta_0} \\ \frac{\delta \log(RM(\beta,\tau))}{\delta\beta_1} \\ \frac{\delta \log(RM(\beta,\tau))}{\delta\tau_0} \\ \frac{\delta \log(RM(\beta,\tau))}{\delta\tau_1} \end{pmatrix} = \begin{pmatrix} \frac{e^{\beta_0}}{1+e^{\beta_0}} - \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} \\ 1 - \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} \\ 0 \\ 1 \end{pmatrix} \tag{A.40}$$

which can be simplified as

$$\Delta^T_{\log(RM(\beta,\tau))} = \begin{pmatrix} p_0(1-p_0) - p_1(1-p_1) \\ -p_1(1-p_1) \\ 0 \\ 1 \end{pmatrix}. \tag{A.41}$$

These derivatives are all continuous for all values of β_0, β_1, τ_0, and τ_1 and at least one of the derivatives (specifically, $\frac{\delta \log(RM(\beta,\tau))}{\delta\tau_1}$) does not equal zero for all values of β_0, β_1, τ_0, and τ_1. Therefore, the delta method may be used. Finally, utilizing the delta method explained in Equation A.79, Σ the asymptotic variance of β_0, β_1, τ_0, and τ_1 from Equation A.30, and the vector of derivatives $\Delta_{\log(RM(\beta,\tau))}$ from Equation A.41, the asymptotic distribution of $\log(RM(\hat{\beta},\hat{\tau}))$ can be established as

$$\sqrt{n}\left(\log(RM(\hat{\beta},\hat{\tau})) - \log(RM(\beta,\tau))\right) \xrightarrow{d} \mathrm{N}(0, \Delta_{\log(RM)}\Sigma\Delta^T_{\log(RM)}) \tag{A.42}$$

where $\Delta_{\log(RM)}\Sigma\Delta^T_{\log(RM)}$ is found in Equation A.45. For the sake of space we begin by defining $\Delta_{\log(RM)}\Sigma$ in Equation A.43 as

$$\Delta_{\log(RM(\beta,\tau))}\Sigma =$$
$$\begin{pmatrix} p_0(1-p_0)-p_1(1-p_1) & -p_1(1-p_1) & 0 & 1 \end{pmatrix}$$
$$\begin{pmatrix} \frac{1}{w_0p_0(1-p_0)} & -\frac{1}{w_0p_0(1-p_0)} & 0 & 0 \\ -\frac{1}{w_0p_0(1-p_0)} & \frac{1}{w_0p_0(1-p_0)}+\frac{1}{w_1p_1(1-p_1)} & 0 & 0 \\ 0 & 0 & \nu^{-1}\frac{1}{w_0p_0} & -\nu^{-1}\frac{1}{w_0p_0} \\ 0 & 0 & -\nu^{-1}\frac{1}{w_0p_0} & \nu^{-1}\left(\frac{1}{w_0p_0}+\frac{1}{w_1p_1}\right) \end{pmatrix} \tag{A.43}$$
$$= \begin{pmatrix} \frac{1}{w_0} & -\frac{1}{w_0}-\frac{1}{w_1} & -\nu^{-1}\frac{1}{w_0p_0} & \nu^{-1}\left(\frac{1}{w_0p0}+\frac{1}{w_1p_1}\right) \end{pmatrix}.$$

Then in Equation A.45 we define the final $\Delta_{\log(RM)}\Sigma\Delta^T_{\log(RM)}$ as

$$\Delta_{\log(RM)}\Sigma\Delta^T_{\log(RM)}$$
$$= \begin{pmatrix} \frac{1}{w_0} & -\frac{1}{w_0}-\frac{1}{w_1} & -\nu^{-1}\frac{1}{w_0p_0} & \nu^{-1}\left(\frac{1}{w_0p0}+\frac{1}{w_1p_1}\right) \end{pmatrix} \begin{pmatrix} p_0(1-p_0)-p_1(1-p_1) \\ -p_1(1-p_1) \\ 0 \\ 1 \end{pmatrix}$$
$$= \frac{p_0(1-p_0)}{w_0} + \frac{p_1(1-p_1)}{w_1} + \nu^{-1}\left(\frac{1}{w_0p_0}+\frac{1}{w_1p_1}\right) \tag{A.44}$$

From this result, a variance of $\log(RM)$ can be defined for finite sample sizes. Specifically, since $w_0 = \frac{n_0}{n}$ and $w_1 = \frac{n_1}{n}$,

$$n\Delta_{\log(RM)}\Sigma\Delta^T_{\log(RM)} =$$
$$Var(\log(RM(\hat{\beta},\hat{\tau})) = \frac{p_0(1-p_0)}{n_0} + \frac{p_1(1-p_1)}{n_1} + \nu^{-1}\left(\frac{1}{n_0p_0}+\frac{1}{n_1p_1}\right). \tag{A.45}$$

This result was then used in Chapter 3 to create a Wald type hypothesis test based on $\log(RM)$ and was also used in Section 5.2.1 to create a 95% confidence interval of RM by first finding a 95% confidence interval of $\log(RM(\beta,\tau))$ using this asymptotic variance and then exponentiating the confidence limits.

Similar methods (not shown) were used to find asymptotic distributions for

DM, $\log(RM_{SS})$, and $\log(RM_{MAR})$ when adjusting for covariates. These tests are used in Chapter 4.

A.2 Zero-Inflated Log-Normal Difference of Means and Ratio of Means Test Derivations

In this appendix we follow the same pattern as in Appendix Section A.2. First we will outline the ZILN model likelihood and the asymptotic distribution of the related parameters β_0, β_1, τ_0, τ_1, and σ^2. Then we will use the delta method to derive the asymptotic distribution of $DM(\beta, \tau, \sigma^2)$).

A.2.1 Zero-Inflated Log-Normal Likelihood Set-up

As was the case for ZIG regression, ZILN regression consists of a two-part model with the first part using logistic regression to model the probability of a non-zero outcome ($p(X_i, \beta)$ where $\text{logit}(p(X_i, \beta)) = X_i\beta$). For ZILN, in the second part of the model the non-zero outcomes are modeled such that $\log(y_i) \sim \text{N}(X_i\tau, \sigma^2)$, where X_i is the vector of covariates for subject i and y_i is the semicontinuous outcome for subject i. In the two group comparison without covariate adjustment scenario being explored here $X_i' = (1, x_i)$ where $x_i = 0, 1$ is a group indicator.

This leads to an overall likelihood of as found in Equation 2.11. Taking the log of Equation 2.11 leads to the log-likelihood which is:

$$\begin{aligned} l(\beta, \tau, \sigma^2) = &\sum_{i=1}^{n} -\log(1 + e^{\beta_0 + \beta_1 x_i}) + \mathbf{I}(y_i > 0)(\beta_0 + \beta_1 x_i) \\ &+ \sum_{i=1}^{n} -\mathbf{I}(y_i > 0)\log(y_i) - \mathbf{I}(y_i > 0)\frac{1}{2}\log(2\pi\sigma^2) \\ &+ \sum_{i=1}^{n} -\mathbf{I}(y_i > 0)\frac{(\log(y_i) - (\tau_0 + \tau_1 x_i))^2}{2\sigma^2} \end{aligned} \tag{A.46}$$

This log-likelihood can then be used to create a score vector $U(\beta, \tau, \sigma^2)$ which is defined as a vector of the derivatives of $l(\beta, \tau, \sigma^2)$ in terms of β_0, β_1, τ_0, τ_1, and σ^2.

$U(\beta, \tau, \sigma^2)$ is defined in Equations A.47 through A.52.

$$U(\beta, \tau, \sigma^2)' = \begin{pmatrix} \frac{\delta l(\beta,\tau,\sigma^2)}{\delta \beta_0} & \frac{\delta l(\beta,\tau,\sigma^2)}{\delta \beta_1} & \frac{\delta l(\beta,\tau,\sigma^2)}{\delta \tau_0} & \frac{\delta l(\beta,\tau,\sigma^2)}{\delta \tau_1} & \frac{\delta l(\beta,\tau,\sigma^2)}{\delta \sigma^2} \end{pmatrix} \tag{A.47}$$

with the elements of $U(\beta, \tau, \sigma^2)$ defined as

$$\frac{\delta l(\beta, \tau, \sigma^2)}{\delta \beta_0} = \sum_{i=1}^{n} \mathbf{I}(y_i > 0) - \frac{exp^{\beta_0 + \beta_1 x_i}}{1 + exp^{\beta_0 + \beta_1 x_i}}, \tag{A.48}$$

$$\frac{\delta l(\beta, \tau, \sigma^2)}{\delta \beta_1} = \sum_{i=1}^{n} \left(\mathbf{I}(y_i > 0) - \frac{exp^{\beta_0 + \beta_1 x_i}}{1 + exp^{\beta_0 + \beta_1 x_i}} \right) x_i, \tag{A.49}$$

$$\frac{\delta l(\beta, \tau, \sigma^2)}{\delta \tau_0} = \sum_{i=1}^{n} -\mathbf{I}(y_i > 0) \left(\frac{\ln(y_i) - (\tau_0 + \tau_1 x_i)}{\sigma^2} \right), \tag{A.50}$$

$$\frac{\delta l(\beta, \tau, \sigma^2)}{\delta \tau_1} = \sum_{i=1}^{n} -\mathbf{I}(y_i > 0) \left(\frac{\ln(y_i) - (\tau_0 + \tau_1 x_i)}{\sigma^2} \right) x_{1i}, \tag{A.51}$$

and

$$\frac{\delta l(\beta, \tau, \sigma^2)}{\delta \sigma^2} = \sum_{i=1}^{n} -\mathbf{I}(y_i > 0) \frac{1}{2\sigma^2} + \mathbf{I}(y_i > 0) \left(\frac{(\ln(y_i) - (\tau_0 + \tau_1 x_i))^2}{\sigma^4} \right). \tag{A.52}$$

To find the information matrix $\mathrm{I}(\beta, \tau, \sigma^2)$ we will find the second derivative of $l(\beta, \tau, \sigma^2)$. The information matrix is defined as follows:

$$\mathrm{I}(\beta, \tau, \sigma^2) = - \begin{pmatrix} \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_0\delta\beta_0} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_0\delta\beta_1} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_0\delta\tau_0} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_0\delta\tau_1} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_0\delta\sigma^2} \\ \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_1\delta\beta_0} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_1\delta\beta_1} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_1\delta\tau_0} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_1\delta\tau_1} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_1\delta\sigma^2} \\ \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_0\delta\beta_0} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_0\delta\beta_1} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_0\delta\tau_0} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_0\delta\tau_1} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_0\delta\sigma^2} \\ \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_1\delta\beta_0} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_1\delta\beta_1} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_1\delta\tau_0} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_1\delta\tau_1} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_1\delta\sigma^2} \\ \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\sigma^2\delta\beta_0} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\sigma^2\delta\beta_1} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\sigma^2\delta\tau_0} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\sigma^2\delta\tau_1} & \frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\sigma^2\delta\sigma^2} \end{pmatrix}. \tag{A.53}$$

Using the first derivative found in the score vector, we find the second derivatives to fill in the matrix defined in Equation A.53. Starting with the first derivatives in

Equations A.48 to A.52, the second derivatives are found below. Again we will show only the bottom triangle of the matrix as the entire matrix can be defined with those derivatives alone.

Starting with Equation A.48 for the first derivative of $l(\beta, \tau, \sigma^2)$ with regards to β_0, Equations A.54 to A.58 show second derivatives of $l(\beta, \tau, \sigma^2)$ by taking the derivative of $\frac{\delta l(\beta,\tau,\sigma^2)}{\delta\beta_0}$ from Equation A.48 in terms of β_0 and β_1, τ_0, τ_1, and then σ^2. Equation A.48 does not contain τ_0, τ_1, or σ^2 which then leads to 0 values for the second derivatives in Equations A.56 to A.58 which deal with those parameters. The second derivatives $\frac{\delta^2 l(\beta,\tau)}{\delta\beta_0\delta\beta_0}$, $\frac{\delta^2 l(\beta,\tau)}{\delta\beta_0\delta\beta_1}$, and $\frac{\delta^2 l(\beta,\tau)}{\delta\beta_0\delta\beta_1}$ are identical to the second derivatives found in Appendix Section A.1.1 for ZIG.

$$\begin{aligned}\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_0\delta\beta_0} &= \sum_{i=1}^{n} -\frac{exp^{\beta_0+\beta_1 x_i}}{(1+exp^{\beta_0+\beta_1 x_i})^2} = \sum_{i=1}^{n} -p_i(1-p_i) \\ &= -n_0 p_0(1-p_0) - n_1 p_1(1-p_1)\end{aligned} \tag{A.54}$$

$$\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_0\delta\beta_1} = \sum_{i=1}^{n} -x_i\frac{exp^{\beta_0+\beta_1 x_i}}{(1+exp^{\beta_0+\beta_1 x_i})^2} = \sum_{i=1}^{n} -x_i p_i(1-p_i) = -n_1 p_1(1-p_1) \tag{A.55}$$

$$\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_0\delta\tau_0} = 0 \tag{A.56}$$

$$\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_0\delta\tau_1} = 0 \tag{A.57}$$

$$\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_0\delta\sigma^2} = 0 \tag{A.58}$$

Second derivatives involving at least one β_1 term follow from Equation A.49 and are outlined in Equations A.59 to A.62. As the first derivative of the log-likelihood with respect to β_1 does not contain τ_0, τ_1, or σ^2, the second derivatives with respect to β_1 and τ_0, τ_1, or σ^2 are equal to 0.

$$\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_1\delta\beta_1} = \sum_{i=1}^{n} -x_i^2 \frac{exp^{\beta_0+\beta_1 x_i}}{(1+exp^{\beta_0+\beta_1 x_i})^2} = \sum_{i=1}^{n} -x_1^2 p_i(1-p_i) = -n_1 p_1(1-p_1) \tag{A.59}$$

$$\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_1\delta\tau_0} = 0 \tag{A.60}$$

$$\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_1\delta\tau_1} = 0 \tag{A.61}$$

$$\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\beta_1\delta\sigma^2} = 0 \tag{A.62}$$

The lower triangle portion of the third column of the information matrix of second derivatives involvingτ_0 are shown in Equations A.63 to A.71. These follow from the first derivative of $l(\beta,\tau,\sigma^2)$ in terms of τ_0 shown in Equation A.50.

$$\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_0\delta\tau_0} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{I}(y_i>0) \tag{A.63}$$

$$\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_0\delta\tau_1} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{I}(y_i>0)x_i \tag{A.64}$$

$$\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_0\delta\sigma^2} = -\frac{2}{\sigma^4}\sum_{i=1}^{n}\mathbf{I}(y_i>0)\left(\log(y_i)-\tau_0-\tau_1 x_i\right) \tag{A.65}$$

The second derivatives involving τ_1 (except for those already outlined) are shown in Equation A.72 and Equation A.73. These follow from the first derivative of $l(\beta,\tau,\sigma^2)$ in terms of τ_1.

$$\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_1\delta\tau_1} = \frac{1}{\sigma^2}\sum_{i=1}^{n} -\mathbf{I}(y_i>0)x_i^2 \tag{A.66}$$

$$\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_0\delta\sigma^2} = -\frac{2}{\sigma^4}\sum_{i=1}^{n} -\mathbf{I}(y_i > 0)(\log(y_i) - \tau_0 - \tau_1 x_i) \tag{A.67}$$

Finally, the second derivative of $l(\beta,\tau,\sigma^2)$ in terms of σ^2 is:

$$\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\sigma^2\delta\sigma^2} = \frac{1}{2\sigma^4}\sum_{i=1}^{n}\mathbf{I}(y_i > 0) - \frac{1}{\sigma^6}\sum_{i=1}^{n}\mathbf{I}(y_i > 0)(\ln(y_i) - (\tau_0 + \tau_1 x_i))^2. \tag{A.68}$$

Taking the expectation of these second derivatives yields the expected information, $\mathcal{I}(\beta,\tau,\sigma^2)$. The first few columns of the information matrix (derivatives involving the β parameters) contain no y_i's and so for them the Fisher's information and the expected information are equivalent. For the other columns, the expected information will be created after taking the following expectations:

$$E\left(\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_0\delta\tau_0}\right) = -\frac{1}{\sigma^2}E\left(\sum_{i=1}^{n}\mathbf{I}(y_i > 0)\right) = -\frac{1}{\sigma^2}\sum_{i=1}^{n} p_i = -\frac{1}{\sigma^2}n_0 p_0 + n_1 p_1 \tag{A.69}$$

$$E\left(\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_0\delta\tau_1}\right) = -\frac{1}{\sigma^2}E\left(\sum_{i=1}^{n}\mathbf{I}(y_i > 0)x_i\right) = -\frac{1}{\sigma^2}\sum_{i=1}^{n} p_i x_i = -\frac{1}{\sigma^2}n_1 p_1 \tag{A.70}$$

$$E\left(\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_0\delta\sigma^2}\right) = -\frac{2}{\sigma^4}E\left(\sum_{i=1}^{n}\mathbf{I}(y_i > 0)\left(\log(y_i) - \tau_0 - \tau_1 x_i\right)\right) = 0 \tag{A.71}$$

$$E\left(\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_1\delta\tau_1}\right) = \frac{1}{\sigma^2}E\left(\sum_{i=1}^{n} -\mathbf{I}(y_i > 0)x_i^2\right) = -\frac{1}{\sigma^2}\sum_{i=1}^{n} p_0 x_i^2 = \frac{1}{\sigma^2}n_1 p_1 \tag{A.72}$$

$$E\left(\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\tau_0\delta\sigma^2}\right) = -\frac{2}{\sigma^4}E\left(\sum_{i=1}^{n} -\mathbf{I}(y_i > 0)(\log(y_i) - \tau_0 - \tau_1 x_i)\right) = 0 \tag{A.73}$$

and,

$$E\left(\frac{\delta^2 l(\beta,\tau,\sigma^2)}{\delta\sigma^2\delta\sigma^2}\right) = \frac{1}{2\sigma^4}E\left(\sum_{i=1}^{n}\mathbf{I}(y_i>0)\right) - \frac{1}{\sigma^6}E\left(\sum_{i=1}^{n}\mathbf{I}(y_i>0)(\ln(y_i)-(\tau_0+\tau_1 x_i))^2\right) = \frac{n_0p_0+n_1p_1}{2\sigma^4} - \frac{\sigma^2(n_0p_0+n_1p_1)}{\sigma^6} = -\frac{n_0p_0+n_1p_1}{2\sigma^4} \tag{A.74}$$

Using these results, the parameters from a ZILN model have the following joint distributional properties:

$$\sqrt{n}\left(\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\tau}_0\\ \hat{\tau}_1\\ \hat{\sigma}^2\end{pmatrix} - \begin{pmatrix}\beta_0\\ \beta_1\\ \tau_0\\ \tau_1\\ \sigma^2\end{pmatrix}\right) \xrightarrow{d} \mathrm{N}(0,\Sigma) \tag{A.75}$$

where $\Sigma = \mathrm{n}\mathcal{I}^{-1}(\beta,\tau,\sigma^2)$ with n being the overall sample size and

$$\Sigma = \begin{pmatrix} \frac{n}{n_0p_0(1-p_0)} & -\frac{n}{n_0p_0(1-p_0)} & 0 & 0 & 0\\ -\frac{n}{n_0p_0(1-p_0)} & \frac{n}{n_0p_0(1-p_0)} + \frac{n}{n_1p_1(1-p_1)} & 0 & 0 & 0\\ 0 & 0 & \sigma^2\frac{n}{n_0p_0} & -\sigma^2\frac{n}{n_0p_0} & 0\\ 0 & 0 & -\sigma^2\frac{n}{n_0p_0} & \sigma^2\left(\frac{n}{n_0p_0}+\frac{n}{n_1p_1}\right) & 0\\ 0 & 0 & 0 & 0 & 2\sigma^4\frac{n}{n_0p_0+n_1p_1} \end{pmatrix}. \tag{A.76}$$

This can be written to be free of sample size by writing Σ in terms of weights where $w_0 = \frac{n_0}{n}$ is the proportion of the sample that is in group 0 and $w_1 = \frac{n_1}{n}$ is the proportion of the sample that is in group 1.

$$\Sigma = \begin{pmatrix} \frac{1}{w_0p_0(1-p_0)} & -\frac{1}{w_0p_0(1-p_0)} & 0 & 0 & 0 \\ -\frac{1}{w_0p_0(1-p_0)} & \frac{1}{w_0p_0(1-p_0)} + \frac{1}{w_1p_1(1-p_1)} & 0 & 0 & 0 \\ 0 & 0 & \frac{\sigma^2}{w_0p_0} & -\frac{\sigma^2}{w_0p_0} & 0 \\ 0 & 0 & -\sigma^2\frac{1}{w_0p_0} & \sigma^2\left(\frac{1}{w_0p_0} + \frac{1}{w_1p_1}\right) & 0 \\ 0 & 0 & 0 & 0 & 2\sigma^4\frac{1}{w_0p_0+w_1p_1} \end{pmatrix}. \tag{A.77}$$

It can be also written in a block-diagonal form as

$$\Sigma_{\beta,\tau,\sigma^2} = \begin{pmatrix} n(X'D_{p(1-p)}X)^{-1} & \mathbf{0} & 0 \\ \mathbf{0} & \sigma^2(X'D_pX)^{-1} & 0 \\ \mathbf{0} & \mathbf{0} & 2\sigma^4\frac{1}{w_0p_0+w_1p_1} \end{pmatrix} \tag{A.78}$$

A.2.2 Approximate Distribution for Difference of Means

The Delta method states that given $\sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{d} N\left(0, \Sigma\right)$ and a function $\mathrm{g}(\theta)$ with first derivatives $\Delta = \left(\frac{\delta \mathrm{g}(\theta)}{\delta\theta_i}\right)$, if $\frac{\delta \mathrm{g}(\theta)}{\delta\theta_i}$ is continuous for all i, and $\frac{\delta \mathrm{g}(\theta)}{\delta\theta_i} \neq 0$ for some i then[31, 32],

$$\sqrt{n}\left(\mathrm{g}(\hat{\theta}) - \mathrm{g}(\theta)\right) \xrightarrow{d} \mathrm{N}(0, \Delta\Sigma\Delta^T). \tag{A.79}$$

In the ZILN model set up in the preceding pages, $\theta = (\beta_0, \beta_1, \tau_0, \tau_1, \sigma^2)$, and $\mathrm{g}(\theta) = \mathrm{g}(\beta_0, \beta_1, \tau_0, \tau_1, \sigma^2)$ could be any function of β_0, β_1, τ_0, τ_1, and σ^2. For example, the difference of means between two groups can be written as function of β_0, β_1, τ_0, τ_1, and σ^2. Then the delta method mays be applied to find its asymptotic distribution.

The difference of means between two groups under the ZILN model could be expressed as follows:

$$\begin{aligned}
\mathrm{g}(\beta,\tau,\sigma^2) = DM &= E(Y_i|X=1) - E(Y_i|X=0) \\
&= E(Y_i|Y_i>0, X=1)P(Y_i>0|X=1) \\
&- E(Y_i|Y_i>0, X=0)P(Y_i>0|X=0) \\
&= e^{\tau_0+\tau_1+\frac{1}{2}\hat{\sigma}^2}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} - e^{\tau_0+\frac{1}{2}\hat{\sigma}^2}\frac{e^{\beta_0}}{1+e^{\beta_0}}.
\end{aligned} \tag{A.80}$$

Then a vector of derivatives, Δ_{DM}, could be constructed such that

$$\Delta_{DM}^T = \begin{pmatrix} \frac{\delta DM}{\delta\beta_0} \\ \frac{\delta DM}{\delta\beta_1} \\ \frac{\delta DM}{\delta\tau_0} \\ \frac{\delta DM}{\delta\tau_1} \\ \frac{\delta DM}{\delta\sigma^2} \end{pmatrix} = \begin{pmatrix} e^{\tau_0+\tau_1+\frac{1}{2}\hat{\sigma}^2}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} - e^{\tau_0+\frac{1}{2}\hat{\sigma}^2}\frac{e^{\beta_0}}{(1+e^{\beta_0})^2} \\ e^{\tau_0+\tau_1+\frac{1}{2}\hat{\sigma}^2}\frac{e^{\beta_0+\beta_1}}{(1+e^{\beta_0+\beta_1})^2} \\ e^{\tau_0+\tau_1+\frac{1}{2}\hat{\sigma}^2}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} - e^{\tau_0+\frac{1}{2}\hat{\sigma}^2}\frac{e^{\beta_0}}{1+e^{\beta_0}} \\ e^{\tau_0+\tau_1+\frac{1}{2}\hat{\sigma}^2}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} \\ \frac{1}{2}e^{\tau_0+\tau_1+\frac{1}{2}\hat{\sigma}^2}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} - \frac{1}{2}e^{\tau_0+\frac{1}{2}\hat{\sigma}^2}\frac{e^{\beta_0}}{1+e^{\beta_0}} \end{pmatrix} \tag{A.81}$$

which can be simplified as follows where p_j is the probability of a non-zero outcome for group j and μ_j is the mean of the non-zero values for $j = 0, 1$:

$$\Delta_{DM}^T = \begin{pmatrix} p_1(1-p_1)\mu_1 - p_0(1-p_0)\mu_0 \\ p_1(1-p_1)\mu_1 \\ p_1\mu_1 - p_0\mu_0 \\ p_1\mu_1 \\ \frac{1}{2}p_1\mu_1 - \frac{1}{2}p_0\mu_0 \end{pmatrix}. \tag{A.82}$$

These derivatives are all continuous for all values of β_0, β_1, τ_0, τ_1, and σ^2. A second conditions holds since $\frac{\delta DM}{\delta\tau_1} = e^{\tau_0+\tau_1+\sigma^2}\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$ is greater than zero for all values of β_0, β_1, τ_0, τ_1, and σ^2. This is clearly true since $e^{\tau_0+\tau_1}$ is greater than zero for all τ_0 and τ_1 values and $\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$ is greater than zero for all values of β_0 and β_1. Since these conditions are met, the delta method may be used.

Then, utilizing the delta method explained in Equation A.79, Σ the asymptotic variance of β_0, β_1, τ_0, τ_1, and σ^2 from Equation A.76, and the vector of derivatives

Δ_{DM} from Equation A.82, the asymptotic distribution of DM can be established as

$$\sqrt{n}\left(DM(\hat{\beta},\hat{\theta},\hat{\sigma}^2) - DM(\beta,\theta,\sigma^2)\right) \xrightarrow{d} \mathrm{N}(0, \Delta_{DM}\Sigma\Delta_D M^T) \tag{A.83}$$

where $\Delta_{DM}\Sigma\Delta_{DM}^T$ is found in Equation A.85. For the sake of space we begin by defining $\Delta_{DM}\Sigma$ in Equation A.84 as

$$\begin{aligned}
&\Delta_{DM}\Sigma = \\
&\begin{pmatrix} p_1(1-p_1)\mu_1 - p_0(1-p_0)\mu_0 & p_1(1-p_1)\mu_1 & p_1\mu_1 - p_0\mu_0 & p_1\mu_1 & \frac{1}{2}p_1\mu_1 - \frac{1}{2}p_0\mu_0 \end{pmatrix} \\
&\times \begin{pmatrix} \frac{1}{w_0p_0(1-p_0)} & -\frac{1}{w_0p_0(1-p_0)} & 0 & 0 & 0 \\ -\frac{1}{w_0p_0(1-p_0)} & \frac{1}{w_0p_0(1-p_0)} + \frac{1}{w_1p_1(1-p_1)} & 0 & 0 & 0 \\ 0 & 0 & \frac{\sigma^2}{w_0p_0} & -\frac{\sigma^2}{w_0p_0} & 0 \\ 0 & 0 & -\sigma^2\frac{1}{w_0p_0} & \sigma^2\left(\frac{1}{w_0p_0} + \frac{1}{w_1p_1}\right) & 0 \\ 0 & 0 & 0 & 0 & 2\sigma^4\frac{1}{w_0p_0+w_1p_1} \end{pmatrix} \\
&= \begin{pmatrix} -\frac{\mu_0}{w_0} & \frac{\mu_0}{w_0} + \frac{\mu_1}{w_1} & -\frac{p_0\mu_0\sigma^2}{w_0p_0} & \frac{p_0\mu_0\sigma^2}{w_0p_0} + \frac{p_1\mu_1\sigma^2}{w_1p_1} & \frac{\sigma^4}{w_0p_0+w_1p_1}(p_1\mu_1 - p_0\mu_0) \end{pmatrix}.
\end{aligned} \tag{A.84}$$

Then in Equation A.85 we define the final $\Delta_{DM}\Sigma\Delta_{DM}^T$ as

$$\begin{aligned}
&\Delta_{DM}\Sigma\Delta_{DM}^T = \\
&= \begin{pmatrix} -\frac{\mu_0}{w_0} & \frac{\mu_0}{w_0} + \frac{\mu_1}{w_1} & -\frac{p_0\mu_0\sigma^2}{w_0p_0} & \frac{p_0\mu_0\sigma^2}{w_0p_0} + \frac{p_1\mu_1\sigma^2}{w_1p_1} & \frac{\sigma^4}{w_0p_0+w_1p_1}(p_1\mu_1 - p_0\mu_0) \end{pmatrix} \\
&\times \begin{pmatrix} p_1(1-p_1)\mu_1 - p_0(1-p_0)\mu_0 \\ p_1(1-p_1)\mu_1 \\ p_1\mu_1 - p_0\mu_0 \\ p_1\mu_1 \\ \frac{1}{2}p_1\mu_1 - \frac{1}{2}p_0\mu_0 \end{pmatrix} \\
&= \frac{\mu_0^2p_0(1-p_0)}{w_0} + \frac{\mu_1^2p_1(1-p_1)}{w_1} + \frac{\mu_0^2p_0\sigma^2}{w_0} + \frac{\mu_1^2p_1\sigma^2}{w_1} + \frac{\sigma^4}{2(w_0p_0+w_1p_1)}(p_1\mu_1 - p_0\mu_0)^2.
\end{aligned} \tag{A.85}$$

From this result, a variance of $DM(\hat{\beta},\hat{\tau},\hat{\sigma}^2)$ can be defined for finite sample sizes. Specifically, since $w_0 = \frac{n_0}{n}$ and $w_1 = \frac{n_1}{n}$,

$$\begin{aligned} n\Delta_{DM}\Sigma\Delta_{DM}^T =& Var(DM(\hat{\beta}, \hat{\tau})) \\ =& \frac{\mu_0^2 p_0(1-p_0)}{n_0} + \frac{\mu_1^2 p_1(1-p_1)}{n_1} \\ &+ \frac{\mu_0^2 p_0 \sigma^2}{n_0} + \frac{\mu_1^2 p_1 \sigma^2}{n_1} \\ &+ \frac{\sigma^4}{2(n_0 p_0 + n_1 p_1)}(p_1\mu_1 - p_0\mu_0)^2. \end{aligned} \tag{A.86}$$

This variance can then be used to create Wald confidence intervals or Wald-type hypothesis tests as was done in Chapter 3.

Similar methods (not shown) were used to find the asymptotic distribution $\log(RM)$ for the non-adjusting ZILN regression case (used in Chapter 3) as well as DM, $\log(RM_{SS})$, and $\log(RM_{MAR})$ when adjusting for covariates (used in Chapter 4).

APPENDIX B
R CODE

B.1 Introduction

This appendix, includes the R code which was used for the simulations and the data examples. In practice, for the analysis portion of these codes it may be possible to combine canned R functions and some additional coding for a shorter set of code that may also have better convergence properties.

The Fisher scoring algorithm for the logistic regression and the gamma regression parts of the ZIG model, code for the estimation of the log-normal part of the ZILN model, and code for AIC estimation are show in Section B.2. Then in Section B.3 code for the DM and RM tests without covariate adjustment are include for both ZIG and ZILN models. Code for DM, RM_{SS}, and RM_{MAR} are show in Section B.4. And, code for calculating AIC for ZIG and ZILN models is shown in Section B.5.

B.2 Model Estimation Code

B.2.1 Fisher Scoring Algorithms

```
#Step-Halving
step.halve<-function(d,loglik.old,loglik.new,x,y,bhat,bdiff)
{eps <- 1e-04
 if (loglik.old <= loglik.new + eps)
 {step.ok <- 1
  return(c(step.ok,bhat,loglik.new))
 }
 else
 {bhat <- bhat - bdiff
  for (step in 1:10)
  {bdiff <- bdiff/2
   bhat <- bhat + bdiff
   if (d=="gamma")
   {muhat<-exp(x%*%bhat)
    loglik.new <- sum(-log(muhat)-y/muhat)
```

```
    }
    if (d=="bern")
    {niln<-log(1/(1+exp(x %*% bhat)))
     niln[is.na(niln)] <- 0
     loglik.new<- crossprod(x %*% bhat,y) - sum(niln)
    }
    if(loglik.new + eps >= loglik.old)
    {step.ok <- 2
     return(c(step.ok,bhat,loglik.new))
    }
   }
   step.ok <- 0
   c(step.ok,bhat,loglik.new)
  }
 }
#Fisher Scoring for binomial
glm.bin<-function(x,y,nit,tol=0.0000001)
{if (sum(y)==0)
   cat("\n\nAll the Ys are zeros!\n\n")
  else
   y.b<-ifelse(y==0,0,1)
  #Initial estimate of intercept
  b0.b <- log(mean(y.b)/(1-mean(y.b)))
  #Setup for Iteration Routine
  bdiff.b <- 1
  if (length(x)==length(y))
  {nb<-1
   bhat.b<-t(b0.b)
  }
  else
  {nb<-dim(x)[2]
   bhat.b <- c(b0.b,rep(0,nb-1))
  }
  it <- 0
  space <- " "
  niln<-log(1/(1+exp(x%*%bhat.b)))
  niln[is.na(niln)] <- 0
  loglik.null.b <- crossprod(x%*%bhat.b,y.b)+sum(niln)
  loglik.old.b <- loglik.null.b

  #Finding MLEs and standard errors
  while (it <= nit && max(abs(bdiff.b)) > tol)
  {it <- it+1
   p <- exp(x%*%bhat.b)/(1+exp(x%*%bhat.b))
   inf <- t(x)%*%(diag(length(p))*as.vector(p*(1-p)))%*% x
```

```
  resid.b <- y.b-p
  score.b<-t(x)%*%(resid.b)
  bhat.b.old<-bhat.b
  bhat.b <- bhat.b+solve(inf)%*%score.b
  p<-exp(x%*%bhat.b)/(1+exp(x%*%bhat.b))
  niln<-log(1/(1+exp(x%*%bhat.b)))
  niln[is.na(niln)] <- 0
  loglik.b <- crossprod(x%*%bhat.b,y.b)+sum(niln)
  bdiff.b<-bhat.b-bhat.b.old

  st.res <- step.halve('bern',loglik.old.b,loglik.b,x,y.b,
                      bhat.b,bdiff.b)
  if (st.res[1] == 2)
  {bhat.b <- st.res[2:(nb+1)]
   loglik.b <- st.res[(nb+2)] }
   if (st.res[1] == 0)
   {cat("Step halving failed to find better estimates\n")
    break }
   loglik.old.b <- loglik.b
  }
  #Calculate Results
  if (it > nit)
  {cat("Failed to converge in",nit,"steps\n")
   iconvg <- 0
  }
  else
   if (st.res[1] == 0)
    cat("Step halving failed\n")
  else
  {niln<-log(1/(1+exp(x%*%bhat.b)))
   niln[is.na(niln)] <- 0
   #log-likelihood at mle
   loglik.b <- crossprod(x%*%bhat.b,y.b) + sum(niln)
   inf <- t(x)%*%(diag(length(p))*as.vector(p*(1-p)))%*%x
   resid.b <- y.b - p
   score.b<-t(x)%*%resid.b
   infinv <- solve(inf)
   sd1.b <- sqrt(diag(infinv))
  }
 return(list(llb=loglik.b,b.b=bhat.b,sd1.b=sd1.b,
            score.b=score.b,invinf.b=infinv))
}

#Fisher Scoring for gamma
glm.gamma<-function(x,y,nit,tol=0.00000001)
```

```
{y.c<-y[y!=0]    #non-zero y-values
 #Initial estimate of intercept
 b0 <- log(mean(y.c))
 #Saturated log-likelihood/v (scaled; c(y,v) dropped)
 loglik.sat <- sum(-log(y.c)-1)
 #Setup for Iteration Routine
 bdiff <- 1
 nc<-dim(as.matrix(x))[2]#number of columns in the x matrix
 #x.c contains the rows in x corresponding to non-zero y's
 x.c<-matrix(x[y!=0],nrow=length(y.c),ncol=nc)
 bhat <- c(b0,rep(0,nc-1))
 it <- 0
 space <- " "
 muhat<-exp(x.c%*%bhat)
 #Null log-likelihood/ v (scaled version; dropping c(y,v))
 #Note:scaled v cancels out in maximization
 #      c(y,v) doesn't impact maximization
 loglik.null <- sum(-log(muhat)-y.c/muhat)
 deviance.null <- 2*(loglik.sat - loglik.null)
 loglik.old <- loglik.null

 #Finding MLEs and standard errors
 while ( it <= nit && max(abs(bdiff))  > tol   )
 {it <- it+1
  muhat<-exp(x.c%*%bhat)
  resid <- y.c - muhat
  bhat.old <- bhat
  inf<-crossprod(x.c,x.c)#unscaled
  score<-t(x.c)%*%(resid/muhat)#unscaled
  bhat<-bhat.old+solve(inf)%*%score
  muhat.old<-muhat
  muhat<-exp(x.c%*%bhat)
  loglik <- sum(-log(muhat)-y.c/muhat)
  bdiff<-bhat-bhat.old
  #Step-Halving
  st.res <- step.halve('gamma',loglik.old,loglik,x.c,y.c,
                       bhat,bdiff)
  if (st.res[1] == 2)
  {bhat <- st.res[2:(nc+1)]
   loglik <- st.res[(nc+2)]
  }
  if (st.res[1] == 0)
  {cat("Step halving failed to find better estimates\n")
   break
  }
```

```
    loglik.old <- loglik
  }
 #Calculate Results
  if (it > nit)
  {cat("Failed to converge in",nit,"steps\n")
   iconvg <- 0
  }
  else
   if (st.res[1] == 0)
    cat("Step halving failed\n")
  else
  {iconvg <- 1
   muhat<-exp(x.c%*%bhat)
   loglik <- sum(-log(muhat)-y.c/muhat)
   deviance <- 2*(loglik.sat - loglik)
   vinv.mle<-deviance/(length(y.c)-nc)#bias corrected.
   covb.mle<-vinv.mle*solve(crossprod(x.c,x.c))

   sdb.mle<-sqrt(diag(covb.mle))
   resid.c <- y.c - muhat
   score<-t(x.c)%*%(resid.c/muhat)
  }
return(list(llc=loglik,b.c=bhat,sd1.c.mle=sdb.mle,
              disp.mle=vinv.mle,invinf.c=covb.mle,
              score.c=score, dev=deviance,
              ncont=length(y.c),resid.c=resid.c))
}

#Putting the two parts together
zig.glm<-function(x,y,nit,tol=0.00000001)
{glmb<-glm.bin(x,y,nit,tol)
 glmg<-glm.gamma(x,y,nit,tol)
 llzig<-glmb$llb+glmg$llc
 py<- exp(x%*%glmb$b.b)/(1+exp(x%*%glmb$b.b))
 muhat<-exp(x%*%glmg$b.c)
 pred.y<-py*muhat
 resid<-y-pred.y
return(list(llb=glmb$llb,llc=glmg$llc,llzic=llzig,
              b.b=glmb$b.b,b.c=glmg$b.c,sd1.b=glmb$sd1.b,
              sd1.c.mle=glmg$sd1.c.mle,disp.mle=glmg$disp.mle,
              invinf.b=glmb$invinf.b,invinf.c=glmg$invinf.c,
              score.b=glmb$score.b,score.c=glmg$score.c,
              dev.c=glmg$dev,ncont=glmg$ncont,
              resid.c=glmg$resid.c,pred.y=pred.y, resid=resid))
}
```

B.2.2 Log-Normal Regression Estimation

```
lm.lnorm<-function(x,y)
{
 if (sum(y)==0)
  cat("\n\nAll the Ys are zeros!\n\n")
 y.c<-y[y!=0]
 if (length(x)==length(y))
  x.c<-matrix(x[y!=0],nrow=length(y.c),ncol=1)
 else
  x.c<-matrix(x[y!=0],nrow=length(y.c),ncol=dim(x)[2])
 nc<-dim(x.c)[2]
 invinf.unsc<-solve(crossprod(x.c,x.c))#unscaled
 bhat<-invinf.unsc%*%crossprod(x.c,log(y.c))
 resid.c<-log(y.c)-x.c%*%bhat
 sigma2<-crossprod(resid.c)/(length(y.c)-nc)
 sdb.c<-sqrt(sigma2*diag(invinf.unsc))
 llc<-crossprod(resid.c)
return(list(llc=llc,b.c=bhat,sdb.c=sdb.c,sigma2=sigma2,
             invinf.c=as.numeric(sigma2)*invinf.unsc,
             resid.c=resid.c))
}


#Putting ZILN pieces together

ziln.glm<-function(x,y,nit)
{
 glmb<-glm.bin(x,y,nit)
 glmln<-lm.lnorm(x,y)
 llziln<-glmb$llb+glmln$llc
 py<- exp(x%*%glmb$b.b)/(1+exp(x%*%glmb$b.b))
 muhat<-exp(x%*%glmln$b.c)
 pred.y<-py*muhat
 resid<-y-pred.y
return(list(llb=glmb$llb,llc=glmln$llc,llzic=llziln,
             b.b=glmb$b.b,b.c=glmln$b.c,
             sd1.b=glmb$sd1.b, sd1.c=glmln$sdb.c,
             disp=glmln$sigma2,invinf.b=glmb$invinf.b,
             invinf.c=glmln$invinf.c,resid.c=glmln$resid.c,
             pred.y=pred.y, resid=resid))
}
```

B.3 Code for Mean-Based Two Group Comparison Without Covariate Adjustment

B.3.1 Code for Mean-Based Tests for Zero Inflated Gamma

```
zig.uni<-function(mod,n0,n1)
{n<-n0+n1
 p1<-exp(mod$b.b[1]+mod$b.b[2])/(1+exp(mod$b.b[1]+mod$b.b[2]))
 p0<-exp(mod$b.b[1])/(1+exp(mod$b.b[1]))
 n0c<-n0*p0
 n1c<-n1*p1
 nc<-n0c+n1c
 M0<-exp(mod$b.c[1])*p0
 M1<-exp(mod$b.c[1]+mod$b.c[2])*p1

 DM<-exp(mod$b.c[1]+mod$b.c[2])*p1-exp(mod$b.c[1])*p0
 VDM<-exp(2*mod$b.c[1])*p0*(1-p0)/n0+
           exp(2*mod$b.c[1]+2*mod$b.c[2])*p1*(1-p1)/n1+
         p0^2*exp(2*mod$b.c[1])*mod$disp.mle/n0c+
         p1^2*exp(2*(mod$b.c[1]+mod$b.c[2]))*mod$disp.mle/n1c
           #Note: p0^2/n0c = p0/n0
 DMtest<-DM^2/VDM
 DM.p<-1-pchisq(DMtest,1)

 lRM<-mod$b.c[2] + mod$b.b[2] + log(1-p1) - log(1-p0)
 VlRM<-mod$disp.mle*(nc/(n0c*n1c))+
                   (1-p0)/(n0*p0)+(1-p1)/(n1*p1)
 #Note: nc/(n0c*n1c)=(n0*p0+n1*p1)/(n0*p0*n1*p1)=
 #        1/(n0*p0)+1/(n1*p1)
 lRMtest<-lRM^2/VlRM
 lRM.p<-1-pchisq(lRMtest,1)
return(list(M0=M0,M1=M1,DM=DM,VDM=VDM,
                 DMtest=DMtest,DM.p=DM.p,
                 lRM=lRM,VlRM=VlRM,lRMtest=lRMtest,lRM.p=lRM.p))
}
wilcoxon<-function(yw,xw)
{r<-rank(yw); s2r<-var(r);
 r0<-mean(r[xw==0]);r1<-mean(r[xw==1]);
 Wlcx<-(r0-r1)^2/(s2r*(1/length(r[xw==0])+1/length(r[xw==1])))
 return(Wlcx)
}
```

B.3.2 Code for Mean-Based Tests for Zero Inflated Log-Normal

```
ziln.uni<-function(mod,n0,n1)
{n<-n0+n1
 p1<-exp(mod$b.b[1]+mod$b.b[2])/(1+exp(mod$b.b[1]+mod$b.b[2]))
 p0<-exp(mod$b.b[1])/(1+exp(mod$b.b[1]))
 n0c<-n0*p0
 n1c<-n1*p1
 nc<-n0c+n1c
 M0<-exp(mod$b.c[1]+mod$disp/2)*p0
 M1<-exp(mod$b.c[1]+mod$b.c[2]+mod$disp/2)*p1

 DM<-M1-M0
 VDM<-exp(2*mod$b.c[1]+mod$disp)*p0*(1-p0)/n0+
       exp(2*mod$b.c[1]+2*mod$b.c[2]+mod$disp)*p1*(1-p1)/n1+
       p0^2*exp(2*mod$b.c[1]+mod$disp)*mod$disp/n0c+
       p1^2*exp(2*(mod$b.c[1]+
                   mod$b.c[2]+mod$disp/2))*mod$disp/n1c+
       mod$disp^2/(2*nc)*(exp(mod$b.c[1]+
                          mod$b.c[2]+mod$disp/2)*p1-
                       exp(mod$b.c[1]+mod$disp/2)*p0)^2
 #Note: p0^2/n0c = p0/n0

 DMtest<-DM^2/VDM
 DM.p<-1-pchisq(DMtest,1)

 lRM<-mod$b.c[2] + mod$b.b[2] + log(1-p1) - log(1-p0)
 VlRM<-mod$disp*(nc/(n0c*n1c))+(1-p0)/(n0*p0)+(1-p1)/(n1*p1)
 #Note: nc/(n0c*n1c)=(n0*p0+n1*p1)/(n0*p0*n1*p1)=
 #                    1/(n0*p0)+1/(n1*p1)
 lRMtest<-lRM^2/VlRM
 lRM.p<-1-pchisq(lRMtest,1)

return(list(M0=M0,M1=M1,
            DM=DM,VDM=VDM,DMtest=DMtest,DM.p=DM.p,
            lRM=lRM,VlRM=VlRM,lRMtest=lRMtest,lRM.p=lRM.p))
}
```

B.4 Code for Two Group Comparison With Covariate Adjustment

```
require(Matrix)
mzg.DMRM.adj<-function(x,mod)
{#Test coded for 2nd variable only DM test
 xg1<-cbind(x[,1],rep(1,length(x[,2])),x[,3])
 xg0<-cbind(x[,1],rep(0,length(x[,2])),x[,3])
 mu.xg1<-exp(xg1%*%mod$b.c); mu.xg0<-exp(xg0%*%mod$b.c);
 p.xg1<-(exp(xg1%*%mod$b.b)/(1+exp(xg1%*%mod$b.b)))
 p.xg0<-(exp(xg0%*%mod$b.b)/(1+exp(xg0%*%mod$b.b)))
 p<-(exp(x%*%mod$b.b)/(1+exp(x%*%mod$b.b)))
 M1<-mu.xg1*p.xg1;   M0<-mu.xg0*p.xg0;
 Dp<-(diag(length(p))*as.vector(p))
 invinf.c<-mod$disp.mle*solve(t(x)%*%Dp%*%x)
 Ibcinv<-as.matrix(bdiag(mod$invinf.b,invinf.c))
 r1mp<-mean((1-p.xg1)/(1-p.xg0))

 DMbar<-mean(M1-M0)
 lRMss<-log(mean(M1/M0))
 lRMmar<-log(mean(M1)/mean(M0))
 M0mn<-mean(M0);  M1mn<-mean(M1);

 DMderb0<-mean(p.xg1*(1-p.xg1)*mu.xg1-p.xg0*(1-p.xg0)*mu.xg0)
 DMderb1<-mean(p.xg1*(1-p.xg1)*mu.xg1)
 DMderb2<-mean(x[,2]*(p.xg1*(1-p.xg1)*mu.xg1-
                        p.xg0*(1-p.xg0)*mu.xg0))
 DMdert0<-DMbar
 DMdert1<-mean(p.xg1*mu.xg1);  DMdert2<-mean(x[,2]*DMbar);
 DMder<-t(c(DMderb0,DMderb1,DMderb2,DMdert0,DMdert1,DMdert2))
 DMbarvar<-DMder%*%Ibcinv%*%t(DMder)
 DMtest<-DMbar/sqrt(DMbarvar)

 lRMssderb0<-mean((1-p.xg1)*(exp(xg0%*%mod$b.b)-
                              p.xg1/p.xg0))/r1mp
 lRMssderb1<-mean((1-p.xg1)^2/(1-p.xg0))/r1mp
 lRMssderb2<-mean(x[,3]*((1-p.xg1)*(exp(xg0%*%mod$b.b)-
                                     p.xg1/p.xg0))))/r1mp
 lRMssdert0<-0;    lRMssdert1<-1;    lRMssdert2<-0;
 lRMssder<-t(c(lRMssderb0,lRMssderb1,lRMssderb2,
               lRMssdert0,lRMssdert1,lRMssdert2))
 lRMssvar<-lRMssder%*%Ibcinv%*%t(lRMssder)
 lRMsstest<-lRMss/sqrt(lRMssvar)
```

```
  lRMmarderb0<-mean(p.xg1*(1-p.xg1)*mu.xg0)/
                  mean(p.xg1*mu.xg0)-
                 mean(p.xg0*(1-p.xg0)*mu.xg0)/mean(p.xg0*mu.xg0)
  lRMmarderb1<-mean(p.xg1*(1-p.xg1)*mu.xg0)/mean(p.xg1*mu.xg0)
  lRMmarderb2<-mean(x[,3]*p.xg1*(1-p.xg1)*mu.xg0)/
                  mean(p.xg1*mu.xg0)-
                 mean(x[,3]*p.xg0*(1-p.xg0)*mu.xg0)/
                 mean(p.xg0*mu.xg0)
  lRMmardert0<-0;        lRMmardert1<-1;
  lRMmardert2<-mean(x[,3]*p.xg1*exp(xg0%*%mod$b.c))/
                  mean(p.xg1*exp(xg0%*%mod$b.c))-
                 mean(x[,3]*p.xg0*exp(xg0%*%mod$b.c))/
                  mean(p.xg0*exp(xg0%*%mod$b.c))
  lRMmarder<-t(c(lRMmarderb0,lRMmarderb1,lRMmarderb2,
                   lRMmardert0,lRMmardert1,lRMmardert2))
  lRMmarvar<-lRMmarder%*%Ibcinv%*%t(lRMmarder)
  lRMmartest<-lRMmar/sqrt(lRMmarvar)

 return(list(M0=M0mn, M1=M1mn, DMbar=DMbar, DMbarvar=DMbarvar,
             DMtest=DMtest, lRMss=lRMss, lRMssvar=lRMssvar,
             lRMsstest=lRMsstest,lRMmar=lRMmar,
             lRMmarvar=lRMmarvar,lRMmartest=lRMmartest))
}

mzln.DMRM.adj<-function(x, mod)
{xg1<-cbind(x[,1],rep(1,length(x[,2])),x[,3])
 xg0<-cbind(x[,1],rep(0,length(x[,2])),x[,3])

 mu.xg1<-exp(xg1%*%mod$b.c+.5*as.numeric(mod$disp));
 mu.xg0<-exp(xg0%*%mod$b.c+.5*as.numeric(mod$disp));
 p.xg1<-(exp(xg1%*%mod$b.b)/(1+exp(xg1%*%mod$b.b)))
 p.xg0<-(exp(xg0%*%mod$b.b)/(1+exp(xg0%*%mod$b.b)))
 p<-(exp(x%*%mod$b.b)/(1+exp(x%*%mod$b.b)))
 M1<-mu.xg1*p.xg1;     M0<-mu.xg0*p.xg0;
 Dp<-(diag(length(p))*as.vector(p))
 invinf.c<-as.numeric(mod$disp)*solve(t(x)%*%Dp%*%x)
 Ibcinv<-as.matrix(bdiag(mod$invinf.b,invinf.c,
                         2*as.numeric(mod$disp)^2/sum(p)))
 r1mp<-mean((1-p.xg1)/(1-p.xg0))

 DMbar<-mean(M1-M0)
 lRMss<-log(mean(M1/M0))
```

```
lRMmar<-log(mean(M1)/mean(M0))
M0mn<-mean(M0);          M1mn<-mean(M1);

DMderb0<-mean(p.xg1*(1-p.xg1)*mu.xg1-p.xg0*(1-p.xg0)*mu.xg0)
DMderb1<-mean(p.xg1*(1-p.xg1)*mu.xg1)
DMderb2<-mean(x[,3]*(p.xg1*(1-p.xg1)*mu.xg1-
                     p.xg0*(1-p.xg0)*mu.xg0))
DMdert0<-DMbar;           DMdert1<-mean(p.xg1*mu.xg1);
DMdert2<-mean(x[,3]*DMbar);
DMdersigma2<-.5*DMbar

DMder<-t(c(DMderb0,DMderb1,DMderb2,
           DMdert0,DMdert1,DMdert2,DMdersigma2))
DMbarvar<-DMder%*%Ibcinv%*%t(DMder)
DMtest<-DMbar/sqrt(DMbarvar)

lRMssderb0<-mean((1-p.xg1)*(exp(xg0%*%mod$b.b)-
                             p.xg1/p.xg0))/r1mp
lRMssderb1<-mean((1-p.xg1)^2/(1-p.xg0))/r1mp
lRMssderb2<-mean(x[,3]*((1-p.xg1)*(exp(xg0%*%mod$b.b)-
                             p.xg1/p.xg0))))/r1mp
lRMssdert0<-0;      lRMssdert1<-1;      lRMssdert2<-0;
lRMssdersigma2<-0
lRMssder<-t(c(lRMssderb0,lRMssderb1,lRMssderb2,lRMssdert0,
              lRMssdert1,lRMssdert2,lRMssdersigma2))
lRMssvar<-lRMssder%*%Ibcinv%*%t(lRMssder)
lRMsstest<-lRMss/sqrt(lRMssvar)

lRMmarderb0<-mean(p.xg1*(1-p.xg1)*mu.xg0)/
             mean(p.xg1*mu.xg0)-
            mean(p.xg0*(1-p.xg0)*mu.xg0)/mean(p.xg0*mu.xg0)
lRMmarderb1<-mean(p.xg1*(1-p.xg1)*mu.xg0)/mean(p.xg1*mu.xg0)
lRMmarderb2<-mean(x[,3]*p.xg1*(1-p.xg1)*mu.xg0)/
             mean(p.xg1*mu.xg0)-
            mean(x[,3]*p.xg0*(1-p.xg0)*mu.xg0)/
             mean(p.xg0*mu.xg0)
lRMmardert0<-0;       lRMmardert1<-1;
lRMmardert2<-mean(x[,3]*p.xg1*exp(xg0%*%mod$b.c))/
             mean(p.xg1*exp(xg0%*%mod$b.c))-
            mean(x[,3]*p.xg0*exp(xg0%*%mod$b.c))/
             mean(p.xg0*exp(xg0%*%mod$b.c))
lRMmardersigma2<-0
```

```
 lRMmarder<-t(c(lRMmarderb0,lRMmarderb1,lRMmarderb2,
                    lRMmardert0,lRMmardert1,lRMmardert2,
                    lRMmardersigma2))
 lRMmarvar<-lRMmarder%*%Ibcinv%*%t(lRMmarder)
 lRMmartest<-lRMmar/sqrt(lRMmarvar)
return(list(M0=M0mn, M1=M1mn, DMbar=DMbar, DMbarvar=DMbarvar,
             DMtest=DMtest, lRMss=lRMss, lRMssvar=lRMssvar,
             lRMsstest=lRMsstest, lRMmar=lRMmar,
             lRMmarvar=lRMmarvar,lRMmartest=lRMmartest))
}
```

B.5 AIC Calculation for Zero Inflated Gamma and Zero Inflated Log-Normal

```
ZIG.AIC<-function(y,mod)
{y.c<-y[y!=0]
 n.c<-length(y.c)
 vhat<-(1/mod$disp.mle)
 zig.full_llc<-vhat*mod$llc+(vhat-1)*sum(log(y.c))+
                  n.c*vhat*log(vhat)-n.c*log(gamma(vhat))
 np<-length(mod$b.b)+length(mod$b.c)
 ZGAIC<-2*np-2*(zig.full_llc+mod$llb)
 return(ZGAIC)
}
ZILN.AIC<-function(y,mod)
{y.c<-y[y!=0]
 n.c<-length(y.c)
 s2<-mod$disp
 ziln.full_llc<--(n.c/2)*log(2*pi)-(n.c/2)*log(s2)-
              (1/(2*s2))*mod$llc-sum(log(y.c))
 np<-length(mod$b.b)+length(mod$b.c)
 ZLNAIC<-2*np-2*(ziln.full_llc+mod$llb)
return(ZLNAIC)
}
```

REFERENCES

[1] Tu, W. and Zhou, X.-H. A Wald test comparing medical costs based on log-normal distributions with zero valued costs. *Statistics in Medicine*, **1999**. 18(20), 2749–2761.

[2] Zhou, X.-H. and Tu, W. Comparison of several independent population means when their samples contain log-normal and possibly zero observations. *Biometrics*, **1999**. 55(2), pp. 645–651.

[3] Duan, N.; Manning, W.; Morris, C.; and Newhouse, J. A comparison of alternative models for the demand for medical care. *Journal of Business & Economic Statistics*, **1983**. 1(2), 115–126.

[4] Lee, A. H.; Zhao, Y.; Yau, K. K.; and Xiang, L. How to analyze longitudinal multilevel physical activity data with many zeros? *Preventive Medicine*, **2010**. 51(6), 476–481.

[5] Feueverger, A. Some methods of analysis for weather experiments. *Biometrika*, **1979**. 66(3), 655–658.

[6] Hallstrom, A. P. A modified Wilcoxon test for non-negative distributions with a clump of zeros. *Statistics In Medicine*, **2010**. 29(3), 391–400.

[7] Chai, H. S. and Bailey, K. R. Use of log-skew-normal distribution in analysis of continuous data with a discrete component at zero. *Statist. Med.*, **2008**. 27(18), 3643–3655.

[8] Taylor, S. and Pollard, K. Hypothesis tests for point-mass mixture data with application to 'omics data with many zero values. *Statistical Applications In Genetics and Molecular Biology*, **2009**. 8(1), 8.

[9] Uc, E. Y.; Rizzo, M.; Dastrup, E.; Sparks, J. D.; Anderson, S. W.; and Dawson, J. D. Impaired curve negotiation in drivers with Parkinson's disease. *Turkish Journal of Neurology*, **2009**. 15 (1), 10–18.

[10] Berk, K. N. and Lachenbruch, P. A. Repeated measures with zeros. *Statistical Methods In Medical Research*, **2002**. 11(4), 303–316.

[11] Zhou, X. H. and Tu, W. Z. Interval estimation for the ratio in means of log-normally distributed medical costs with zero values. *Computational Statistics & Data Analysis*, **2000**. 35(2), 201–210.

[12] McLerran, D. http://listserv.uga.edu/cgi-bin/wa?A2=ind0805A&L=sas-l&P=R20779, **2008**.

[13] Tooze, J. A.; Grunwald, G. K.; and Jones, R. H. Analysis of repeated measures data with clumping at zero. *Statistical Methods In Medical Research*, **2002**. 11(4), 341–355.

[14] Lachenbruch, P. A. Comparisons of two-part models with competitors. *Statistics In Medicine*, **2001**. 20(8), 1215–1234.

[15] Moulton, L. H.; Curriero, F. C.; and Barroso, P. F. Mixture models for quantitative HIV RNA data. *Statistical Methods In Medical Research*, **2002**. 11(4), 317–325.

[16] Su, L.; Tom, B. D. M.; and Farewell, V. T. Bias in 2-part mixed models for longitudinal semicontinuous data. *Biostatistics*, **2009**. 10(2), 374–389.

[17] McCullagh, P. and Nelder, J. A. *Generalized Linear Models*. Chapman and Hall, **1989**.

[18] Hills, M. Likelihood. In *Encyclopedia of Biostatistics*. John Wiley & Sons, Ltd, **2005**.

[19] Land, C. E. Evaluation of approximate confidence interval estimation methods for log-normal means. *Technometrics*, **1972**. 14(1), 145–&.

[20] Dominici, F. and Zeger, S. L. Smooth quantile ratio estimation with regression: estimating medical expenditures for smoking-attributable diseases. *Biostatistics*, **2005**. 6(4), 505–519.

[21] Iman, R. L. and Conover, W. J. Use of the rank transform in regression. *Technometrics*, **1979**. 21(4), 499–509.

[22] Rizzo, M.; Shi, Q.; Dawson, J.; Anderson, S.; Kellison, I.; and Pietras, T. Stops for cops: Impaired response implementation for older drivers with cognitive decline. *Transportation Research Record: Journal of the Transportation Research Board*, **2005**. 1922(-1), 1–8.

[23] Dastrup, E.; Lees, M. N.; Bechara, A.; Dawson, J. D.; and Rizzo, M. Risky car following in abstinent users of MDMA. *Accident Analysis and Prevention*, **2010**. 42(3), 867–873.

[24] Uc, E. Y.; Rizzo, M.; Anderson, S. W.; Dastrup, E.; Sparks, J. D.; and Dawson, J. D. Driving under low-contrast visibility conditions in Parkinson disease. *Neurology*, **2009**. 73(14), 1103–1110.

[25] Uc, E. Y.; Rizzo, M.; Anderson, S. W.; Qian, S.; Rodnitzky, R. L.; and Dawson, J. D. Visual dysfunction in Parkinson disease without dementia. *Neurology*, **2005**. 65(12), 1907–1913.

[26] Fox, J. and Weisberg, S. *An R Companion to Applied Regression*. SAGE Publications, **2010**.

[27] Chen, Y. H. and Zhou, X. H. Interval estimates for the ratio and difference of two log-normal means. *Statistics In Medicine*, **2006**. 25(23), 4099–4113.

[28] Beran, R. Minimum hellinger distance estimates for parametric models. *Annals of Statistics*, **1977**. 5(3), 445–463.

[29] Basu, A. and Harris, I. R. Robust predictive-distributions for exponential-families. *Biometrika*, **1994**. 81(4), 790–794.

[30] Olsen, M. K. and Schafer, J. L. A two-part random-effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association*, **2001**. 96(454), 730–745.

[31] Cox, C. Delta method. In *Encyclopedia of Biostatistics*. John Wiley & Sons, Ltd, **2005**.

[32] Casella, G. and Berger, R. *Statistical inference*. Thomson Learning, **2002**.