2022

Power and Sample Size Calculations for Linear Mixed Models of Longitudinal Data Using the Kenward-Roger Adjusted Wald Test

Tarrant Oliver McPherson
University of Alabama at Birmingham

POWER AND SAMPLE SIZE CALCULATIONS FOR LINEAR MIXED MODELS OF LONGITUDINAL DATA USING THE KENWARD-ROGER ADJUSTED WALD TEST

by

TARRANT OLIVER MCPHERSON

LLOYD J. EDWARDS, COMMITTEE CHAIR
INMACULADA ABAN
TIMOTHY M. BEASLEY
IKJAE LEE
NENGJUN YI

A DISSERTATION

Submitted to the graduate faculty of The University of Alabama at Birmingham,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

BIRMINGHAM, ALABAMA

2022

POWER AND SAMPLE SIZE CALCULATIONS FOR LINEAR MIXED MODELS OF LONGITUDINAL DATA USING THE KENWARD-ROGER ADJUSTED WALD TEST

TARRANT OLIVER MCPHERSON

BIOSTATISTICS

ABSTRACT

The linear mixed model has become a popular technique for the analysis of longitudinal data, but Wald test statistics of fixed effects for these models frequently lack well defined distributions. A common approach to this problem uses the Kenward-Roger adjustment, which attempts to approximate the distribution of the Wald statistic by matching its moments obtained via Taylor expansion to those of an F distribution. However, this approach only matches moments obtained under the null hypothesis of no effect and cannot currently be used to approximate the distribution of the test statistic under some alternative hypothesis. This limitation prevents a straightforward approach to calculating power for the Kenward-Roger adjusted Wald statistic. In chapter 2, we introduce a novel power calculation that extends the original methodology of Kenward and Roger to obtain an approximate noncentral distribution of this adjusted Wald statistic from which power for tests of linear trend can then be calculated. This method is then extended to calculate expected power for designs with anticipated rates of missing follow-up data in Chapter 3, and finally to the calculation of sample size for such designs in Chapter 4. A variety of other techniques are also examined and compared to this method, with the newly developed method consistently outperforming other approaches in the calculation of both power and sample size.

Key Words: Power, Sample Size, Mixed Models, Kenward-Roger, Longitudinal Data

ACKNOWLEDGMENTS

This work could not have been achieved without the support of numerous individuals whose collective advise and effort provided me the means and encouragement to complete this project.

My Advisor, Dr. Lloyd Edwards, was instrumental in guiding my research. His knowledge of linear mixed models and the analysis of longitudinal data proved invaluable in solidifying my own understanding of these subjects and challenging me to pursue new and creative directions. His patience allowed me to explore many avenues and learn from dead ends, while his guidance ensured I never became lost.

The members of my Committee – Drs. Aban, Beasley, Lee, and Yi – all provided great insights and criticisms of my work in this dissertation. Additionally, they have been a great source of learning and encouragement by challenging me in the classroom and inviting me to collaborate on projects that furthered my professional growth.

Dr. Gary Cutter introduced me to the field of biostatistics. His continued encouragement, leadership, and desire to help his students and colleagues succeed motivated me to continue moving forward and inspired self-confidence needed to complete this process.

My fellow students were tremendous in their encouragement and positivity. In particular, Anastasia Hartzes and Vincent Laufer motivated me to work hard and furthered my interest and curiosity in the field of statistics.

Lastly, my parents were an unending source of love, support, and encouragement. Without the foundation they provided and their continued interest in my success, this work would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
COMMON ANALYSIS AND POWER METHODS FOR LONGITUDINAL DATA

I. Introduction

Designing medical experiments involves collaboration across a wide array of disciplines. Common questions encountered during the design phase concern both the ethics and feasibility of an experiment. What should a treatment be compared to? How will the success of a treatment be measured? What logistical or recruitment hurdles might a study face? Appropriate answers to these questions often require input from multiple perspectives, and a key piece of information required for such answers is the number of subjects an experiment will require.

Determining power and sample size of a study is critical for several reasons. In prospective studies, an accurate sample size determination informs researchers of recruitment feasibility and prevents overspending. The risks to human and animal subjects are likewise minimized by exposing a minimal number of subjects to treatments with unknown side effects. In retrospective studies, knowing power for a given sample size helps inform researchers if a set of data can potentially answer a research question and is therefore worth spending effort to analyze. Similarly, knowledge of how power changes with sample size may help researchers determine whether a complicated imputation is worth performing in the presence of missing data.

Power can be thought of as the probability that the null statistical hypothesis is rejected assuming the test statistic is distributed according to some function of the minimal clinically significant effect size. Therefore, a proper power or sample size calculation should match the type of analysis planned for the experiment. For many common types of statistical analyses, the power or sample size calculation is straightforward, as the distribution of the test statistic given the effect size is well known. For instance, to calculate the power for a one-sample t-test with null hypothesis $\mu = 0$ and alternative hypothesis $\mu \neq 0$, all that is required is the minimal effect size μ, the variance of the effect size σ^2, and the number of subjects in an experiment n. It is then well known that the test statistic under this parameterization follows a $t_{n-1,\omega}$ distribution with $n - 1$ degrees of freedom and noncentrality parameter $\omega = \frac{\mu}{\sigma/\sqrt{n}}$, and the probability a test statistic with such a distribution falls in the rejection region of a $t_{n-1,0}$ distribution under the null hypothesis can be easily calculated. Additional modifications to the power calculation may likewise be easily implemented. For instance, an anticipated dropout rate *r* may be accommodated by dividing *n* by 1-*r*. However, power calculations for longitudinal designs are often not so straightforward.

Longitudinal data analysis examines data collected on the same unit (subject for our discourse) at multiple time points. These sequential observations are often correlated within subjects, and so approaches requiring independence among all observations may be invalid for the analysis of longitudinal data. Instead, longitudinal data is often more appropriately analyzed using techniques that specifically accommodate related observations taken on the same subject. The conceptual framework for such analytical methods extends back to the foundations of statistics itself, where the idea that

observational errors could be probabilistically described was already being explored by the likes of Bernoulli, Laplace, and Gauss. Small errors were thought to be more probable than large errors, and errors within subjects were seen to be more related than errors between subjects. Early attempts to accommodate such errors typically relied on naïve Bayes type estimation procedures, although by the early part of the 20th century likelihood-based methods had become more common, perhaps in part due to the establishment of sufficiency of likelihood estimators (Stigler, 2007).

Among the first modern attempts at analyzing correlated data was a "mixed" effect ANOVA with random relatives effect (to account for correlation among relatives) performed by Fisher in a study on Mendelian inheritance (Fisher, 1919). Under the assumption of compound symmetric within-subjects covariance provided by assuming random subject effect, this approach easily generalized the paired t-test to multiple observations and was quickly applied to the analysis of longitudinal data. However, the mandate of compound symmetry is often overly restrictive for longitudinal data, which led to other methods being explored, as well as certain generalizations relaxing this assumption (for example the Greenhouse Geisser correction (Greenhouse & Geisser 1959)). Most notably, MANOVA approaches were developed as the theory related to multivariate normal distributions expanded, with the advantage of these approaches being a lack of assumption about covariance structure, but with the disadvantage being that within-subjects data must be non-missing and observed at equal time points for all subjects (Fitzmaurice and Molenberghs, 2008).

In the following sections, we will explore a few of these models and their generalizations in rigorous detail. Such an understanding is necessary to appreciate the

utility of the linear mixed model for longitudinal data and importance of the Kenward-Roger test statistic in performing inference for these models. Specifically, the Kenward-Roger test statistic was designed to match the Wald-type statistics for many of these models when exact null distributions for such statistics are known and restricted maximum likelihood estimation is used. Consequently, an understanding of the behavior of Wald-type test statistics in these models under the noncentral case will be necessary in guiding the development of power and sample size calculations for the Kenward-Roger test statistic.

II. The General Linear Multivariate Model

A. Definition and Model Structure

We will first examine the General Linear Multivariate Model (GLMM). For this model, we will largely be adopting the notation presented in Mardia, Kent, and Bibby (MKB, 1979) except for using $\boldsymbol{\varepsilon}$ instead of $\boldsymbol{U}$ to represent the error term. We express this model as

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

with $\boldsymbol{Y}$ an $(n\ x\ \tau)$ matrix, $\boldsymbol{X}$ an $(n\ x\ q)$ matrix, $\boldsymbol{\beta}$ a $(q\ x\ \tau)$ matrix, and $\boldsymbol{\varepsilon}$ an $(n\ x\ \tau)$ matrix. The GLMM can be presented as:

$$\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1\tau} \\ y_{21} & y_{22} & \cdots & y_{2\tau} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{n\tau} \end{bmatrix}$$

$$= \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1q} \\ x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nq} \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1\tau} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2\tau} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{q1} & \beta_{q2} & \cdots & \beta_{q\tau} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1\tau} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2\tau} \\ \vdots & \vdots & \vdots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{n\tau} \end{bmatrix}$$

or more succinctly as

$$\begin{bmatrix} \boldsymbol{y}_1' \\ \boldsymbol{y}_2' \\ \vdots \\ \boldsymbol{y}_n' \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1' \\ \boldsymbol{\beta}_2' \\ \vdots \\ \boldsymbol{\beta}_q' \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1' \\ \boldsymbol{\varepsilon}_2' \\ \vdots \\ \boldsymbol{\varepsilon}_n' \end{bmatrix}$$

with n representing the number of subjects, τ representing the number of observations per subject, and q representing the number of independent variables in the model (including intercept). Additionally, $\boldsymbol{y}_i' \sim N_\tau(\boldsymbol{x}_i'\boldsymbol{\beta}, \boldsymbol{\Sigma})$ and $\boldsymbol{\varepsilon}_i' \sim N_\tau(\boldsymbol{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ being a τ *x* τ covariance matrix and $\boldsymbol{y}_i'$ are uncorrelated.

B. Likelihood Function and Estimation

The probability density function (PDF) for $\boldsymbol{y}_i$ is given as:

$$f_{\boldsymbol{Y}_i}(\boldsymbol{y}_i) = (2\pi)^{-\tau/2}\,|\boldsymbol{\Sigma}|^{-1/2} exp\left\{-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\beta}'\boldsymbol{x}_i)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{\beta}'\boldsymbol{x}_i)\right\}$$

And so for the joint PDF we have:

$$f_{\boldsymbol{Y}_{1,\ldots,n}}(\boldsymbol{y}_{1,\ldots,n}) = (2\pi)^{-n\tau/2}\,|\boldsymbol{\Sigma}|^{-n/2} exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\beta}'\boldsymbol{x}_i)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{\beta}'\boldsymbol{x}_i)\right\}$$

Thus, the PDF for $\boldsymbol{Y}$ is**:**

$$f_{\boldsymbol{Y}}(\boldsymbol{Y}) = (2\pi)^{-n\tau/2}\,|\boldsymbol{\Sigma}|^{-n/2} exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i' - \boldsymbol{x}_i'\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i' - \boldsymbol{x}_i'\boldsymbol{\beta})'\right\}$$

Noting that

$$\sum_{i=1}^{n}(\boldsymbol{y}_i' - \boldsymbol{x}_i'\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i' - \boldsymbol{x}_i'\boldsymbol{\beta})' = tr[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})']$$

then

$$f_{\boldsymbol{Y}}(\boldsymbol{Y}) = (2\pi)^{-n\tau/2}\,|\boldsymbol{\Sigma}|^{-n/2} exp\{tr[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})']\}.$$

Therefore, the log-likelihood function for $\boldsymbol{Y}$ is:

$$l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) = -\frac{n\tau}{2} log(2\pi) - \frac{n}{2} log|\boldsymbol{\Sigma}| - \frac{1}{2} tr[(\boldsymbol{Y} - \boldsymbol{X\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta})'].$$

The maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$ will be independent of $\boldsymbol{\Sigma}$ while the reverse is not true. Specifically, from proof 1 in Appendix I we have the MLEs for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ as:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-\mathbf{1}}\boldsymbol{X'Y}$$

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\big(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\big)'\big(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\big).$$

C. Hypothesis Testing

The general linear hypothesis is given by H: $\boldsymbol{C'\beta U}$**= Θ.** Several different test statistics are frequently utilized for testing hypotheses in the GLMM, none of which are uniformly most powerful in all cases, although all reduce to the same value in many scenarios encountered in the analysis of longitudinal data. A comparison of several of these statistics is presented in Ateş (2019), with relative performance depending on covariance homogeneity, observational balance, and adherence to normality assumptions. Among the most popular of these statistics is the Hotelling-Lawley Trace (HLT) statistic first examined in Hotelling (1931), later generalized to cases with more than two groups Lawley(1938), and given its more-or-less present formalization in Hotelling (1951). We will restrict our focus to this test statistic, as it has an approachable form and is explicitly tracked by the Kenward-Roger statistic. The form of the HLT test statistic is given as:

$$F = \lambda * tr\left\{\left[\boldsymbol{U'}\widehat{\boldsymbol{\beta}}'\boldsymbol{C}(\boldsymbol{C'}(\boldsymbol{X'X})^{-\mathbf{1}}\boldsymbol{C})^{-\mathbf{1}}\boldsymbol{C'}\widehat{\boldsymbol{\beta}}\boldsymbol{U}\right]\left[n\boldsymbol{U'}\widehat{\boldsymbol{\Sigma}}\boldsymbol{U}\right]^{-1}\right\}$$

Where $\boldsymbol{C}$ is a $(q \times a)$ between subjects contrast matrix and $\boldsymbol{U}$ is a $(\tau \times b)$ within subjects contrast matrix. To obtain the distribution of this test statistic, we will first derive the distributions of the quantities $\boldsymbol{H} = \boldsymbol{U'}\widehat{\boldsymbol{\beta}}'\boldsymbol{C}(\boldsymbol{C'}(\boldsymbol{X'X})^{-\mathbf{1}}\boldsymbol{C})^{-\mathbf{1}}\boldsymbol{C'}\widehat{\boldsymbol{\beta}}\boldsymbol{U}$ and $\boldsymbol{E} = \left[n\boldsymbol{U'}\widehat{\boldsymbol{\Sigma}}\boldsymbol{U}\right]$.

(i) Distribution of H : From corollary 3 of Singull and Koski (2012) we have that if $\boldsymbol{Y}\sim N_{\tau}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ and $\boldsymbol{A}$ is a symmetric idempotent matrix of rank *r*, then $\boldsymbol{Y'AY}\sim W_{\tau}(r,\boldsymbol{\Sigma},\boldsymbol{\mu'A\mu})$, a τ *x* τ Wishart distributed random matrix with *r* degrees of freedom and non-centrality parameter $\boldsymbol{\mu'A\mu}$.

Now, $\boldsymbol{H} = \boldsymbol{U'\widehat{\beta}'C(C'(X'X)^{-1}C)^{-1}C'\widehat{\beta}U}$

$$= \boldsymbol{U'Y'X(X'X)^{-1}C(C'(X'X)^{-1}C)^{-1}C'(X'X)^{-1}X'YU}$$

Note

$$\boldsymbol{[X(X'X)^{-1}C(C'(X'X)^{-1}C)^{-1}C'(X'X)^{-1}X'][X(X'X)^{-1}C(C'(X'X)^{-1}C)^{-1}C'(X'X)^{-1}X']}$$

$$= \boldsymbol{X(X'X)^{-1}C(C'(X'X)^{-1}C)^{-1}C'(X'X)^{-1}X'}$$

And so $\boldsymbol{X(X'X)^{-1}C(C'(X'X)^{-1}C)^{-1}C'(X'X)^{-1}X'}$ is idempotent (proof of symmetry is trivially obtained by taking the transpose). Additionally, since it is idempotent, its rank is equal to its trace, which is

$$\boldsymbol{tr[X(X'X)^{-1}C(C'(X'X)^{-1}C)^{-1}C'(X'X)^{-1}X']}$$

$$= \boldsymbol{tr[C'(X'X)^{-1}X'X(X'X)^{-1}C(C'(X'X)^{-1}C)^{-1}]}$$

$$= \boldsymbol{tr[C'(X'X)^{-1}C(C'(X'X)^{-1}C)^{-1}]}$$

$$= \boldsymbol{tr[I}_a\boldsymbol{]}$$

$$= a$$

So, $\boldsymbol{YU}\sim N_b(\boldsymbol{X\beta U},\boldsymbol{U'\Sigma U})$ and therefore

$\boldsymbol{H}\sim W_b(a,\boldsymbol{U'\Sigma U},\boldsymbol{U'\beta'X'X(X'X)^{-1}C(C'(X'X)^{-1}C)^{-1}C'(X'X)^{-1}X'X\beta U})$ reducing to

$$\boldsymbol{H}\sim W_b(a,\boldsymbol{U'\Sigma U},\boldsymbol{U'\beta'C(C'(X'X)^{-1}C)^{-1}C'\beta U})$$

(ii) Distribution of E: First, let $\boldsymbol{Q} = \boldsymbol{I}_{nxn} - \boldsymbol{X(X'X)^{-1}X'}$. Then $\boldsymbol{Q}$ is symmetric and

$$\boldsymbol{Q'Q} = \boldsymbol{I} - \boldsymbol{X(X'X)^{-1}X'} - \boldsymbol{X(X'X)^{-1}X'} + \boldsymbol{X(X'X)^{-1}X'X(X'X)^{-1}X'}$$

$$= \boldsymbol{I} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$$

$$= \boldsymbol{Q}$$

Now, $\boldsymbol{E} = \boldsymbol{U}'\big(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\big)'\big(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\big)\boldsymbol{U} = \boldsymbol{U}'\boldsymbol{Y}'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{Y}\boldsymbol{U} = \boldsymbol{U}'\boldsymbol{Y}'\boldsymbol{Q}\boldsymbol{Y}\boldsymbol{U}$ with $\boldsymbol{Q}$ symmetric and idempotent of rank $= tr(\boldsymbol{I}_{nxn}) - tr(\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}) = tr(\boldsymbol{I}_{NxN}) - tr\big(\boldsymbol{I}_{qxq}\big) = N - q$, and so:

$$\boldsymbol{E} \sim W_b(N - q, \boldsymbol{U}'\boldsymbol{\Sigma}\boldsymbol{U}, \boldsymbol{U}'\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{Q}\boldsymbol{X}\boldsymbol{\beta}\boldsymbol{U})$$

$$= W_b(n - q, \boldsymbol{U}'\boldsymbol{\Sigma}\boldsymbol{U}, \boldsymbol{U}'\boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{X}\boldsymbol{\beta}\boldsymbol{U})$$

$$= W_b(n - q, \boldsymbol{U}'\boldsymbol{\Sigma}\boldsymbol{U}, \boldsymbol{0})$$

(iii) Independence of H and E : From MKB theorem 3.4.6 (Craig's theorem), we have that if $\boldsymbol{Y} \sim N_\tau(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are symmetric idempotent matrices, then $\boldsymbol{Y}'\boldsymbol{A}_1\boldsymbol{Y}$ and $\boldsymbol{Y}'\boldsymbol{A}_2\boldsymbol{Y}$ are jointly independent if $\boldsymbol{A}_1\boldsymbol{A}_2 = \boldsymbol{0}$. Letting $\boldsymbol{A}_1 = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}(\boldsymbol{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and $\boldsymbol{A}_2 = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ we have

$$\boldsymbol{A}_1\boldsymbol{A}_2 = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}(\boldsymbol{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')$$

$$= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}(\boldsymbol{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' -$$

$$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}(\boldsymbol{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$$

$$= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}(\boldsymbol{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' -$$

$$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}(\boldsymbol{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$$

$$= \boldsymbol{0}$$

So, since $\boldsymbol{H} = \boldsymbol{U}'\boldsymbol{Y}'\boldsymbol{A}_1\boldsymbol{Y}\boldsymbol{U}$ and $\boldsymbol{E} = \boldsymbol{U}'\boldsymbol{Y}'\boldsymbol{A}_2\boldsymbol{Y}\boldsymbol{U}$, they are independent.

(iv) Distribution of $tr[HE^{-1}]$: When $b = 1$ (i.e. when the within-subjects contrast matrix $\boldsymbol{U}$ is a single column) then:

$$tr[\boldsymbol{HE}^{-1}] = \boldsymbol{HE}^{-1} = \frac{\chi_a^2(\boldsymbol{U'\beta'C(C'(X'X)^{-1}C)^{-1}C'\beta U})}{\chi_{n-q}^2}$$

Therefore, in this case,

$$tr[\boldsymbol{HE}^{-1}] \sim \frac{a}{n-q} F_{a,\ n-q,\ \omega}$$

with $\omega = \boldsymbol{U'\beta'C(C'(X'X)^{-1}C)^{-1}C'\beta U}$

When $a = 1$ (i.e. when the between-subjects contrast matrix $\boldsymbol{C}$ is a single column), then $\boldsymbol{HE}^{-1}$ is of rank 1 and its trace is equal to its only nonzero eigenvalue. Thus, from extension of MKB definition 3.7.2 and equation 3.7.15 describing the distribution of the largest eigenvalue of the product of two such independent Wishart matrices (note the trace of $\boldsymbol{HE}^{-1}$ and the trace of $\boldsymbol{E}^{-1}\boldsymbol{H}$ are equal) we have, under the null hypothesis with non-centrality parameter $\omega = 0$:

$$tr[\boldsymbol{HE}^{-1}] \sim \frac{b}{n-q-b+1} F_{b,\ n-q-b+1}$$

Under the alternative hypothesis the noncentrality parameter is not always well defined. Obviously, $\boldsymbol{U'\beta'C(C'(X'X)^{-1}C)^{-1}C'\beta U}$ cannot be used as the noncentrality parameter because it is a matrix and does not directly conform to the scalar value demanded by a noncentral F distribution. Approximations to the distribution under the alternative hypothesis are still a subject of research (e.g. Johnstone and Nadler, 2017). However, in certain situations such as the test of a vector of means at each of τ time points for one or between two groups, the noncentrality parameter can be given as that of the noncentral Hotelling T^2 distribution. The results are important: in the one sample T^2 the noncentrality parameter is $\omega = \boldsymbol{\mu'\Sigma^{-1}\mu}$ whereas in a test of the means between two groups with sizes n_1 and n_2 the noncentrality parameter is given as $\omega = \frac{n_1 n_2}{n_1+n_2}(\boldsymbol{\mu_1 - \mu_2})'\boldsymbol{\Sigma^{-1}}(\boldsymbol{\mu_1 - \mu_2})$

(Gupta, 2006) suggesting ω is obtained by scaling the effect size much like λ scales the test statistic to form an F distribution.

(v) Distribution of the Wald test statistic: We have seen that in two exact cases, the test statistic follows an F distribution. Specifically, when $b = 1$:

$$\boldsymbol{F} = \lambda * tr[\boldsymbol{H}\boldsymbol{E}^{-1}] = \frac{n-q}{a} tr[\boldsymbol{H}\boldsymbol{E}^{-1}] \sim F_{a,\ n-q,\ \omega}$$

$$\omega = \boldsymbol{U}'\boldsymbol{\beta}'\boldsymbol{C}(\boldsymbol{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{\beta}\boldsymbol{U}$$

and when $a = 1$ under the null hypothesis:

$$\boldsymbol{F} = \lambda * tr[\boldsymbol{H}\boldsymbol{E}^{-1}] = \frac{n-q-b+1}{b} tr[\boldsymbol{H}\boldsymbol{E}^{-1}] \sim F_{b,\ n-q-b+1}$$

However, when these two cases fail, the distribution of the test statistic has no known distribution. Instead, it is common to assume the distribution follows an F statistic and to approximate the degrees of freedom of the distribution and possibly scale factor. Fortunately, as will be demonstrated in the following example, the hypothesis tests of individual fixed effects of interest in balanced linear mixed models (for this document, "balanced" means no missing data and all subjects are observed at the same time points) analyzing longitudinal data can frequently be expressed as a multivariate linear model with a Wald test statistic distributed identically to both special cases previously discussed, i.e. both a=1 and b=1.

D. Example, Group by Time Effect

Suppose we are analyzing the results of a study examining protein supplementation on lean body mass following bariatric surgery. Patients were randomized to either a routine diet of 1500 Kcal per day following surgery (group 1), or a

diet of 1500 Kcal per day with 400 Kcal coming from supplemental protein powder (group 2). Post-surgery, patients have biometric data including lean body mass measured once every 3 months for a year providing 5 observations per subject (months 0, 3, 6, 9, 12). Of primary interest is whether the trajectory of lean body mass index over time is different between the two groups (equivalent to the group by time effect in the mixed model). There were 10 subjects in each study group with no missing data and no mistimed observations.

The between and within-subjects contrast matrices to test if there is a difference between groups in trajectory of lean body mass over time are:

$$\boldsymbol{C} = [0 \quad 1]'$$

$$\boldsymbol{U} = [-2 \quad -1 \quad 0 \quad 1 \quad 2\,]'$$

The design matrix for subject *i* will be:

$$\boldsymbol{x}_i' = [1 \;\; g_i\,]$$

with $g_i = 0$ if subject *i* in group 1 and $g_i = 0$ if subject *i* in group 2. The hypothesis that no difference exists will then be rejected if $\frac{n-q}{a} tr[\boldsymbol{HE}^{-1}] = 18 * tr[\boldsymbol{HE}^{-1}]$ falls in the rejection region of an $F_{1,18}$ distribution.

E. Strengths and Limitations

Compared to other methods of analyzing longitudinal data, the GLMM has two attractive qualities. First, while $\widehat{\boldsymbol{\Sigma}}$ is a biased estimator of $\boldsymbol{\Sigma}$, this bias has no impact on the test statistic. Secondly, the test statistics have known distributions for the special cases described, which happen to correspond with testing individual effects of interest for most researchers.

However, the GLMM has several major limitations, the two most obvious of which are that all data must be observed (non-missing) and that all subjects must be observed at the same timepoints. Such demands are clear from the structure presented in Section II.A, as $\boldsymbol{y}_i'$ must have τ elements and $\boldsymbol{\beta}$ takes no consideration of the time of observation (only order of observation is considered) in mapping from $\boldsymbol{X}$ to $\boldsymbol{Y}$. Lastly, the structure of the model is unfamiliar to many researchers who are often unaccustomed to the idea of an outcome as a joint set of observations. As a result of these limitations, other more flexible methods for analyzing longitudinal data have become popular alternatives.

III. The General Linear Model (GLM)

The next type of model commonly used for the analysis of longitudinal data is the General Linear Model (GLM). This model offers increased flexibility over the GLMM and is capable of explicitly estimating the effect of time as a continuous function on which the outcome depends. However, this model will be shown to have its own limitations, particularly with regards to estimation and hypothesis testing.

A. Definition and Model Structure

The GLM is specified as follows:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

In this instance $i = 1,\ldots,n$ refers to one of the subjects in the model. $\boldsymbol{Y}_i$ is the m_i x 1 matrix consisting of the outcome measurements for subject i at each of the $j = 1,\ldots,m_i$ times the subject was measured. Additionally, $\boldsymbol{Y}_i \sim N_{n_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$ and $\boldsymbol{Y}_i$ are independent. $\boldsymbol{X}_i$ is the m_i x r fixed effects design matrix, one of whose columns is the vector of

timepoints at which subject *i* was measured (*r* has been used instead of *q* to differentiate from the GLMM). $\boldsymbol{\beta}$ is the *r* x 1 vector of fixed effect parameters. $\boldsymbol{\varepsilon}_i$ is the m_i x 1 matrix giving the "error" or amounts by which the outcome for patient *i* at observation time *j* deviates from its expected value. Thus $\boldsymbol{\varepsilon}_i \sim N_{m_i}(0, \boldsymbol{\Sigma_i})$. So, structurally the model appears as:

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{im_i} \end{bmatrix} = \begin{bmatrix} x_{i11} & x_{i21} & \dots & x_{ir1} \\ x_{i12} & x_{i22} & \dots & x_{ir2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{i1m_i} & x_{i2m_i} & \dots & x_{irm_i} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{im_i} \end{bmatrix}$$

Stacking these observations as $\boldsymbol{Y} = (\boldsymbol{Y}_1\, \boldsymbol{Y}_2 \dots \boldsymbol{Y}_n)'$ we can express the model in shorthand form:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

Typically, the first column of the design matrix will be a column of ones representing the intercept term, but this is not a requirement.

B. Likelihood function and Estimation

The pdf of $\boldsymbol{Y}_i$ is given as:

$$f_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i) = (2\pi)^{-m_i/2}\, |\boldsymbol{\Sigma_i}|^{-1/2} exp\left\{-\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right\}$$

In the simplest case, assume that $m_i = \tau$ for all subjects. Then $\boldsymbol{\Sigma} = \boldsymbol{\Sigma_Y} = \boldsymbol{I}_{nxn} \otimes \boldsymbol{\Sigma_i}$ and the pdf of $\boldsymbol{Y}$ is given as:

$$f_{\boldsymbol{Y}_{1,\dots,n}}(\boldsymbol{Y}_{1,..,n}) = f_{\boldsymbol{Y}}(\boldsymbol{Y}) = (2\pi)^{-n\tau/2}\, |\boldsymbol{\Sigma_i}|^{-n/2} exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right\}$$

$$= (2\pi)^{-n\tau/2}\, |\boldsymbol{\Sigma}|^{-1/2} exp\left\{-\tfrac{1}{2}(\boldsymbol{Y} - \boldsymbol{X\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta})\right\}$$

Therefore, the log-likelihood function for $\boldsymbol{Y}$ is:

$$l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) = -\frac{n\tau}{2} log(2\pi) - \frac{1}{2} log|\boldsymbol{\Sigma}| - -\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta})$$

From proof 2 in Appendix I, the MLEs for this balanced case are

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} \text{ and } \widehat{\boldsymbol{\Sigma}}_i = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{X\beta})(\boldsymbol{Y} - \boldsymbol{X\beta})'$$

when assuming $\boldsymbol{\beta}$ or $\boldsymbol{\Sigma_i}$ is fixed and known, respectively. Unfortunately, these quantities both depend on each other, and so the following expressions are often given instead as the "feasible" estimators obtained from the conditional likelihood function for each parameter:

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Y} \text{ and } \widehat{\boldsymbol{\Sigma}}_i = \frac{1}{n}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'$$

For cases when the number of observations differs among subjects, the estimates may not be so easily expressed. However, even if these estimates are based on balanced data, several problems exist that complicate analyses based on this model.

C. Issues with Likelihood Estimates

The problems with the estimates in the previous section II.B are three-fold.

(i)Bias of ML Estimators Typically maximum likelihood estimates are biased. A simple example readily shows the bias in the ML estimate of estimate of $\boldsymbol{\Sigma}$. Assume all subjects are observed τ times and all observations within and between subjects are independent such that $\boldsymbol{\Sigma}_i = \sigma^2 \boldsymbol{I}_{\tau\, x\, \tau}$ and $\boldsymbol{\Sigma_Y} = \sigma^2 \boldsymbol{I}_{n\tau\, x\, n\tau}$. Then setting $\frac{\partial}{\partial \boldsymbol{\Sigma}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) = 0$ and solving, conditional on $\boldsymbol{\beta}$, we have:

$$\frac{n}{2} tr[\boldsymbol{I}_{\tau x \tau}] = \frac{1}{2} tr\left[\sum_{i=1}^{n} \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\right]$$

$$n\tau = \sigma^{-2} tr\left[\sum_{i=1}^{n} \boldsymbol{I}_{\tau x \tau}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\right]$$

$$n\tau = \sigma^{-2} tr\left[\sum_{i=1}^{n} (\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{I}_{\tau x \tau}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right]$$

$$n\tau = \sigma^{-2}(\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})$$

$$\sigma^2 = \frac{1}{n\tau}(\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})$$

Additionally, conditional on $\boldsymbol{\Sigma}$

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} \\ &= \sigma^2\sigma^{-2}(\boldsymbol{X}'\mathbf{I}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbf{I}^{-1}\boldsymbol{Y} \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\end{aligned}$$

Therefore,

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n\tau}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) \\ &= \frac{1}{n\tau}(\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y})'(\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y})\end{aligned}$$

Proceeding as in A.III we have $n\tau\hat{\sigma}^2 = \boldsymbol{Y}'\boldsymbol{Q}\boldsymbol{Y} \sim W_1(n\tau - r, \sigma^2\boldsymbol{I}_{n\tau\, x\, n\tau}, \mathbf{0})$, since

$$\boldsymbol{Y}_i \sim N_1(\boldsymbol{X}_i\boldsymbol{\beta}, \sigma^2)$$

From MKB corollary 3.4.2.1 and theorem 3.4.4 b we have:

$$\boldsymbol{Y}'\boldsymbol{Q}\boldsymbol{Y} \sim \sum_{i=1}^{n\tau - r} \sigma^2\chi_1^2$$

And so

$$E(\hat{\sigma}^2) = \frac{n\tau - r}{n\tau}\sigma^2 < \sigma^2$$

Note that this bias is typically not an issue in exact cases where the degrees of freedom in the F-statistic correct for this bias in the variance estimator, but in non-exact cases this bias can pose more of an issue.

(ii) Interdependence of Estimators The estimates often depend on each other, thus preventing a general closed form solution.

(iii)Hierarchical distribution of $\widehat{\boldsymbol{\beta}}$ In the case that $\widehat{\boldsymbol{\Sigma}}_i$ could be estimated independently of $\boldsymbol{\beta}$ so that $\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Y}$ (as we'll see is the case in REML estimation), there would still be the issue that the distribution of $\widehat{\boldsymbol{\beta}}$ would hierarchically depend on the distribution of $\widehat{\boldsymbol{\Sigma}}_i$ and would therefore be difficult to determine. Thus, test statistics based on this estimate can also have distributions that are difficult to express.

D. REML Estimation

To solve issues *(i)* and *(ii)* from previous section III.C we can instead estimate $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Sigma}}_i$ via Restricted Maximum Likelihood Estimation (REML). Essentially, we utilize the fact that the error term $\boldsymbol{\varepsilon}_i$ has the same covariance as $\boldsymbol{y}_i$ but whose mean does not depend on $\boldsymbol{\beta}$. So instead of estimating $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Sigma}}_i$ based on $\boldsymbol{Y}$ we will estimate based on the residuals of the model**.** Interdependencies are introduced among observed residuals in the estimation process (for instance, in OLS estimation we can determine the value of the n^{th} residual if we know the previous n - 1 residuals), and so we need to find a basis for these residuals that expresses them in terms of the original error space. To do so, we can define:

$\boldsymbol{H'} = \boldsymbol{X}(\boldsymbol{X'}\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X'}\boldsymbol{\Sigma}^{-1}$ (with $\boldsymbol{H}$ an $n\tau$ *x* $n\tau$ matrix)

$\boldsymbol{Q'} = \boldsymbol{I} - \boldsymbol{H'}$ (note $\boldsymbol{Q'X} = \boldsymbol{X} - \boldsymbol{X}(\boldsymbol{X'}\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X'}\boldsymbol{\Sigma}^{-1}\boldsymbol{X} = \boldsymbol{0}$)

$\boldsymbol{Q}$ is idempotent and its rank = trace = $n\tau$-r. Since $\boldsymbol{Q}$ is a square matrix it can be expressed via "QR" factorization as $\boldsymbol{Q} = \boldsymbol{MA}$ where $\boldsymbol{M}$ is rank $n\tau$-r and $\boldsymbol{A}$ is an upper triangular matrix. Additionally, the columns of $\boldsymbol{M}$ form an orthogonal basis and so

$$\boldsymbol{R_i} = \boldsymbol{M_i'Y}_i - \boldsymbol{M_i'X}_i\boldsymbol{B} = \boldsymbol{M_i'Y}_i \sim N_\tau(\boldsymbol{0}, \boldsymbol{M_i'\Sigma M}_i)$$

Note $\boldsymbol{M'\Sigma M}$ is of the same block diagonal shape as $\boldsymbol{\Sigma}$ but with the last r rows containing only zeros (since $\boldsymbol{M}$ is rank $n\tau$-r and has columns forming orthogonal basis). We can therefore express the likelihood function of $\boldsymbol{R}$ as:

$$l(\boldsymbol{\Sigma}|\boldsymbol{R}) = -\frac{np}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\log|\boldsymbol{M_i'\Sigma}_i\boldsymbol{M}_i| - \frac{1}{2}\sum_{i=1}^{n}\boldsymbol{R_i'}(\boldsymbol{M_i'\Sigma}_i\boldsymbol{M}_i)^{-1}\boldsymbol{R}_i$$

The above expression still has the problem that $\boldsymbol{M}$ depends on $\boldsymbol{Q}$ and therefore on the value of $\boldsymbol{\Sigma}$, which is an unknown quantity we are trying to estimate. However, this function can be equivalently expressed as (Gurka, 2006):

$l_{REML}(\boldsymbol{\Sigma}|\boldsymbol{Y}) =$

$$-\frac{n\tau - r}{2}\log(2\pi) + \frac{1}{2}\log|\boldsymbol{X'X}| - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\log|\boldsymbol{X'}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}| - \frac{1}{2}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)$$

Where $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X'}\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X'}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}$

From the above expression, it is now obvious that REML allows the estimation of $\boldsymbol{\Sigma}$ first and then of $\boldsymbol{\beta}$. Thus, the dependency issue in (ii) from III.C is avoided. To see how the bias issue in (i) is mitigated, we can calculate the REML estimate of $\boldsymbol{\Sigma}$ from part III.C.*i*. Here $\boldsymbol{Q}$ is idempotent and its rank = trace = $n\tau$-r and $\boldsymbol{\Sigma}_i = \sigma^2\boldsymbol{I}_{\tau\,x\,\tau}$

$$\frac{\partial}{\partial\boldsymbol{\Sigma}}l(\boldsymbol{\Sigma}|\boldsymbol{R}) = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\boldsymbol{\Sigma_i}}log|\boldsymbol{M}_i{}'\boldsymbol{\Sigma}_i\boldsymbol{M}_i|\right) - \frac{1}{2}\left(tr\left[\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\Sigma_i}}(\boldsymbol{M}_i{}'\boldsymbol{\Sigma}_i\boldsymbol{M}_i)^{-1}\boldsymbol{R_i}\boldsymbol{R_i}{}'\right]\right)$$

$$= -\frac{1}{2}\frac{\partial}{\boldsymbol{\Sigma}}(log|\boldsymbol{M}'\sigma^2\boldsymbol{I}_{n\tau\, x\, n\tau}\boldsymbol{M}|) - \frac{1}{2}\left(tr\left[\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\Sigma_i}}(\boldsymbol{M}'\sigma^2\boldsymbol{I}_{\tau\, x\, \tau}\boldsymbol{M})^{-1}\boldsymbol{R_i}\boldsymbol{R_i}'\right]\right)$$

Noting $\boldsymbol{M}_i$ is orthogonal and thus invertible, so $(\boldsymbol{M}'\boldsymbol{\Sigma}\boldsymbol{M})^{-1} = \boldsymbol{M}^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{M}')^{-1}$

$$= -\frac{1}{2}tr[\boldsymbol{M}'\boldsymbol{M}] + \frac{1}{2}tr[\sigma^{-2}\boldsymbol{I}_{n\tau\, x\, n\tau}(\boldsymbol{M}')^{-1}\boldsymbol{R}\boldsymbol{R}'\boldsymbol{M}^{-1}]$$

$$= -\frac{n\tau - r}{2} + \frac{1}{2}tr[\sigma^{-2}\boldsymbol{I}_{n\tau\, x\, n\tau}(\boldsymbol{M}')^{-1}\boldsymbol{R}\boldsymbol{R}'\boldsymbol{M}^{-1}]$$

Recalling $\boldsymbol{R} = \boldsymbol{M}'\boldsymbol{Y} - \boldsymbol{M}'\boldsymbol{X}\boldsymbol{B}$

$$n\tau - r = tr[\sigma^2(\boldsymbol{M}')^{-1}(\boldsymbol{M}'\boldsymbol{Y} - \boldsymbol{M}'\boldsymbol{X}\boldsymbol{B})(\boldsymbol{M}'\boldsymbol{Y} - \boldsymbol{M}'\boldsymbol{X}\boldsymbol{B})'\boldsymbol{M}^{-1}]$$

$$n\tau - r = tr[\sigma^2(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})(\boldsymbol{Y}' - \boldsymbol{X}'\boldsymbol{B}')]$$

$$n\tau - r = tr[\sigma^2(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})']$$

$$n\tau - r = tr[\sigma^2(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})]$$

$$n\tau - r = \sigma^{-2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})$$

$$\sigma^2 = \frac{1}{n\tau - r}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})$$

As before, $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ and so

$$\hat{\sigma}^2 = \frac{1}{n\tau - r}(\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y})'(\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y})$$

And therefore

$$E(\hat{\sigma}^2) = \frac{n\tau - r}{n\tau - r}\sigma^2 = \sigma^2$$

In this case, the REML variance estimate is unbiased and therefore solves the bias issue in (i) from III.C. In general, the bias of the REML estimate $\widehat{\boldsymbol{\Sigma}}$ is bounded by $O(n^{-2})$ whereas the ML estimate is bounded by $O(n^{-1})$ (Tang, 2017). Additionally, while in this simple case the issue described in (iii) from III.C is avoided, $\widehat{\boldsymbol{\beta}}$ will often remain a function of $\widehat{\boldsymbol{\Sigma}}$ even under REML estimation. Thus, the typical estimate of $\widehat{var}(\widehat{\boldsymbol{\beta}}) =$

$\left(\boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}\right)^{-1}$ will underestimate the true value of $var\left(\widehat{\boldsymbol{\beta}}\right)$ and therefore lead to inflated type-I error rates in Wald-type F tests.

E. Advantages and Disadvantages of the GLM

The general linear model with repeated measures offers several advantages over the GLMM. Most important are the ability to explicitly estimate the effect of time on an outcome as a continuous function, as well as the lack of demand that subjects have equal numbers of observations or be observed at equal time points.

However, the added flexibility of the general linear model with repeated measures tends to create problems with estimation and hypothesis testing, particularly when data are unbalanced. Additionally, the types of covariance structures accommodated by this model are typically those which are functions of the order of the observations rather than of the time at which each observation was taken. For instance, an AR(1) covariance structure specifies a constant decay of correlation regardless of when observations were taken. A linear exponent autoregressive (LEAR) covariance structure mitigates this problem to an extent by scaling the decay rate by the amount of time between observations, but this structure adjusts the decay rate equally irrespective of when observations occur and therefore may not well accommodate data with many subjects observed in widely varying time periods (see Simpson et al., 2010 for details on these covariance structures). Thus, while the GLM may accommodate unbalanced data, its ability to do so is often subject to the validity of assumptions demanded by the types of covariance structures modeled, which impacts power via the variance estimates.

IV. The Linear Mixed Model

The linear mixed model (LMM) has much of the same advantages and disadvantages of the GLM. However, for the purposes of longitudinal data analysis, the LMM is capable of modeling covariance structures that are continuous functions of time and therefore provide an added level of flexibility.

A. Definition and Model Structure

The Linear Mixed Model (LMM) is specified as follows:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$$

The structure of this model is quite similar to that of the GLM. Again, $\boldsymbol{Y}_i$ is the m_i x 1 matrix consisting of the outcome measurements for subject *i,* $i = 1,\ldots,n,$ at each of the $j = 1,\ldots,m_i$ times the subject was measured, and with all $\boldsymbol{Y}_i$ independent. $\boldsymbol{X}_i$ is the m_i x r fixed effects design matrix, one of whose columns is the vector of timepoints at which subject *i* was measured. $\boldsymbol{\beta}$ is the r x 1 vector of fixed effect parameters. $\boldsymbol{e}_i$ is the m_i x 1 matrix giving the "error" or amounts by which the outcome for patient *i* at observation time *j* deviates from its expected value. Thus $\boldsymbol{e}_i \sim N_{m_i}(0, \boldsymbol{\Sigma_{ei}})$. For the remainder of this document we will assume $\boldsymbol{\Sigma_{ei}} = \sigma^2\mathbf{I}_{m_i x m_i}$.

Additionally, in this model we have two new components: $\boldsymbol{Z}_i$ is the m_i x 1 (random intercept only) or m_i x 2 (random intercept and slope) random effects design matrix given as:

$$\boldsymbol{Z}_i = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad \text{or} \quad \boldsymbol{Z}_i = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_{m_i} \end{bmatrix}$$

$\boldsymbol{b}_i$ is the 1 x 1 or 2 x 1 vector of random effect parameters to be estimated given by $\boldsymbol{b}_i = (b_{i0})$ or by $\boldsymbol{b}_i = (b_{i0}\ b_{i1})'$ where b_{i0} is the increment in intercept for subject *i* beyond that of the mean model and b_{i1} is the increment in outcome slope for subject *i* beyond that of the mean model. For random intercept only models, $\boldsymbol{\Sigma_{bi}} = \sigma_{11}^2$ whereas for models with both random intercept and random slope we will have unstructured covariance:

$$\boldsymbol{\Sigma_{bi}} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{bmatrix}$$

For the purpose of inference on the fixed effects, the values of the random effects themselves are typically of little interest. Rather, the benefit of this structure in longitudinal data analysis is that it allows the subject specific deviations from the population average to be continuous functions of time and so too the variance of these deviations. This model assumes a linear relationship between $\boldsymbol{Y}_i$ and $\boldsymbol{X}_i$ and that $\boldsymbol{e}_i \sim N_{n_i}(0, \boldsymbol{\Sigma_{ei}})$ is independent of $\boldsymbol{b}_i \sim N_{1\ or\ 2}(0, \boldsymbol{\Sigma_{bi}})$. Therefore, $\boldsymbol{Y}_i \sim N_{n_i}\left(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{Y}_i}\right)$ where:

$$\boldsymbol{\Sigma_{Yi}} = var(\boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i) = var(\boldsymbol{Z}_i\boldsymbol{b}_i) + var(\boldsymbol{e}_i) = \boldsymbol{Z}_i\boldsymbol{\Sigma_{bi}}\boldsymbol{Z}_i' + \sigma^2\mathbf{I}_{n_i x n_i}$$

B. Estimation

Laird and Ware (1982) conceptualized the LMM as a two-stage hierarchical model. Specifically, they specified

Stage 1: $\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$ with $\boldsymbol{b}_i$ fixed and $\boldsymbol{e}_i \sim N_{n_i}\left(0, \sigma^2\mathbf{I}_{n_i x n_i}\right)$

Stage 2: $\boldsymbol{b}_i \sim N_k(0, \boldsymbol{\Sigma_{bi}})$

The marginal distribution of $\boldsymbol{Y}_i$ can then be obtained as

$$f_{\boldsymbol{Y}}(\boldsymbol{Y}_i) = \int_{-\infty}^{\infty} f_{\boldsymbol{Y}}(\boldsymbol{Y}_i|\boldsymbol{b}_i) f_{\boldsymbol{b}}(\boldsymbol{b}_i)\ d\boldsymbol{b}_i$$

For purposes of parameter estimation however they suggest a Bayesian approach treating both $\boldsymbol{b}_i$ and $\boldsymbol{\beta}$ as random variables. Specifically, they suggest for priors $\boldsymbol{\beta} \sim N_r(0, \boldsymbol{\Gamma})$ with $\boldsymbol{\Gamma}$ infinitely large and $\boldsymbol{b}_i \sim N_k(0, \boldsymbol{\Sigma_{bi}})$, giving the marginal distribution of $\boldsymbol{Y}_i$:

$$f_{\boldsymbol{Y}}(\boldsymbol{Y}_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\boldsymbol{Y}}(\boldsymbol{Y}_i|\boldsymbol{\beta}, \boldsymbol{b}_i) f_{\boldsymbol{\beta}}(\boldsymbol{\beta}) f_{\boldsymbol{b}}(\boldsymbol{b}_i)\ d\boldsymbol{\beta} d\boldsymbol{b}_i$$

They note that maximizing $f_{\boldsymbol{Y}}(\boldsymbol{Y})$ with respect to variance components of $\boldsymbol{\Sigma_Y}$ returns the REML estimates of $\boldsymbol{\Sigma}_{\boldsymbol{Y}_i}$ obtained for $\boldsymbol{Y}_i \sim N_{m_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{Y}_i})$. Additionally, the Empirical Bayes Estimate of $\boldsymbol{\beta}$ is the expected value of the posterior distribution of $\boldsymbol{\beta}$ conditioned on $\boldsymbol{\Sigma}_{\boldsymbol{Y}_i} = \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}_i}$ and is equivalent to the REML estimate of $\boldsymbol{\beta}$ obtained when assuming $\boldsymbol{\beta}$ fixed and $\boldsymbol{Y}_i \sim N_{m_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{Y}_i})$. The prediction (estimate) of $\boldsymbol{b}_i$ is often of little interest but can likewise be obtained as the expected value of the posterior distribution of $\boldsymbol{b}_i$.

The result of this equivalence between Empirical Bayes and REML estimation allows us to utilize the likelihood-based estimation techniques established in (III.D). As such, estimation for the LMM is actually quite similar to that of the GLM. Unfortunately, the frequent lack of closed form solution and bias in the estimate of the variance of $\widehat{\boldsymbol{\beta}}$ even under REML estimation means the Wald test statistics still lack known distributions in many instances.

C. The Kenward-Roger Approximation

The Kenward-Roger approximation (Kenward and Roger, 1997) attempts to address both the bias in the REML estimator of the variance of $\widehat{\boldsymbol{\beta}}$ and the frequent lack of

known distribution of the Wald test statistic for testing H0: $\boldsymbol{C'\beta = C'\beta_H}$ vs Ha: $\boldsymbol{C'\beta \neq C'\beta_H}$ in linear mixed models (and GLM), which we can express as:

$$\mathrm{W} = \frac{1}{rank(\boldsymbol{C})}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_H\right)'\boldsymbol{C}\left(\boldsymbol{C'}\widehat{\boldsymbol{\Phi}}\boldsymbol{C}\right)^{-1}\boldsymbol{C'}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_H\right) \qquad (1.1)$$

As such, there are essentially two steps to the approximation.

(i) Step 1: Modifying the REML estimate of the variance of $\hat{\beta}$ The variance of $\widehat{\boldsymbol{\beta}}$ is given as $\widehat{\boldsymbol{\Phi}} = \left(\boldsymbol{X'}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}}^{-1}\boldsymbol{X'}\right)^{-1}$ and frequently underestimates the true variance of $\widehat{\boldsymbol{\beta}}$ for two previously mentioned reasons:

(1) $\widehat{\boldsymbol{\Phi}}$ hierarchically depends on the random variable of $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}}$

(2) The REML estimate $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}}$ is often biased to some extent, though less so than the ML estimate

Issue (1) had previously been tackled by Kackar and Harville, (1984) by expressing $var\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{\Phi} + \boldsymbol{\Lambda}$ where $\boldsymbol{\Phi} = \left(\boldsymbol{X'}\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\boldsymbol{X'}\right)^{-1}$ and $\boldsymbol{\Lambda} = var\left(\widehat{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}}\right) - \widehat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}_{\boldsymbol{Y}})\right)$. They then used a Taylor expansion to obtain an approximate value for $\boldsymbol{\Lambda}$. Kenward and Roger similarly used a Taylor expansion of $\widehat{\boldsymbol{\Phi}}$ to obtain its bias as a function of the REML estimate $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}}$. Combining these two results, Kenward and Roger obtained the adjusted estimate of the variance of $\widehat{\boldsymbol{\beta}}$ given as: $\widehat{\boldsymbol{\Phi}}_{\mathrm{A}} = \widehat{\boldsymbol{\Phi}} + 2\widehat{\boldsymbol{\Lambda}}$. The calculation of $\widehat{\boldsymbol{\Lambda}}$ is provided in Appendix II part A, but for now it suffices to simply note an adjusted estimate is made to the REML based estimate of $var\left(\widehat{\boldsymbol{\beta}}\right)$ which reduces the bias in the estimate from $\mathrm{O}(n^{-2})$ to $\mathrm{O}\left(n^{-5/2}\right)$ (Alnosaier, 2007).

(ii) Step 2: Approximating the distribution of the Wald test statistic The Kenward-Roger method now aims to provide an approximation to the distribution of the adjusted Wald type F statistic:

$$W = \frac{1}{rank(\boldsymbol{C})}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_H)'\boldsymbol{C}(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_A\boldsymbol{C})^{-1}\boldsymbol{C}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_H)$$

To outline the approach, they first use a second order Taylor Expansion to obtain approximations to the mean and variance of the F statistic. However, there is an error term in any expansion that will result in inaccuracies particularly when the sample size is small. Consequently, they later modify the results to match the mean and variance known in certain exact cases. For the initial step of the Taylor Expansion, they approximate the moments which may be given by the following well-known theorem:

Theorem 1 (Mean and Variance of Normal Quadratic Form):

Let $Y = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$ with $\boldsymbol{X} \sim \boldsymbol{N}_l(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then,

$\mathbf{E}(Y) = tr(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}$

$\mathbf{V}(Y) = 2tr[(\boldsymbol{A}\boldsymbol{\Sigma})^2] + 4\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\mu}$

With the condition that X must be a column vector and Y a scalar

In terms of Theorem 1 we have $\frac{1}{l}\mathrm{F} = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}$ where $rank(\boldsymbol{C}) = l$ and

$\boldsymbol{X} = \boldsymbol{C}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_H) \rightarrow \boldsymbol{\mu} = \boldsymbol{C}'(\boldsymbol{\beta} - \boldsymbol{\beta}_H)$

$\boldsymbol{A} = (\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_A\boldsymbol{C})^{-1}$

$\boldsymbol{\Sigma} = \mathbf{var}(\boldsymbol{C}'\widehat{\boldsymbol{\beta}}) = \boldsymbol{C}'\mathbf{var}(\widehat{\boldsymbol{\beta}})\boldsymbol{C} = \boldsymbol{C}'(\boldsymbol{\Phi} + \boldsymbol{\Lambda})\boldsymbol{C}$

Therefore

$$\mathrm{E}(W) = tr\left[(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_A\boldsymbol{C})^{-1}(\boldsymbol{C}'(\boldsymbol{\Phi} + \boldsymbol{\Lambda})\boldsymbol{C})\right] + (\boldsymbol{\beta} - \boldsymbol{\beta}_H)'\boldsymbol{C}(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_A\boldsymbol{C})^{-1}\boldsymbol{C}'(\boldsymbol{\beta} - \boldsymbol{\beta}_H)$$

$$\mathrm{V}(W) = 2tr\left[\left(\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}(\boldsymbol{C}'(\boldsymbol{\Phi}+\boldsymbol{\Lambda})\boldsymbol{C})\right)^2\right]$$

$$+ 4(\boldsymbol{\beta}-\boldsymbol{\beta}_H)'\boldsymbol{C}\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}(\boldsymbol{C}'(\boldsymbol{\Phi}+\boldsymbol{\Lambda})\boldsymbol{C})\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}\boldsymbol{C}'(\boldsymbol{\beta}-\boldsymbol{\beta}_H)$$

Unfortunately, each of these terms depends on $\widehat{\boldsymbol{\Phi}}$ and therefore on $\widehat{\boldsymbol{\Sigma}}$, and so we must use the hierarchical identities:

$$\mathrm{E}(W) = \mathrm{E}\left(\mathrm{E}\left(W|\widehat{\boldsymbol{\Sigma}}\right)\right)$$

$$\mathrm{V}(W) = \mathrm{E}\left(\mathrm{V}\left(W|\widehat{\boldsymbol{\Sigma}}\right)\right) + \mathrm{V}\left(\mathrm{E}\left(W|\widehat{\boldsymbol{\Sigma}}\right)\right)$$

Kenward and Roger only approximate these moments under the null hypothesis. So,

$$\mathrm{E}(\mathrm{W}) = \mathrm{E}\left(\mathrm{E}\left(F|\widehat{\boldsymbol{\Sigma}}\right)\right)$$

$$= \mathrm{E}\left(tr\left[\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}(\boldsymbol{C}'(\boldsymbol{\Phi}+\boldsymbol{\Lambda})\boldsymbol{C})\right]\right)$$

$$+ \mathrm{E}\left((\boldsymbol{\beta}-\boldsymbol{\beta}_H)'\boldsymbol{C}\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}\boldsymbol{C}'(\boldsymbol{\beta}-\boldsymbol{\beta}_H)\right)$$

$$= \mathrm{E}\left(tr\left[\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}(\boldsymbol{C}'(\boldsymbol{\Phi}+\boldsymbol{\Lambda})\boldsymbol{C})\right]\right)$$

Taking the expectation of the Taylor expansion of $tr\left[\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}(\boldsymbol{C}'(\boldsymbol{\Phi}+\boldsymbol{\Lambda})\boldsymbol{C})\right]$ around $\boldsymbol{\sigma}$ they arrive at

$$\mathrm{E}(W) = 1 + \frac{\mathrm{A}_2}{l} + O\left(n^{-\frac{3}{2}}\right)$$

where l =rank(C). A similar process for the variance leads to

$$\mathrm{V}(W) = \frac{2}{l}\left[1 + \frac{1}{2l}(\mathrm{A}_1 + \mathrm{A}_2)\right] + O\left(n^{-\frac{3}{2}}\right)$$

Where:

$$\mathrm{A}_1 = \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\, tr(\boldsymbol{\theta}\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi})\, tr\left(\boldsymbol{\theta}\boldsymbol{\Phi}\mathbf{P}_j\boldsymbol{\Phi}\right)$$

$$A_2 = \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij} tr\left(\boldsymbol{\theta\phi P}_i\boldsymbol{\phi\theta\phi P}_j\boldsymbol{\phi}\right)$$

$w_{ij} = cov\left(\hat{\sigma}_i, \hat{\sigma}_j\right) = ij^{th}$ entry of the inverse of the expected information matrix

$$\boldsymbol{\theta} = \mathbf{C}(\mathbf{C}'\boldsymbol{\phi}\mathbf{C})^{-1}\mathbf{C}'$$

$$\mathbf{P}_i = -\mathbf{X}'\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i}\boldsymbol{\Sigma}^{-1}\mathbf{X}$$

Lastly, they solve for values of λ and m such that $\lambda F = F^* \sim F(l, m)$ which they do by matching the obtained quantities to the moments of the F distribution. The solution obtained does not match the correct values in certain known cases (ANOVA and Hotelling T^2 tests), which is unsurprising because the Taylor expansion has an error term which is dropped from the mean and variance expressions. To accommodate this, they modified their solution to the mean and variance expressions so that the exact value is returned for these known cases. The derivations of these solutions are provided in rigorous detail in Alnosaier (2007).

We can now show an equivalent example to that in previous section I.D, but using the LMM with random intercept and slope and using the Kenward-Roger test statistic for the group by time effect. Now, the contrast matrix for the test of the hypothesis H0: $\mathbf{C}'\boldsymbol{\beta} = \mathbf{0}$ will be

$$\boldsymbol{C} = [0 \;\; 0 \;\; 0 \;\; 1]'$$

And for subject *i* the design matrix will be:

$$\boldsymbol{X}_i = \begin{bmatrix} 1 & g_i & 0 & g_i * 0 \\ 1 & g_i & 3 & g_i * 3 \\ 1 & g_i & 6 & g_i * 6 \\ 1 & g_i & 9 & g_i * 9 \\ 1 & g_i & 12 & g_i * 12 \end{bmatrix}$$

where again $g_i = 0$ if subject *i* belongs to control group and $g_i = 1$ if subject *i* belongs to the treatment group. Likewise, we'll have within subjects covariance matrix:

$$\boldsymbol{\Sigma_{bi}} = \begin{bmatrix} \sigma_{00}^2 & \sigma_{01} \\ \sigma_{01} & \sigma_{11}^2 \end{bmatrix}$$

with σ_{00}^2 being the variance of random intercept, σ_{11}^2 that of random slope, and σ_{01} the covariance between random intercept and random slope.

And so the variance of $\boldsymbol{Y}_i$ will be $\boldsymbol{Z}_i\boldsymbol{\Sigma_{bi}}\boldsymbol{Z}_i{}' + \sigma^2\mathbf{I}_{5x5}$ with

$$\boldsymbol{Z}_i = \begin{bmatrix} 1 & 0 \\ 1 & 3 \\ 1 & 6 \\ 1 & 9 \\ 1 & 12 \end{bmatrix}$$

Now, using the Kenward-Roger F statistic, F_{KR}, we will have $F_{KR} = 18 * tr[\boldsymbol{HE}^{-1}]$ from I.D and will be tested using the α level rejection region of an $F_{1,18}$ distribution, which we've seen to be the exact value for the test of the group by time interaction. Thus, while in general the Wald test statistic *W* for the LMM does not have an easily calculable distribution, the Kenward-Roger adjustment has caused the exact value for this hypothesis known from the GLMM to be returned.

V. Existing Power Calculation Methods

As described, the KR approach only approximates the distribution of the Wald statistic under the null hypothesis, which means calculating power for this adjusted statistic is not straightforward. Existing power calculations for the KR adjusted Wald test can be described as belonging to at least one of three approaches: equivalent, parallel, or simulated. Equivalent approaches calculate power for a test statistic (such as the HLT statistic in the GLMM) that is known to be equivalent to the KR statistic in certain cases. Parallel approaches calculate power for some other LMM Wald statistic approximation, with the idea being that power for the KR statistic should be similar. Lastly, simulated

approaches provide an estimate of power by simulating the specified trial a number of times and observing the proportion of times the null hypothesis is rejected.

One straightforward exact approach is to use the equivalence between the HLT statistic in the GLMM and KR statistic in the LMM. A 2018 article by Chi et al. provides a detailed set of criteria under which the LMM be recast as a GLMM and the KR statistic obtained via REML estimation will be equivalent to the HLT trace test statistic[2]. The resulting implications are powerful and suggest the tests of hypotheses of any individual fixed effects in the linear mixed model with balanced data (balanced again meaning equal observation numbers and times for all subjects with no time-varying covariates) has an exact distribution for which power can easily be calculated by using existing multivariate techniques. When data are unbalanced or hypotheses of interest involve more complicated contrasts, such as those involving more than two groups or those involving multiple effects such as omnibus ANOVA type tests, the HLT and KR statistics no longer must be equivalent and the distribution of the HLT must itself be approximated. However, using existing power calculations for the HLT in such instances still provides an intuitive parallel approach to calculating power for the KR adjusted Wald test.

An alternative parallel calculation was developed by Kreidler (2014). While the KR method matches the values of the first two moments of W_{A}obtained via Taylor expansion up to those of an F distribution, the method described by Kriedler takes the two step approach of of

(1) approximating $\left(\boldsymbol{C}'\left(\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}'\right)^{-1}\boldsymbol{C}\right)^{-1}$ as a Wishart distribution and

(2) making the assumption that $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\beta}}$ are independent, so that

$\widehat{\boldsymbol{\beta}}'\boldsymbol{C}\left(\boldsymbol{C}'\left(\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}\right)^{-1}\boldsymbol{C}\right)^{-1}\boldsymbol{C}'\widehat{\boldsymbol{\beta}}$ has a known correspondence with an F distribution for
testing H0: $\boldsymbol{C}'\boldsymbol{\beta}$=0

Specifically, in step 1, $\widehat{\boldsymbol{\Sigma}}$ is the REML estimate obtained after recasting the LMM as the GLMM in section A.I. For balanced mixed models, a single estimate of $\widehat{\boldsymbol{\Sigma}}$ is obtained as shown in A.II and so for the LMM: $\boldsymbol{C}'\left(\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}\right)^{-1}\boldsymbol{C}\sim W_a\left(N-q,\boldsymbol{C}'\left(\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}\right)^{-1}\boldsymbol{C}\right)$ (the proof being similar to that of $\boldsymbol{E}$ in the multivariate model). However, when data are unbalanced, as can occur with missing or mistimed data, the LMM is recast as a collection of m multivariate linear models grouped by observational pattern. For each multivariate model d=1,…,m, $\widehat{\boldsymbol{\Sigma}}_d$ is estimated. Then for each component of the LMM, $\boldsymbol{X}_d'\widehat{\boldsymbol{\Sigma}}_d^{-1}\mathbf{X}_d$ is inverse Wishart distributed, but for the overall LMM, $\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}=\boldsymbol{X}_1'\widehat{\boldsymbol{\Sigma}}_1^{-1}\mathbf{X}_1+\cdots+\boldsymbol{X}_m'\widehat{\boldsymbol{\Sigma}}_m^{-1}\mathbf{X}_m$ has an unknown distribution.

The distribution of $\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}=\boldsymbol{X}_1'\widehat{\boldsymbol{\Sigma}}_1^{-1}\mathbf{X}_1+\cdots+\boldsymbol{X}_m'\widehat{\boldsymbol{\Sigma}}_m^{-1}\mathbf{X}_m$ is instead approximated by matching the moments of the sum to those of an inverse Wishart distribution. The Wishart and inverse Wishart distributions have a variance for each element, but not for the matrix as a whole. So, the expected value of the sum is matched to the expected value of an inverse Wishart matrix, and the variance of the trace of the sum is matched to the variance of the trace of an inverse Wishart matrix. The degrees of freedom N_* and covariance matrix $\widehat{\boldsymbol{\Sigma}}_*^{-1}$ are solved for so that the distribution of $\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}$ is then approximated as a $W_q^{-1}(N_*,\boldsymbol{\Sigma}_*^{-1})$ distribution. Consequently

$\left(\boldsymbol{C}'\left(\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}\right)^{-1}\boldsymbol{C}\right)^{-1}\sim W_a^{-1}(N_*,\boldsymbol{C}'(\mathbf{X}'\boldsymbol{\Sigma}_*^{-1}\mathbf{X})^{-1}\boldsymbol{C})$.

In step 2, known relationships between an F distribution and quadratic forms of $\mathbf{X}'\mathbf{A}\mathbf{X}$ with $\boldsymbol{X}\sim N_\tau(\boldsymbol{\mu},\boldsymbol{\Sigma}_x)$ and $\mathbf{A}\sim W_\tau^{-1}(N,\boldsymbol{\Sigma}_A)$ are used to express the Wald statistic w for

the linear mixed model as a scaled F distributed random variable. Specifically, the assumption is made that $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Sigma}}$ are independent—an assumption suitable for multivariate methods, although which may not hold for the mixed model. So, $w_0 \sim \lambda_0^* F(n_u, N_* - r + a - 2)$ and $w_a \sim \lambda_a^* F(n_u, N_* - r + a - 2, \delta_u)$ where w_0, w_a and λ_0^*, λ_a^* represent Wald statistics and scale factors under the null and alternative hypotheses. Lastly $\mathrm{E}(w_0)$, $\mathrm{E}(w_a)$, $\mathrm{V}(w_a)$ are obtained and matched to a statistic of the form given by Kenward and Roger, namely such that $\lambda w \sim \mathrm{F}(l, \nu, \omega)$.

Power under the alternative that $\mathrm{C}'\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ can then be calculated by defining values of $\alpha, \boldsymbol{\beta}, \boldsymbol{\beta}_0, \boldsymbol{C}$, and $\boldsymbol{X}$ to obtain the distribution of λw under the null hypothesis that $\mathrm{C}'\boldsymbol{\beta} = \boldsymbol{\beta}_0$. The value of $f_{crit} = \mathrm{F}^{-1}(1 - \alpha, l, \nu)$ can be obtained and the power then calculated as $1 - \mathrm{F}(f_{crit}, l, \nu, \omega)$. This method has the advantage of allowing flexibility in anticipated observational pattern, such as allowing for missingness. Moreover, it was found to perform within roughly two-decimal places of accuracy in simulations of a longitudinal analysis with moderate sample sizes. However, the assumption employed by this methd that $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\beta}}$ are assumed to be independent is frequently violated in the mixed model. As such, the variability of $\widehat{\boldsymbol{\beta}}$ may be underestimated by this approach in cases of unbalanced data. The KR approach instead uses an adjusted estimator of $(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$, and so it would also be desirable to have a power calculation for the KR test statistic that accommodates this adjustment. In the following chapter, we will introduce a power calculation that has the same advantages as that introduced by Kriedler, but which addresses some of the limitations of the method for calculating power for the Kenward-Roger test statistic.

Another parallel method is provided by Tang (2017). In this approach, a closed-form estimate, $\widehat{\boldsymbol{V}}$, is found for the KR derived expression of $var(\widehat{\boldsymbol{\beta}})$ under REML estimation and monotone missingness. The power to calculate the test of treatment effect at last visit is obtained as $1 - P\left(t \leq t_{m,\frac{\alpha}{2}}\right)$ where $t = \frac{\beta_{treatment}}{V}$ follows a noncentral t distribution with noncentrality parameter $\omega = \frac{\beta_{treatment}}{V}$ and degrees of freedom m obtained approximately as the fraction of observed information retained at the study visit. However, this methodology has only been established for comparisons between two groups at a particular timepoint. Likewise, the method for determining degrees of freedom in the t-distribution used for power calculations differs from the KR approach.

Perhaps the simplest of such parallel approaches would be to calculate power for the traditional form of the Wald test statistic (equation 1.1) using the residual degrees of freedom. Prior to the development of the KR adjusted statistic, Helms (1991) advocated for such an approach to approximate the distribution of the Wald test statistic and thereby calculate power for unbalanced study designs. Specifically, it was argued the Wald test statistic approximately followed an F distribution with numerator degrees of freedom established in the usual way, denominator degrees of freedom equal to N – rank(**X**||**Z**), with N = $n\tau$ (i.e. the total number of observations) and noncentrality parameter equal to $\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_H\right)' \boldsymbol{C}\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}\boldsymbol{C}\right)^{-1}\boldsymbol{C}'\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_H\right)$. This approach was likewise used by Verbeke and Lessafre (1999) in calculating power for longitudinal mixed models with monotonically missing data, although they specifically noted this method had not been validated for the KR adjusted Wald test. Nevertheless, the simplicity of this approach is appealing and provides a straightforward way to plan for missing observations in power calculations,

and it would be useful to know if such an approach could adequately calculate power for the KR adjusted Wald test.

The final type of power calculation is to simply estimate power based on a number of simulated trials. Unfortunately, this approach can be prohibitively time consuming in a number of ways. First, code must be written to perform the simulation, and doing so for LMMs can be complicated and require several days of writing and validation. While some existing software such as Power and Sample Size Software (PASS, 2015) provides a pre-existing simulated power option for the KR test, we do not find any pre-existing validated packages that easily accommodate for anticipated dropout, missing observations, or other sources of unbalanced data. Regardless, the biggest limitation to this approach is that such simulations can take hours to run. While this issue would not be so problematic if only power were to be calculated, the ultimate question on the mind of many researchers is not power but rather sample size. As a result, calculating sample size would require iteratively performing many simulations, each of which could potentially take many hours, until the correct sample size is obtained. Therefore, a fast method of calculating power would still be preferable to the simulation-based approach. It should, however, be noted that in our opinion a small simulation study to accompany any power calculation approaches may still be beneficial in providing insight into convergence of estimates and what types of covariance structures may be feasible to model.

CHAPTER 2
A NEW POWER CALCULATION METHOD FOR THE KENWARD-ROGER TEST STATISTIC

I. Introduction

To the author's knowledge, no method currently exists that calculates power for the Kenward-Roger test statistic directly for tests of linear functions of time. Existing power calculations are identical to that of the KR statistic in certain cases, but they are technically calculating power for different test statistics which, while hopefully similar, may not always align with the KR statistic. In this chapter, we introduce a power method with the KR statistic as its target. Specifically, the theory and techniques used for the KR statistic described in Chapter I section IV.C is applied to obtain an approximate distribution for the Wald test statistic in the LMM under the alternative which can be used to calculate power.

Unfortunately, an identical approach to that introduced by Kenward and Roger for the null hypothesis of the Wald test is not tractable due to the number of terms which must be calculated and then solved for. Recalling the Wald test statistic for the LMM is given as

$$W = \frac{1}{rank(\boldsymbol{C})}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_H\right)'\boldsymbol{C}\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}\boldsymbol{C}'\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_H\right)$$

then, if the KR approach is to be replicated under the alternative hypotheses, we must obtain the first two moments of *W*:

$$\mathrm{E}(W) = \mathrm{E}\left(\mathrm{E}\left(W|\widehat{\boldsymbol{\Sigma}}\right)\right)$$

$$\mathrm{V}(W) = \mathrm{E}\left(\mathrm{V}\left(W|\widehat{\boldsymbol{\Sigma}}\right)\right) + \mathrm{V}\left(\mathrm{E}\left(W|\widehat{\boldsymbol{\Sigma}}\right)\right),$$

with

$$\mathrm{E}\left(\mathrm{E}\left(W|\widehat{\boldsymbol{\Sigma}}\right)\right) = \mathrm{E}\left(tr\left[\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}\boldsymbol{C}\right)^{-1}\left(\boldsymbol{C}'\boldsymbol{\Phi}\boldsymbol{C}\right)\right]\right) + \mathrm{E}\left((\boldsymbol{\beta}-\boldsymbol{\beta}_H)'\boldsymbol{C}\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}\boldsymbol{C}'(\boldsymbol{\beta}-\boldsymbol{\beta}_H)\right),$$

$$\mathrm{E}\left(\mathrm{V}\left(W|\widehat{\boldsymbol{\Sigma}}\right)\right) = \mathrm{E}\left(2tr\left[\left(\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}\left(\boldsymbol{C}'\boldsymbol{\Phi}\boldsymbol{C}\right)\right)^2\right]\right)$$

$$+ \mathrm{E}\left(4(\boldsymbol{\beta}-\boldsymbol{\beta}_H)'\boldsymbol{C}\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}\left(\boldsymbol{C}'\boldsymbol{\Phi}\boldsymbol{C}\right)\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}\boldsymbol{C}'(\boldsymbol{\beta}-\boldsymbol{\beta}_H)\right)$$

$$\mathrm{V}\left(\mathrm{E}\left(W|\widehat{\boldsymbol{\Sigma}}\right)\right) = V\left(\widehat{\Psi}\right) + V(\widehat{\omega}) + 2Cov\left(\widehat{\Psi},\widehat{\omega}\right)$$

$$\widehat{\Psi} = tr\left(\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}\left(\boldsymbol{C}'\boldsymbol{\Phi}\boldsymbol{C}\right)\right), \quad \widehat{\omega} = (\boldsymbol{\beta}-\boldsymbol{\beta}_H)'\widehat{\boldsymbol{\theta}}_{\mathrm{A}}(\boldsymbol{\beta}-\boldsymbol{\beta}_H).$$

All of these terms would need to be calculated via Taylor expansion, and the results would need to be modified to match the moments of an F distribution so that the exact values are returned in the known cases tracked by the KR statistic (HLT and ANOVA F-tests). Instead, we make a simplifying assumption to improve tractability of obtaining an approximate distribution under the alternative hypothesis.

Specifically, we are unaware of any exact scaled F-distributed test statistics for which null and alternative hypotheses have different degrees of freedom or scale factor. We therefore make the assumption that the KR test statistic has identical degrees of freedom m and scale factor λ under both null and alternative hypotheses. Under this assumption, we can first calculate m and λ using the expected data pattern and population parameters (instead of their estimated values) with the existing method provided by Kenward and Roger. We then only need to calculate $\mathrm{E}(W)$ under the alternative

hypothesis, which is rather straightforward. Doing so, we can easily solve for the noncentrality parameter of the desired F distribution.

II. Derivation of Methods

A. Method 1: Replicating the KR Adjustment for the Alternative Hypothesis

We must first calculate $\mathrm{E}(W)$. We only need to calculate the value of $\mathrm{E}\left((\boldsymbol{\beta}-\boldsymbol{\beta}_H)'\boldsymbol{C}\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}\boldsymbol{C}'(\boldsymbol{\beta}-\boldsymbol{\beta}_H)\right)$, as from Alnosaier (2007) we have:

$$\mathrm{E}\left(tr\left[\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}\boldsymbol{C}\right)^{-1}\left(\boldsymbol{C}'\boldsymbol{\Phi}\boldsymbol{C}\right)\right]\right)=l+\mathrm{A}_2+O\left(n^{-\frac{3}{2}}\right)$$

First, note that for

$\mathrm{H}_0:\ C'(\boldsymbol{\beta}-\boldsymbol{\beta}_H)=C'(\boldsymbol{\beta}-\boldsymbol{\beta})=0$

$\mathrm{H}_a: C'(\boldsymbol{\beta}-\boldsymbol{\beta}_H)=C'(\boldsymbol{\beta}-\mathbf{0})=\boldsymbol{\beta}$, at least for the purpose of calculating power.

Therefore, $\mathrm{E}\left((\boldsymbol{\beta}-\boldsymbol{\beta}_H)'\boldsymbol{C}\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}\boldsymbol{C}'(\boldsymbol{\beta}-\boldsymbol{\beta}_H)\right)=\mathrm{E}\left(\boldsymbol{\beta}'\boldsymbol{C}\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}\boldsymbol{C}'\boldsymbol{\beta}\right)$. After calculating this value via Taylor Expansion as in Appendix II part B. we then obtain:

$$E(\mathrm{W})=\frac{1}{l}(l+\mathrm{A}_2+A_3+\boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta}-\boldsymbol{\beta}'\boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta}\boldsymbol{\beta})+O(n^{-1})$$

Under the assumption that scale factor and degrees of freedom are the same under null and alternative hypotheses, we now only must solve for the noncentrality parameter, ω, to obtain the approximate distribution of the Wald statistic under the alternative hypothesis. Due to the remainder term in the Taylor expansion, the solution will not return the correct value in exact cases, and so as in the Kenward-Roger method for the null hypothesis, we will adjust the expression for E(W) to return the correct expected

value under the alternative hypothesis and then use this new value to obtain the noncentrality parameter for the KR test statistic. First, we will let:

$$\mathrm{E(W)} = \mathrm{E}_0\mathrm{(W)} + \mathrm{E}_a\mathrm{(W)}$$

where

$$\mathrm{E}_0\mathrm{(W)} = \frac{1}{l}(l + \mathrm{A}_2) \quad , \quad \mathrm{E}_a\mathrm{(W)} = \frac{1}{l}(\boldsymbol{\beta'\theta\beta} + A_3 - \boldsymbol{\beta'\theta A\theta\beta})$$

Now, if $\mathrm{F} = \lambda\mathrm{W}$, then $\mathrm{E(F)} = \lambda\mathrm{E(W)} = \lambda\big(\mathrm{E}_0\mathrm{(W)} + \mathrm{E}_a\mathrm{(W)}\big)$. Recall that for the noncentral F distribution with noncentrality parameter ω, numerator degrees of freedom l, and denominator degrees of freedom m, that

$$\mathrm{E(F)} = \frac{m(l+\omega)}{l(m-2)}$$

We can then solve for the noncentrality parameter ω as:

$$\mathrm{E(F)} = \lambda\mathrm{E(W)}$$

$$\frac{m(l+\omega)}{l(m-2)} = \lambda\big(\mathrm{E}_0\mathrm{(W)} + \mathrm{E}_a\mathrm{(W)}\big)$$

Additionally, knowledge about the value of ω in various tests suggest ω may be broken into various components: an effect size, δ, and a scale factor, δ, such that $\omega = \gamma\delta$. For instance, in the test of linear trend given in the example from Chapter 1 section II.D, the effect size may here be given as $\delta = \boldsymbol{\beta'\theta\beta}$ and the scale factor may be given as $\gamma = 1$ so that $\omega = 1(\boldsymbol{\beta'\theta\beta}) = \boldsymbol{\beta'\theta\beta}$. Alternatively, in the two sample Hotelling model briefly mentioned in Chapter 1 section II.C.iv, the effect size is again given as $\delta = \boldsymbol{\beta'\theta\beta}$, but the scale factor is given as $\gamma = \frac{n_1 n_2}{n_1+n_2}$ so that $\omega = \frac{n_1 n_2}{n_1+n_2}\boldsymbol{\beta'\theta\beta}$**.** So:

$$\frac{m(l+\omega)}{l(m-2)} = \frac{m(l+\gamma\nu)}{l(m-2)} = \lambda\big(\mathrm{E}_0\mathrm{(W)} + \mathrm{E}_a\mathrm{(W)}\big)$$

$$\omega = \gamma\delta = l\lambda\left(\frac{m-2}{m}\right)\left(\mathrm{E}_0(\mathrm{W}) + \mathrm{E}_a(\mathrm{W})\right) - l$$

Noting that $\lambda = \frac{m}{\mathrm{E}_0(\mathrm{W})(m-2)}$ we simplify such that

$$\omega = \gamma\delta = l\frac{\mathrm{E}_a(\mathrm{W})}{\mathrm{E}_0(\mathrm{W})}$$

In exact cases, $\mathrm{E}_0(\mathrm{W})$ should equal $\frac{m}{m-2}$ with m being the correct denominator degrees of freedom for the Wald F statistic. However, this does not end up being the case, and so the KR method instead uses $\mathrm{E}_0^*(\mathrm{W}) = \left(1 - \frac{\mathrm{A}_2}{l}\right)^{-1}$ to solve for degrees of freedom and scale parameter under the null hypothesis so that degrees of freedom and scale factor are correct. Under the alternative hypothesis in exact cases:

$$\mathrm{E}_a(\mathrm{W}) = \frac{1}{l}\left(\frac{m+2}{m}\right)\boldsymbol{\beta'\theta\beta}$$

Therefore, in exact cases (see Appendix II part C for details):

$$\omega = \gamma\delta = \ l\frac{\mathrm{E}_a(\mathrm{W})}{\mathrm{E}_0^*(\mathrm{W})} = \frac{\left(\frac{m+2}{m}\right)\boldsymbol{\beta'\theta\beta}}{\frac{m}{m-2}} = \frac{m^2-4}{m^2}\boldsymbol{\beta'\theta\beta} \neq \boldsymbol{\beta'\theta\beta}$$

To ensure the correct noncentrality parameter is returned in exact cases, we make a similar modification to $\mathrm{E}_a(\mathrm{W})$ as to that made to $\mathrm{E}_0(\mathrm{W})$ by Kenward and Roger. Specifically we will let

$$\mathrm{E}_a^*(\mathrm{W}) = \begin{cases} \frac{1}{l}[(\boldsymbol{\beta'\theta\beta} - A_3 + \boldsymbol{\beta'\theta A\theta\beta})^{-1}](\boldsymbol{\beta'\theta\beta} - \boldsymbol{\beta'\theta A\theta\beta})^2 & ,\boldsymbol{\beta'\theta\beta} \neq 0 \\ 0 & ,\boldsymbol{\beta'\theta\beta} = 0 \end{cases}$$

In the above expression, the choice of multiplying by $(\boldsymbol{\beta'\theta\beta} - \boldsymbol{\beta'\theta A\theta\beta})^2$ vs multiplying by $(\boldsymbol{\beta'\theta\beta})^2$ is to some extent arbitrary. However, from Appendix II part B, we see that $\boldsymbol{\beta'\theta\beta} - \boldsymbol{\beta'\theta A\theta\beta}$ is approximately equal to $\boldsymbol{\beta'C(C'\phi_A C)^{-1}C'\beta}$ and therefore to the "effect size" of the test statistic in cases where $\boldsymbol{\phi} \neq \boldsymbol{\phi}_\mathrm{A}$, such as in many unbalanced

designs. So, multiplying by $(\boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta} - \boldsymbol{\beta}'\boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta}\boldsymbol{\beta})^2$ captures more information pertaining to the test statistic.

Note also that in the exact cases of balanced data and single rank contrast

$$\frac{\mathrm{E}_a^*(\mathrm{W})}{\mathrm{E}_a(\mathrm{W})} = \frac{\mathrm{E}_0^*(\mathrm{W})}{\mathrm{E}_0(\mathrm{W})}$$

and so the adjustment to the expected value of the Wald statistic under the alternative hypothesis mirrors that made by Kenward and Roger for the null hypothesis in both form and value. Now,

$$\omega = \gamma\delta = \; l\frac{\mathrm{E}_a^*(\mathrm{W})}{\mathrm{E}_0^*(\mathrm{W})} = \frac{\frac{m}{m-2}\boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta}}{\frac{m}{m-2}} = \boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta}$$

which is the correct value in the exact case when data are balanced and the contrast matrix is rank one.

One problem still remains, however. Specifically, the choices of effect size δ and scale factor γ must be more generally determined. In the exact test for linear trend with rank one contrast we want $\delta = \boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta}$ and $\gamma = 1$. However when the exact distribution of the test statistic is unknown, the desired values of each become less clear. We turn instead to the relationship between the F distribution and noncentrality parameter in the linear model utilized by Muller and Peterson (1984). Specifically, the noncentrality parameter for the distribution of the F statistic in the balanced linear model may be expressed as $\omega = lF$. Now if $F = \lambda W$, then under this framework we would have $\omega = l\lambda W = \lambda l\frac{1}{l}\boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta} = \lambda\boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta}$, suggesting that again $\delta = \boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta}$ but now also that $\gamma = \lambda$.

Currently, our adjustment to the approximation of the noncentrality parameter only returns the effect size for balanced data, but the effect size is not scaled. For

instance, the ANOVA type omnibus F-test of all $\beta_i = 0$ (here the contrast is no longer rank one and $\lambda \neq 1$, nor does the test have an exact known distribution) we have when data are balanced:

$$\omega = \gamma\delta = \ l\frac{\mathrm{E}_a^*(\mathrm{W})}{\mathrm{E}_0^*(\mathrm{W})} = \boldsymbol{\beta'\theta\beta} \neq \lambda\boldsymbol{\beta'\theta\beta}$$

We therefore propose the final modification that:

$$\omega = \ l\lambda\frac{\mathrm{E}_a^*(\mathrm{W})}{\mathrm{E}_0^*(\mathrm{W})}$$

Therefore, in tests of linear trend, the desired (exact if the contrast is single rank) values of $\omega = \gamma\delta = \lambda\boldsymbol{\beta'\theta\beta}$ will be returned in balanced cases.

B. Alternative methods for consideration

While the method just described obtains a value for the noncentrality parameter using the methodology described by Kenward and Roger to obtain the denominator degrees of freedom and scale factor for the F distribution under the null hypothesis, this approach is admittedly complicated. We therefore also introduce three simple intuitive methods that could potentially be used to calculate power for the KR adjusted Wald test. The first of these alternative methods (which we'll refer to as method 2 to distinguish it from the expansion based method ("method1") described in previous section II.A), replicates the approach of Muller and Peterson (1984) in directly using the KR adjusted Wald statistic to obtain the value of the noncentrality parameter. Specifically, under this second method we let the noncentrality parameter simply be (with $\boldsymbol{\Phi}_\mathrm{A} = \boldsymbol{\Phi} + \mathbf{A}$):

$$\omega = lF = l\lambda W_\mathrm{A} = \lambda\boldsymbol{\beta'C}(\boldsymbol{C'\Phi}_\mathrm{A}\boldsymbol{C})^{-\mathbf{1}}\boldsymbol{C'\beta}$$

The distribution of the Wald test statistic under the hypothesized "true" model parameter values is then obtained as $\lambda W_A \sim F(l, m, \omega)$, with λ, l, and $m = m_{KR}$ obtained by using the method described by Kenward and Roger under the null hypothesis using the "true" parameter values in lieu of their estimates. The second of these additional methods (method 3) is identical to method 2, except that the noncentrality parameter is the traditional value in the linear model, i.e.

$$\omega = \boldsymbol{\beta}'\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{\Phi}\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{\beta}$$

which in simulation studies will allow us to examine the impact of the adjustments to the Wald statistic made by Kenward and Roger beyond the impact of only the modifications to the denominator degrees of freedom.

The third of these additional methods (method 4) again uses the traditional value of the linear model. However, the denominator degrees of freedom depend on the random effects to be modeled. If both intercept and slope are modeled as random effects, then the denominator degrees of freedom are set as

$$m = n(1-p) - l - 1$$

where p is the proportion of follow-up observations missing out of the total number possible. For instance, in a design where each subject is supposed to be observed at $\tau = 5$ time points, but in the actual trial 15% of follow-up observations are missing, then m would equal $n(4*0.85+1)/5 - l - 1 = 0.88n - l - 1$. In other words, this method simply scales the denominator degrees of freedom by the proportion of data actually observed. Similarly, when only a random intercept is included in the model, then

$$m = n(1-p)(\tau - 1) - l - 1$$

where again p is the proportion of follow-up observations missing out of the total number possible. The reason the denominator degrees of freedom differs depending on random effects included is that this method tries to make a simple adjustment based on the KR value of m in balanced designs for each of these models. Specifically, when data are balanced, then the KR degrees of freedom for random intercept and slope models will be $m = n - l - 1$, whereas for random intercept only models the degrees of freedom will be $m = n(\tau - 1) - l - 1$.

Lastly, we examined the ability of one simple existing approach (which we'll refer to as method 5) to calculate power for the KR adjusted Wald test. Specifically, we use the method described by Helms (1992) to approximate the distribution of the Wald statistic from which power may be calculated. This method uses the traditional value of the noncentrality parameter in the linear model, but the denominator degrees of freedom are given as:

$$m = N - rank(\boldsymbol{X}||\boldsymbol{Z})$$

where N is the total number of observations actually collected in the study, and "||" is the concatenation operator. So, for a study with no missing observations, $N = n\tau$. Additionally, since in the designs of interest for this dissertation the columns of Z are also columns of X, $rank(\boldsymbol{X}||\boldsymbol{Z}) = rank(\boldsymbol{X})$. A summary of these differing methods is presented in Table 2.1.

Table 2.1: Summary of F distribution approximation methods

Method	Random Effects	$\boldsymbol{m}$	$\boldsymbol{\omega}$
1	Either	m_{KR}	$l\lambda\frac{\mathrm{E}_a^*(\mathrm{W})}{\mathrm{E}_0^*(\mathrm{W})}$
2	Either	m_{KR}	$\lambda\boldsymbol{\beta}'\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{\phi}_{\mathrm{A}}\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{\beta}$
3	Either	m_{KR}	$\boldsymbol{\beta}'\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{\phi}\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{\beta}$
4	Intercept and slope	$n(1-p)-l-1$	$\boldsymbol{\beta}'\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{\phi}\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{\beta}$
	Intercept only	$n(1-p)(t-1)-l-1$	$\boldsymbol{\beta}'\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{\phi}\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{\beta}$
5	Either	$N-rank(\boldsymbol{X}\|\|\boldsymbol{Z})$	$\boldsymbol{\beta}'\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{\phi}\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{\beta}$

m_{KR} is the denominator degrees of freedom calculated by Kenward and Roger
p is the % of data missing
N is the total number of study observations

C. Calculating power

For each of the methods presented in Chapter 2, Section II.B, the value of $f_{crit} = \mathrm{F}^{-1}(1-\alpha, l, m)$ can then be obtained and the power of the KR Wald test calculated as $1 - \mathrm{F}(f_{crit}, l, m, \omega)$.

III. Simulations and Practical Example

A. Simulations

Several simulations were conducted to determine how the methods described in Chapter 2 Section II perform in power calculations for exact and non-exact cases. All simulations were structured such that each design contained only two treatment groups with 10 subjects per group, and 25,000 trials were simulated for each design unless otherwise stated. Power was calculated only for the test of interaction between treatment and time except for in the final set of simulations in which power was calculated for the omnibus ANOVA type test of all fixed effects. Empirical power for each method was then counted as the number of times the p-value for the KR Wald test was less than $\alpha = 0.05$ divided by the total number of converging simulations performed.

For each power calculation, a full design matrix $\boldsymbol{X}_F$ was first generated and then non-baseline values were deleted from the matrix completely at random (using the same seed for all simulation sets) to obtain an unbalanced design matrix $\boldsymbol{X}_U$ with missing values. Power was then calculated using $\boldsymbol{X}_U$ with given covariance parameters. Simulations were then performed using the same $\boldsymbol{X}_U$ on which power calculations were performed. All simulations were performed on models incorporating treatment, time, and treatment by time interactions as main effects. The design matrix for the simulations can be expressed as $\boldsymbol{X} = [\boldsymbol{X}_1\ \boldsymbol{X}_2\ \ldots\ \boldsymbol{X}_N]'$ where $\boldsymbol{X_i} = [\boldsymbol{1}'\quad \boldsymbol{g_i'}\quad \boldsymbol{t_i'}\quad (\boldsymbol{gt})_i']$, $\boldsymbol{1} = [1\quad 1\quad \cdots\quad 1]$, $\boldsymbol{g_i} = [g_i\quad g_i\quad \ldots\quad g_i]$, $\boldsymbol{t_i} = [t_1\quad t_2\quad \ldots\quad t_{n_i}]$, and $(\boldsymbol{gt})_{\boldsymbol{i}} = [g_i t_1\quad g_i t_2\quad \ldots\quad g_i t_{n_i}]$ with g_i and t_j as defined in Chapter 1 section IV.A, but with only non-missing observations present. So, if the second observation for subject *i* was deleted, then t_3 in $\boldsymbol{X}_F$ became t_2 in $\boldsymbol{X}_U$ for that subject. All power calculations and simulations were performed using SAS 9.4 (Copyright © 2016 SAS Institute Inc).

Table 2.2 provides the parameterization of each simulation while Table 2.3 provides the results of each simulation. The "% Observations Missing" in Table 2.3 gives the percentage of values deleted from $\boldsymbol{X}_F$ to form $\boldsymbol{X}_U$ (number of total observations used as denominator, even though only follow-up are missing). The "Trials Converging" column in this table provides the number of simulated trials where the mixed model estimation converged. In Table 2.3, Methods 1-5 refer to the power calculated using the denominator degrees of freedom and noncentrality parameters given Table 2.1.

Table 2.2: Simulation Parameters

Design	C'	β'	σ_{00}	σ_{11}	$*r(int, slp)$	σ_e	Set	t' for X_F
1	[0 0 0 1]	[4 0.5 0.35 3.95]	4	1.15	-0.5	5.85	1a	[1 2 3 4 5]
							1b	[1 2 3 4 5]
							1c	[1 2 3 4 5]
2	[0 0 0 1]	[4 0.5 0.35 3.95]	4	N/A	N/A	5.85	2a	[1 2 3 4 5]
							2b	[1 2 3 4 5]
							2c	[1 2 3 4 5]
3	[0 0 0 1]	[4 0.5 0.35 1.85]	4	1.15	-0.5	5.85	3a	[1 2 3 4 5]
							3b	[2 5 8 14 19]
							3c	[4 7 15 22 34]
4	[0 0 0 1]	[4 0.5 0.35 3.95]	4	N/A	N/A	5.85	4a	[1 2 3 4 5]
							4b	[1 2 3 4]
							4c	[1 2 3]
5	$\begin{bmatrix}0 & 1 & 0 & 0\\0 & 0 & 1 & 0\\0 & 0 & 0 & 1\end{bmatrix}$	[4 0.5 0.35 1.65]	4	1.15	-0.5	5.85	5a	[1 2 3 4 5]
							5b	[1 2 3 4 5]
							5c	[1 2 3 4 5]

$*r(int, slp)$ refers to the correlation between random intercept and random slope

Table 2.3: Simulation Results

Design	Set	% Missing	Trials Converging	Simulated Power	Calculated Power by Method				
					1	2	3	4	5
1	1a	0	16005	0.9614	0.9693	0.9693	0.9693	0.9693	0.9800
	1b	15	14611	0.9164	0.9192	0.9284	0.9337	0.9310	0.9539
	1c	32	16092*	0.7852	0.7746	0.8227	0.8483	0.8353	0.8852
2	2a	0	24936	0.9978	0.9971	0.9971	0.9971	0.9971	0.9972
	2b	15	24811	0.9889	0.9871	0.9876	0.9879	0.9879	0.9883
	2c	32	24518	0.9446	0.9401	0.9437	0.9461	0.9464	0.9478
3	3a	20	14161	0.2788	0.3268	0.3405	0.3491	0.3437	0.3784
	3b	20	18922	0.8644	0.8691	0.8717	0.8734	0.8663	0.9008
	3c	20	19016	0.9063	0.9074	0.9079	0.9083	0.9018	0.9306
4	4a	15	24811	0.9889	0.9871	0.9876	0.9879	0.9789	0.9883
	4b	14	24585	0.8571	0.8542	0.8562	0.8576	0.8579	0.8611
	4c	12	24030	0.4916	0.4816	0.486	0.4889	0.4891	0.4964
5	5a	0	16005	0.8131	0.8118	0.8137	0.8353	0.8167	0.8945
	5b	15	14611	0.7180	0.7118	0.7288	0.7543	0.7225	0.8359
	5c	32	16092*	0.5702	0.5766	0.6119	0.6632	0.5847	0.7492

*31500 trials were simulated to maintain a large number of converging simulations

In the first series of simulations, we examined the performance of the various methods for a random intercept and slope model as the number of observations missing increased. As designed, methods 1 - 4 return the exact power when no missing data is present, and method 5 provides a close approximation. As the amount of missing data increases to 15% in design 1b, the methods diverge slightly, but all perform within 0.05 units of simulated power with methods 1 and 2 providing close estimates. As the amount

of missing data increases to 32% in design 1c, even more divergence is seen. In this case, method 1 obviously outperforms other methods, although method 2 still provides a power estimate within 0.05 units of the simulated power highlighting the impact of the adjustment to the value of the Wald test statistic made by Kenward and Roger.

The second set of simulations is almost identical to the first, except that only a random intercept was fit. The results are good for all methods regardless of missing data prevalence. Additionally, more simulations were able to converge due to not having to estimate slope variance parameters. The results suggest that for random intercept only models (identical to compound symmetric covariance), simple methods may be adequate for power calculations.

In the third set of simulations, the time vector was allowed to vary as something other than $\boldsymbol{t} = [1\ \ 2\ \ 3\ \ 4\ \ 5]'$. The parameters were set the same as those in the first set of simulations except the values of $\boldsymbol{\beta}$ were reduced to keep power from converging to 1 with larger time values. The missing pattern of follow-up observation was held constant with only the time vector allowed to vary. Here, unequal spacing in time seemed to have less impact on each method's performance than did the magnitude of the vector. All methods performed poorest at the lower power values in 3a, which could indicate lower powered tests are more difficult to approximate due to the nonlinear nature of power as a function of the noncentrality parameter. However, even though the methods were less accurate for low power, method 1 still outperforms other methods and maintains accuracy within 0.05 units of power.

The fourth set of simulations has the same parameter values as the second and again only has a random intercept. However, the time vector was allowed to be reduced

from $\boldsymbol{t} = [1\ \ 2\ \ 3\ \ 4\ \ 5]'$ by dropping the number of follow-up observations in $\boldsymbol{X}_F$. The results here echo those of scenario 2 and again suggest the simple approaches perform as well as the more complicated methods in the random intercept only model. Moreover, the number of follow-up observations was found to have little impact on the performance in this method. Unfortunately, a similar experiment evaluating a low number of follow-up observations would be difficult to conduct in a model with random slope because the estimate of slope variance would be difficult to obtain and lead to convergence issues.

The fifth and final set of simulations is almost identical to that of the first, except that now we are interested in the full omnibus F test that $\beta_{\text{group}} = \beta_{\text{time}} = \beta_{\text{group}*\text{time}} = 0$. The effect sizes have likewise been reduced to prevent power from being too close to 1. The results demonstrate that with little missing data, all methods perform well. However, as missingness increases, the degree of performance diverges. With 32% missing, Method 2 performs within a 0.05 unit margin but not as well as Method 1, which demonstrates the superiority of the KR type method based on the Taylor expansion to solve for the noncentrality parameter in these more complicated scenarios. Interestingly, method 4 performs closest to method 1 no matter the amount missing. This proximity arises from the fact that the KR method often shrinks the denominator degrees of freedom only by a small amount and leaves much of the heavy lifting in the F approximation to the adjustment of the Wald statistic itself. Conversely, method 4 makes a comparatively large adjustment to the denominator degrees of freedom that compensates for the lack of adjustment to the Wald statistic and therefore noncentrality parameter.

Altogether the results of these simulations suggest methods 1 and 2 are capable of providing accurate power approximations in the small sample cases for which the

Kenward-Roger test statistic is critical. Moreover, the expansion-based method 1 outperformed simpler method 2 in random intercept and slope models, particularly when the observational pattern became more and more imbalanced due to unobserved follow-up observations. However, the results also suggest that in studies with low anticipated dropout or simple covariance structures, a variety of simple methods - including ones that use residual degrees of freedom and traditional values of the noncentrality parameter for the F distribution of the Wald-test statistic under the alternative hypothesis – are sufficient for calculating power.

B. Practical Example

Wolfinger (1996) examined a study on growth curve data of rats originally presented in Box (1950). In this study, rats were provided one of three treatments: Control, Thyroxin, or Thiouracil. The body weight growth curves of all rats were roughly linear. Of interest: is there a difference in average trajectory of body weight over time for either treatment group compared to the control group? The difference observed in the study was quite large, so for this example we will determine how many rats would be required to detect a smaller difference in trajectory given the variance parameters observed in the study. The model for this example can be specified as follows:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{e}$$

With $\boldsymbol{Y} = [\boldsymbol{Y}_1\ \boldsymbol{Y}_2\ \ldots\ \boldsymbol{Y}_N]'$ and as in chapter 1:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$$

In this example there are 3 treatment groups, and so

$$\boldsymbol{X}_i = \begin{bmatrix} 1 & \delta_i & \psi_i & t_1 & \delta_i t_1 & \psi_i t_1 \\ 1 & \delta_i & \psi_i & t_2 & \delta_i t_2 & \psi_i t_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \delta_i & \psi_i & t_{n_i} & \delta_i t_{n_i} & \psi_i t_{n_i} \end{bmatrix}$$

where δ_i= 1 if rat *i* is on Thyroxin and 0 otherwise, ψ_i= 1 if rat *i* is on Thiouracil and 0 otherwise, and t_j is the time the j^{th} body weight measurement for rat *i* is observed. In the full balanced design, the time vector for each rat will be $t = [t_1 \ t_2 \ \ldots \ t_5]' = [0 \ 1 \ 2 \ 3 \ 4]'$. $\boldsymbol{\beta}$ is the 6 x 1 vector of fixed effect parameters to be estimated given by $\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \beta_2 \ \beta_3 \ \beta_4 \ \beta_5]'$, where β_0 is the mean body weight intercept for rats on control, $\beta_1$is the mean difference in body weight intercept between rats on Thyroxin and rats on control, β_2 is the mean difference in body weight intercept between rats on Thiouracil and rats on control, β_3 is the mean change in body weight per week for rats on control, β_4 is the mean difference in the amount the body weight changes per week between rats on Thyroxin and rats on control, and β_5 is the mean difference in the amount the body weight changes per week between rats on Thiouracil and rats on control. Specifically, in this example we will have:

$$\boldsymbol{\beta} = [52.88 \quad 4.82 \quad -1.08 \quad 26.48 \quad -6.43 \quad 1.0914]'$$

Both intercept and slope will be included as random effects with:

$$\boldsymbol{\Sigma_{bi}} = \begin{bmatrix} 31.6315 & -2.5103 \\ -2.5103 & 15.1184 \end{bmatrix} \qquad \boldsymbol{\Sigma_{ei}} = 18.8556 \cdot \mathbf{I}_{n_i x n_i}$$

As mentioned, of primary interest will be the omnibus hypothesis: is the average change in body weight over time different in either treatment group compared to the control group? The contrast matrix will then be given as:

$$C' = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

We can now calculate power for a trial using these estimates where the observation pattern of the rats is planned to be unbalanced. Specifically, we plan for the rats to have the observation pattern specified in Table 2.4, which shows the planned time of observation for each of the six rats allocated to each treatment group.

Table 2.4: Observation patterns for planned trial

Treatment	Rat	Observation Times
Control	1	0,1,2,3,4
	2	0,1,2,3,4
	3	0,1,2,3,4
	4	0,2,3,4
	5	0,1,3,4
	6	0,1,4
Thyroxin	1	0,1,2,3,4
	2	0,1,2,3,4
	3	0,1,2,3,4
	4	0,1,2,3
	5	0,1,2,3
	6	0,2,3
Thiouracil	1	0,1,2,3,4
	2	0,1,2,3,4
	3	0,2,3,4
	4	0,1,2,4
	5	0,1,2,4
	6	0,1,2,4

We calculated power for each of the 5 methods presented in Chapter 2 Section II.B and compared them to empirical power obtained from 75,000 simulated trials. The results are presented in Table 2.5 and show that the methods using the KR degrees of freedom (methods 1-3) perform best, while the simpler methods 4 and 5 tend to under or overestimate power. However, the estimates provided by method 4 and 5 are not terribly far off from the simulated value, which agrees with the results in Table 2.3 where all

methods tended to perform well in the presence of low to moderate amounts of missing data.

Table 2.5: Simulated empirical vs calculated power for Rats example

Design	% Missing	Trials Converging	Simulated Power	Calculated Power by Method				
				1	2	3	4	5
rats	13	73811	0.7767	0.7738	0.7765	0.7770	0.7593	0.7865

IV. Summary and Discussion

In this chapter, we developed a novel approach to calculating power for the KR adjusted Wald-test that extends the approach of Kenward and Roger from the null to the alternative hypothesis. We also introduced three additional intuitive approaches. These four new approaches, as well as one existing approach based on the residual degrees of freedom, were then used to calculate power for multiple designs. These simulations showed that most of the methods work well, even in small sample studies, provided the degree to which a design is unbalanced is not too large.

Moreover, for random intercept only models, all models provide accurate and similar estimates of power indicating the simpler approaches are adequate in such scenarios. For models with both random intercept and random slope, method 5 consistently overestimated power – likely due in part to the larger degrees of freedom used by this method. While method 4 tended to perform well in many cases, this method has the tendency to underestimate power by a noticeable amount (for instance, in the practical example analyzing rat body weights). Additionally, method 4 may not perform well in designs where subjects have large variations in observational pattern (such as in observational studies where subjects may have very different vectors of observation times), but this potential limitation has not been explored.

Ultimately, method 1 tended to most accurately approximate power for these designs, especially when the observation patterns were highly unbalanced. The high degree of accuracy of this method and others suggests further exploration and the potential extension to the calculation of power, and perhaps sample size, for designs with some anticipated rate of missing data.

CHAPTER 3
PLANNING FOR INCOMPLETE DATA IN POWER CALCULATIONS

I. Introduction

In chapter 2 we demonstrated that power for the KR test statistic in the LMM can be calculated when the observed data pattern is fixed. Calculating power for a more general rate of missingness (i.e. when each observation has some probability of being missing) is less straightforward, but the ability to do so would be desirable in practice. For instance, researchers may desire to calculate power assuming each follow-up observation has a 10% chance of being missing at random. Planning for incomplete data in power calculations for non-longitudinal designs is often rather straightforward: the values in the design matrix are often uniform within subject (for instance group or cluster), and so missing data changes the noncentrality parameter in a predictable way. Conversely, calculating power for such cases in longitudinal studies is particularly challenging, as not every missing data point is of equal importance. For example, Basagaña and Spiegelman (2012) note that timepoints in the middle of a longitudinal study with linear outcome trajectory have comparatively little impact on power compared to beginning and ending timepoints, and so the impact of missing data is not obvious and depends on which observation is missing.

A variety of literature exists incorporating incomplete longitudinal data into power calculations for linear mixed model Wald tests. Most approaches rely either on

asymptotic convergence between *t* and *z* distributions or on multivariate theory and its frequent correspondence with LMM hypotheses. Some methods look at calculating power while accommodating data that are planned to be monotonically missing due to subject dropout, while others plan for data that are assumed to be missing in some general way. We will briefly summarize a few of these approaches as well as highlight the limitations in applying these methods to the KR adjusted Wald test.

II. Existing methods planning for missing data in power calculations for LMMs

Hedecker, Gibbons, and Waterneaux (1999) provide power calculations for the test of group by time interaction when a particular rate of missingness at each follow-up visit is specified. This method simply modified existing power calculations by adjusting covariance and noncentrality parameters based on the expected number of observations at each time point. For instance, they approximate the distribution of the Wald statistic testing the group by time interaction (or other single contrast tests of fixed effects) in the balanced linear mixed model as $\sqrt{W} \sim N\left(\sqrt{W}, 1\right)$. If each time point has an equal probability of missing, *p*, then the Wald statistic would instead follow the distribution $\sqrt{W} \sim N\left(\sqrt{pW}, 1\right)$. This approximation is simple and intuitive, but unfortunately cannot be utilized in the KR adjusted Wald statistic for two primary reasons. First, as noted in the paper, approximating a *t* distribution with a *z* distribution works best when the denominator degrees of freedom in the *t* distribution is at least 30, which will often not be feasible in small sample studies for which the KR adjustment is most necessary. Second, this scaling underrepresents the true amount by which the KR Wald statistic will decrease due to missing data. Specifically in the context of the KR Wald Statistic, when data are

balanced $\mathbf{A} = \mathbf{0}$ (definition of $\mathbf{A}$ given in Appendix II part A), and so $\boldsymbol{\phi} = \boldsymbol{\phi}_{\mathrm{A}}$ and $W = W_A$. Conversely, when data are unbalanced, such as when some data are missing, $\mathbf{A} \neq \mathbf{0}$ and $W \neq W_A$. Therefore, using pW as the noncentrality parameter will ignore the additional reduction of the Wald statistic caused by the elements of $\mathbf{A}$ being nonzero due to imbalanced data.

Wang, Hall, and Kim (2012) and Zhao and Edland (2021) provide power calculations for the test of group by time interaction in the case of monotone missingness during follow-up. Specifically, they obtain $\boldsymbol{\phi}$ as calculated over the $\tau - 1$ observation patterns, with the number of subjects having each determined by dropout probability *p*. They again rely on asymptotic normality of the Wald statistic, however, and so this approach likewise avoids the issue of calculating degrees of freedom needed in smaller sample sizes. Interestingly, as pointed out by Zhao and Edland, under these approaches the power for the test of group by time interaction will be unaffected by the variance of the random intercept making specifying this parameter unnecessary. While true in balanced cases, this fact will not hold for the KR adjusted test in unbalanced designs as the variance parameters will impact the calculations of the denominator degrees of freedom.

Verbeke and Lesaffre (1999) take a similar, but still unique approach in attempting to calculate power while accommodating data missing in a monotone fashion. Specifically, as also described in Galbraith (2002), they let there be τ observations per subject when no dropout is present. Let $\boldsymbol{P} = (p_1, \ldots, p_\tau)$ with p_j be the probability a patient's data is missing after time *j*, and let $\boldsymbol{n} = (n_1, \ldots, n_\tau)$ be the number of patients

who have observations through time *j*. Then power for the LMM Wald test statistic can be calculated using the noncentrality parameter given as

$$\omega = \boldsymbol{\beta}'\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{\Phi}\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{\beta}$$

with

$$\boldsymbol{\Phi} = \left(\sum_{j=1}^{\tau} n_j \boldsymbol{X}'_{(j)} \boldsymbol{\Sigma}^{-1}_{Y_{(j)}} \boldsymbol{X}_{(j)}\right)^{-1}$$

where $\boldsymbol{X}_{(j)}$ and $\boldsymbol{\Sigma}_{Y_{(j)}}$ refers to the $(jn_j)x4$ design and $(jn_j)x(jn_j)$ covariance matrix consisting of patients dropping out after time *j,* and again τ is the maximum number of observations per subject. This expression is provided in a more general format at the end of section 3 of Verbeke and Lesaffre but can be easily shown to take the above form in the case of monotone missingness. The numerator degrees of freedom are chosen in the usual way, and the denominator degrees of freedom are equal to *N – rank(**X**)* as in Helms (1992) (although Galbraith simply uses a *z* approximation to the t statistic).

The observed data $\boldsymbol{n}$ can then be described as a sample drawn from the multinomial distribution given as $multi(n, p_1, \ldots, p_\tau)$. As such, the power for the test in the presence of such dropout is itself a random variable depending on the specific sample of $\boldsymbol{n}$ observed. Verbeke and Lesaffre then calculate power for 1000 different such samples given the model parameters and dropout probabilities to construct an empirical distribution of P(Power ≤ power) for the model power over the possible dropout patterns. This empirical distribution can then be used to evaluate the ability of a design to actually achieve the desired power for the design given some anticipated sample size and rate of

dropout. They also argue this method provides a more holistic criteria (as opposed to a single value such as mean or median power) by which to compare the power of two designs (for instance comparing a design with 5 vs 7 observations). However, this method may be less intuitive to investigators used to working with a single value for power, and it also requires sampling from a full design a large number of times (as mentioned, in this case 1000).

Galbraith uses this same sampling strategy, but instead calculates expected power (focusing on group by time interactions only) rather than focusing on the empirical distribution of power. Specifically, they obtain expected power (given as E[Power($\boldsymbol{n}$)])by averaging calculated power over an (unspecified, but likely 1000 given their reference to Verbeke and Lesaffre) number of samples of observed $\boldsymbol{n}$. Moreover, they investigate the ability to approximate E[Power($\boldsymbol{n}$)] with Power[E($\boldsymbol{n}$)] and E($\boldsymbol{n}$) = $(np_1, \dots, np_\tau)$ where these values are then utilized in the expression of $\boldsymbol{\Phi}$ to calculate the noncentrality parameter and power (and is therefore similar to the approach of Wang, Hall, and Kim (2012) and Zhao and Edland (2021)). They find that, in general, this approximation tends to overestimate E[Power($\boldsymbol{n}$)], but in general these values tend to be similar. Unfortunately, a similar approximation is not feasible for our goal. E($\boldsymbol{n}$) can be easily specified in the case of monotone missingness because only $\tau - 1$ observation patterns are possible for a subject, each with an easily described probability p_j of being observed. Conversely, in the case of general missingness (i.e. each follow-up observation has probability p of being missing) enumerating all the possible observation patterns and their probability of occurrence for a subject becomes overwhelming as the number of repeated measures grows (see Tu et al. 2007 equation 23 and discussion following

equation 26). Additionally, these methods have not, as specifically mentioned by Verbeke and Lesaffre, been designed to calculate power for the KR adjusted Wald-test although they provide a promising framework to extend upon.

Only Ringham et al. (2015) and Josey et al. (2021) provide methods to accommodate some general rate of missingness in calculating (expected) power for a test statistic closely associated with the KR statistic. The roots of this method are grounded in the theory for the GLMM -particularly in the methods developed by Muller and Peterson (1984) and O'Brien and Shieh (1992) to calculate power for the McKeon adjusted HLT statistic, and in methods developed by Catellier and Muller (2000) to further adjust the degrees of freedom in the HLT test with missing data. This method has the added benefit that expected power can be directly calculated rather than being averaged over different sample observation patterns, which in turn reduces computation time.

Specifically, they adjust the degrees of freedom and noncentrality parameter based on the expected number of subjects with complete sets of observations. Treating the probability of each of τ observations being missing as p, then the expected number of n subjects with complete observations is easily obtained as $n_c = n(1-p)^{\tau}$. From Chapter 1 section II.C we have that for single contrast tests of linear trend, the HLT statistic $\left(\frac{n-q}{1} tr[\boldsymbol{HE}^{-1}]\right)$ follows an $F_{1,\ n-q,\ \omega}$ distribution. The method described by Ringham et al. simply replaces the term $n-q$ with $n_c - q$ as the denominator degrees of freedom and replaces ω with $\frac{n_c}{n}\omega$ as the noncentrality parameter to obtain the distribution of the HLT statistic under the alternative hypothesis from which expected power can be calculated for the HLT test with planned rate of missingness p.

Simulations were performed by Ringham et al. to evaluate the performance of the method described. Small to large sample sizes were considered, with either 3 or 6 observations per subject and with missingness probability set as either 0.05 or 0.10. Results were good for designs with 48 or more subjects and 3 observations per subject. However, the method tended to perform worse for designs having 6 observations per subject or fewer than 48 subjects, which is unsurprising as the noncentrality parameter shrinks exponentially with the number of follow-up observations. Additionally, this method only calculates expected power for a close analogue of the KR test, and so it's ability to calculate power for the KR test itself given some missingness rate (especially when that rate is larger than 0.1) is unknown. Lastly, this approach operates on the assumption power is linear around ω such that $\mathrm{E}[\mathrm{Power}|p] \approx \mathrm{Power}[\mathrm{E}(\omega|p)]$, which is not always the case.

III. Calculating Expected Power for the KR Test with Anticipated Rate of Missing Data

A. Expected Power

We propose the simplest solution to the limitations presented in calculating expected power for the KR statistic accommodating some anticipated rate of missing data is to average the power calculated for a number of designs randomly generated according to the specified probability of any follow-up observation being missing. For instance, assume we have a full balanced design matrix $\boldsymbol{X}$ (which we can denote as $\boldsymbol{X}_F$ to indicate all subjects have a full set of τ observations), for the planned study such that for all subjects, for example with an observation per week for 12 weeks

$$\boldsymbol{X}_{fi} = \begin{bmatrix} 1 & \delta_i & 1 & \delta_i 1 \\ 1 & \delta_i & 2 & \delta_i 2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \delta_i & 11 & \delta_i 11 \\ 1 & \delta_i & 12 & \delta_i 12 \end{bmatrix}$$

We then may expect that in an actual trial, each follow-up observation for each subject may have a *p* probability of being missing (a more sophisticated missing data process generated by a specified Markov process is described by Tu et al., but we believe the stated approach is sufficient for typical planning needs). Thus, each subject has an observed design matrix with possibly missing data that we can denote as $\boldsymbol{X}_{ui}$ to represent that data across subjects may be unbalanced. So, for subject 1 we may have a design matrix with

$$\boldsymbol{X}_{u1} = \begin{bmatrix} 1 & \delta_1 & 1 & \delta_1 1 \\ 1 & \delta_1 & 3 & \delta_1 3 \\ 1 & \delta_1 & 4 & \delta_1 4 \\ 1 & \delta_1 & 8 & \delta_1 8 \\ 1 & \delta_1 & 10 & \delta_1 10 \\ 1 & \delta_1 & 12 & \delta_1 12 \end{bmatrix}$$

and, for subject 2 we may have a design matrix with

$$\boldsymbol{X}_{u2} = \begin{bmatrix} 1 & \delta_2 & 1 & \delta_2 1 \\ 1 & \delta_2 & 4 & \delta_2 4 \\ 1 & \delta_2 & 7 & \delta_2 7 \\ 1 & \delta_2 & 9 & \delta_2 9 \\ 1 & \delta_2 & 10 & \delta_2 10 \end{bmatrix}$$

Then if we have *n* subjects in the model, the full observed design matrix $\boldsymbol{X}_U$ is given as

$$\boldsymbol{X}_U = [\boldsymbol{X}_{u1}\ \boldsymbol{X}_{u2}\ \cdots \boldsymbol{X}_{un}]'$$

From any full design matrix $\boldsymbol{X}_F$ there are *Q* possible $\boldsymbol{X}_U$ which could be observed. To calculate expected power for the candidate study given probability *p* of any follow-up

observation being missing $\left(\mathrm{E}(power|p)\right)$, we can average the power over all Q values of $\boldsymbol{X}_U$ weighted by their probability of appearing. Such an approach would be similar to that proposed by Zhao and Edland (2021) and Galbraith (2002) but would naturally calculate power for the KR statistic specifically. Since averaging over all designs (much less determining the probability of each being observed) is not always feasible, the natural question to such an approach is: how many different designs do we need to generate and average power over to get a good estimate of the expected power of the study given some anticipated rate of missing data?

B. A Heuristic Approach to Determining Necessary Computational Complexity

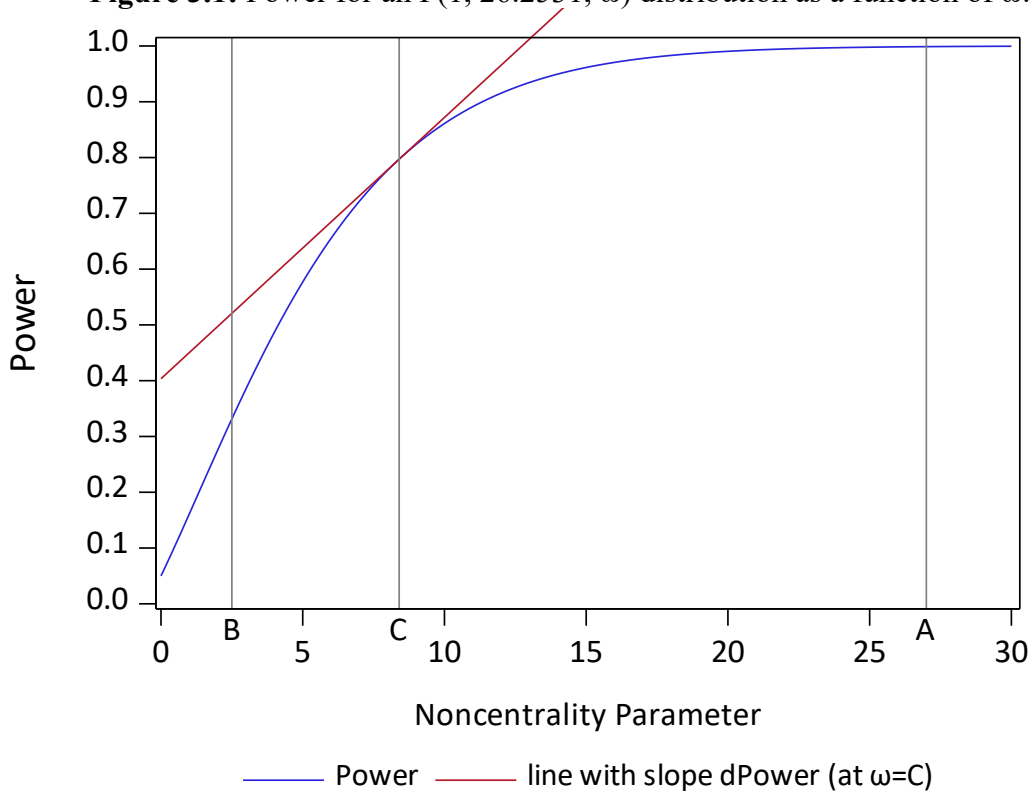
We will develop a heuristic approach to determine whether we have averaged over enough designs with missing data to provide an accurate estimate of expected power given some rate of missing follow-up observations. The approach will rely on the relationship between the power function and noncentrality parameter, as well as on the natural way in which the KR test statistic changes as data becomes more unbalanced. We will then use simulations to evaluate the general ability of the method developed to provide accurate power calculations. The results in Chapter 2 suggest that out of the five methods examined to calculate power for the KR adjusted Wald-test, method 1, which calculating power by extending the methodology of Kenward and Roger to the alternative hypothesis, performed best. Therefore, we will initially focus on calculating expected power with some probability of any follow-up observation being missing by using an extension of method 1. However, the methodology developed will also be applicable to methods 2-5, and so these methods will again be compared in section III of this chapter.

Test power as a function of the noncentrality parameter is well known to be nonlinear. For instance, in Figure 3.1, as hypothesized treatment effect size tends to infinity, the power of the test to reject the null hypothesis of no difference in effect of time by treatment approaches 100% asymptotically. After a certain point, a change in effect size will result in a comparatively negligible change in power. Intuitively this relationship means that, given anticipated missingness probability p of each planned follow-up observation in the study, if the expected value of the noncentrality parameter (which we can denote as $E(\omega|p)$) for a test statistic is high, power calculated for most possible observation patterns is unlikely to deviate substantially from the expected power of the study. For instance, if $E(\omega|p)$ = A in Figure 3.1, then observed power for the study will be similar for most $\boldsymbol{X}_U$. Conversely, if $E(\omega|p)$ is located at a point where the power curve changes substantially with the noncentrality parameter, then differences in observation pattern can have a big impact on calculated power. For instance, if $E(\omega|p)$ = B in Figure 3.1, then observed power for the study may vary greatly depending on $\boldsymbol{X}_U$.

As such, one of the components that will help us determine how many generated $\boldsymbol{X}_U$ we need to average over to estimate expected study power is the derivative of the power function with respect to the noncentrality parameter around $E(\omega|p)$ (we will call this value *dPower*). If in a study we have $E(\omega|p)$ = C as in Figure 3.1, then *dPower* is equal to the slope of the line tangent to the power function at C. The quantity *dPower* then provides us with some idea of how many $\boldsymbol{X}_M$ we need to average over to provide a stable estimate of $E(power|p)$. Values of *dPower* would range from 0 to ∞, with a small value of *dPower* indicating only a few number $\boldsymbol{X}_U$ are needed as each would provide a similar power value, while a large value of *dPower* would indicates more $\boldsymbol{X}_U$ should be

averaged over as each may provide widely different power values. An algorithm to calculate *dPower* is provided in Appendix III.

Figure 3.1: Power for an F(1, 26.2531, ω) distribution as a function of ω.



Of course, having study power be sensitive to changes in the noncentrality parameter means little if the noncentrality parameter will change by only a very small amount from one choice of $\boldsymbol{X}_U$ to another. For instance, if again E($\omega|p$) = C as in Figure 3.1 and ω_U only varies by a range of 0.01 between most probable (given p) values of $\boldsymbol{X}_U$, power calculated for each choice of $\boldsymbol{X}_U$ should still provide a similar estimate of E($\omega|p$) despite the high value of *dPower* (compared to what it would be at point A, for instance). One example of such a scenario may arise in testing the group by time interaction with a small effect size but large number of time points. Note the expression for ω provided in Chapter 2 Table 2.1 for method 1, which shall be used to obtain the value of the noncentrality parameter in this chapter, depends on the value $\boldsymbol{\beta'\theta A\theta\beta}$. The elements of **A**

become increasingly nonzero as data become more imbalanced, and so we can intuitively view $\boldsymbol{\beta'\theta A\theta\beta}$ as reflecting the magnitude by which ω is affected by the variability in observational pattern between subjects. Values of $\boldsymbol{\beta'\theta A\theta\beta}$ would range from 0 to ∞, with larger average values of $\boldsymbol{\beta'\theta A\theta\beta}$ indicating a higher degree of variability in the noncentrality parameter depending on choice of $\boldsymbol{X}_U$, whereas smaller average values indicate the noncentrality parameter does not vary much from one choice of $\boldsymbol{X}_U$ to another.

These two quantities, *dPower* and $\boldsymbol{\beta'\theta A\theta\beta}$, taken together provide us with some idea of both how much the noncentrality parameter will change from one choice of $\boldsymbol{X}_U$ to another and how much the calculated power will change with the noncentrality parameter. We can combine these two components into a single quantity "G" to provide an intuitive measure of whether a small or large number of $\boldsymbol{X}_U$ need to be averaged over so that

$$\widehat{\mathrm{E}}(power|p)_K = \frac{1}{K}\sum_{i=1}^{K} Power\left(\omega|\boldsymbol{X}_{U_i}\right)$$

(with K being the number of $\boldsymbol{X}_U$ over which power is averaged) provides a stable estimate of E(*power*|*p*). Specifically, for the sake of computational speed we can calculate $\widehat{\mathrm{E}}(G|p)$ as the average of 10 G_U each calculated from a randomly generated $\boldsymbol{X}_U$ such that

$$\widehat{\mathrm{E}}(G|p) = \frac{1}{10}\sum_{U=1}^{10} G_U = \frac{1}{10}\sum_{U=1}^{10}[1000 \cdot dPower_U \cdot (\boldsymbol{\beta'\theta A\theta\beta})_U]$$

A small value of $\widehat{\mathrm{E}}(G|p)$ would indicate that $\widehat{\mathrm{E}}(power|p)_K$ calculated over only a small number K of $\boldsymbol{X}_U$ should provide a stable estimate of $\mathrm{E}(power|p)$, whereas a larger value

of $\widehat{\mathrm{E}}(G|p)$ would indicate a large number of $\boldsymbol{X}_U$ need to be considered. Note the quantity $dPower_U \cdot (\boldsymbol{\beta}'\boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta}\boldsymbol{\beta})_U$ is multiplied by 1000 just to keep the value of G_U from being too small. Additionally, the product between $dPower_U \cdot (\boldsymbol{\beta}'\boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta}\boldsymbol{\beta})_U$ was chosen instead of their sum to make each term equally impactful (for instance, a large value of $dPower_U$ and a small value of $(\boldsymbol{\beta}'\boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta}\boldsymbol{\beta})_U$ would negate each other to yield a moderate value of G_U).

A simulation study was performed to help evaluate what values of $\widehat{\mathrm{E}}(G|p)$ may indicate only a small number of $\boldsymbol{X}_U$ need to be generated. 2100 designs of the form specified in Chapter 1 Section IV.A were created by randomly generating population parameter values with the conditions that all model standard deviations (within subjects, random intercept, and random slope) were less than 100 with $\rho(\sigma_{11}, \sigma_{22}) \in (0, 0.5]$, and that the group by time interaction effect parameter β_3 be no more than 2.95 times the standard deviation of the random slope in order to keep power from being too large. Only models with both random intercept and slope were generated since the results from Chapter 2 Section III suggest power for random intercept only models can be accurately calculated with simple methods (e.g. those described by Galbraith or Zhao and Edland but for a general missingness pattern). The values of time at each observation were allowed to vary between 1 and 75, each subsequent time value being greater than the last, with the number of observations varying between 3 and 12. Lastly, each design was assigned a probability, $p \in [0.05, 0.5]$, of any given follow-up observation being missing. Model parameters β_0, β_1, and $\beta_2$were fixed at 4, 0.9, and 3.45 respectively with only the group by time effect β_3 being randomly generated as only this effect was examined. As an example, Table 3.1 provides the first 5 of 2100 designs generated.

Table 3.1: Sample of Generated Designs

Design	σ_e	σ_{11}	σ_{22}	$\rho(\sigma_{11},\sigma_{22})$	$\boldsymbol{t}'$ for $\boldsymbol{X}_F$	1-p	β_3
1	1.70284	0.83074	0.48964	0.24706	[1 6 9 14 25 29 43 52 59 65 72]	0.94	0.57702
2	3.12824	6.48894	1.24757	0.19173	[6 15 28 29]	0.91	0.48217
3	73.74644	4.37835	13.18203	0.17312	[6 10 14 17 19]	0.87	3.14720
4	14.68027	2.78509	2.42391	0.30144	[2 5 11 13 22 25 28 32 40]	0.66	1.15831
5	1.40316	0.38730	0.26408	0.43773	[1 16 25 30 32]	0.63	0.74793

From each design, 200 different observational patterns $\boldsymbol{X}_U$ were randomly generated according to p, with the seed generating these $\boldsymbol{X}_U$ changing for each design. $\widehat{\mathrm{E}}(G|p)$ was then calculated over the first 10 $\boldsymbol{X}_U$. Expected power was then calculated by averaging power calculated by method 1 over the first 10 $\boldsymbol{X}_U$, which will be denoted as $\widehat{\mathrm{E}}(power|p)_{10}$, as well as over 25, 50, 100, and 200 $\boldsymbol{X}_U$, which will be denoted as $\widehat{\mathrm{E}}(power|p)_K$ with K=25, 50, 100, or 200. Finally, we examined the relationship between values of G and the average absolute difference between $\widehat{\mathrm{E}}(power|p)_{10}$ and $\widehat{\mathrm{E}}(power|p)_K$ given as:

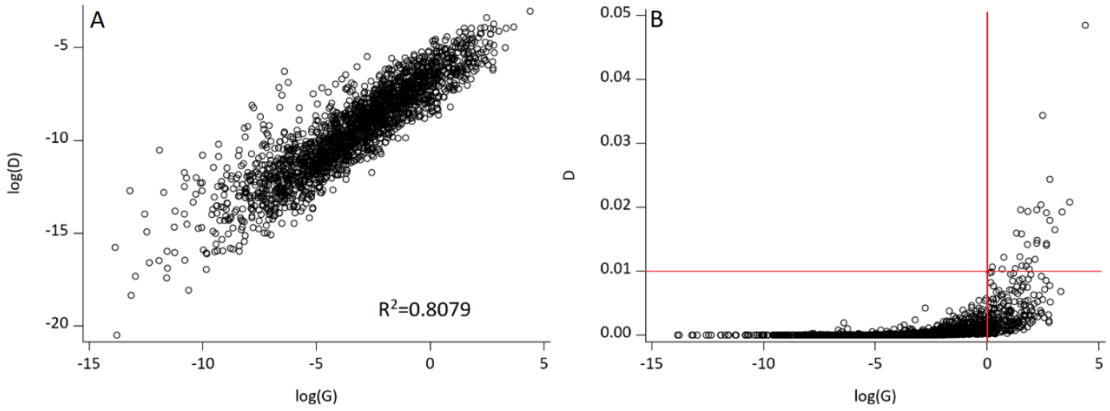
$$D = \frac{1}{4}\sum_{K=25}^{200} \left|\widehat{\mathrm{E}}(power|p)_{10} - \widehat{\mathrm{E}}(power|p)_K\right|$$

One value of $\widehat{\mathrm{E}}(G|p)$ was obtained as -2.64192E-12 due to SAS overflow issues and was excluded from the study for having a negative value. The results, shown in Figure 2, show a rather straightforward association between $\widehat{\mathrm{E}}(G|p)$ and the consistency between $\widehat{\mathrm{E}}(power|p)_{10}$ and $\widehat{\mathrm{E}}(power|p)_{K\epsilon\{25,50,100,200\}}$ as estimators of E($power|p$).

As seen in Figure 3.2.A, there is a clear relationship on the logarithmic scale between $\widehat{\mathrm{E}}(G|p)$ (simply given as log(G) in the figure) and the average absolute difference between $\widehat{\mathrm{E}}(power|p)_{10}$ and $\widehat{\mathrm{E}}(power|p)_{K\epsilon\{25,50,100,200\}}$ with lower values of $\widehat{\mathrm{E}}(G|p)$ being associated with a lower difference. Additionally, Figure 3.2.B shows that no meaningful difference (in this case a difference greater than 0.01) exists between the

two power estimates for $\widehat{\mathrm{E}}(G|p)$ values less than 1. These results hold even stronger with an R^2 value of 0.8497 when $\widehat{\mathrm{E}}(G|p)$ is calculated over 200 different $\boldsymbol{X}_U$ instead of 10. The results presented in Figure 3.2.A suggest that when $\widehat{\mathrm{E}}(G|p)$ is less than 1, only a few different $\boldsymbol{X}_U$ (for instance 10 or 25) need to be generated in order to obtain a stable estimate of E(*power*|*p*), which means that in such cases an estimate can be calculated quickly and require few computational resources.

Figure 3.2: Average deviation between $\widehat{\mathrm{E}}(power|p)_{10}$ and $\widehat{\mathrm{E}}(power|p)_{K\epsilon\{25,50,100,200\}}$



C. Simulation Performance

A subsequent simulation study was performed to evaluate the accuracy of $\widehat{\mathrm{E}}(power|p)_{10}$ and $\widehat{\mathrm{E}}(power|p)_{200}$ as estimators of E(*power*|*p*). Specifically, 400 designs were randomly selected from the initial 2100 mentioned in section z, with the condition that 200 have $\widehat{\mathrm{E}}(G|p)$ values greater than 1 and 200 have $\widehat{\mathrm{E}}(G|p)$ values less than 1. For each design selected, 20000 trials were simulated, and follow-up observations were deleted from the trial at random according to specified probability *p* for each design with seed generating outcome values and missing observations changing with each design. Empirical expected power, $\widetilde{\mathrm{E}}(power|p)$, for each design was then calculated as

the number of times the null hypothesis was rejected (with α = 0.05 in this study). Results of the simulation study are shown in Figure 3.3 and Table 3.2.

Figure 3.3: Calculated for Simulated power for 2100 designs

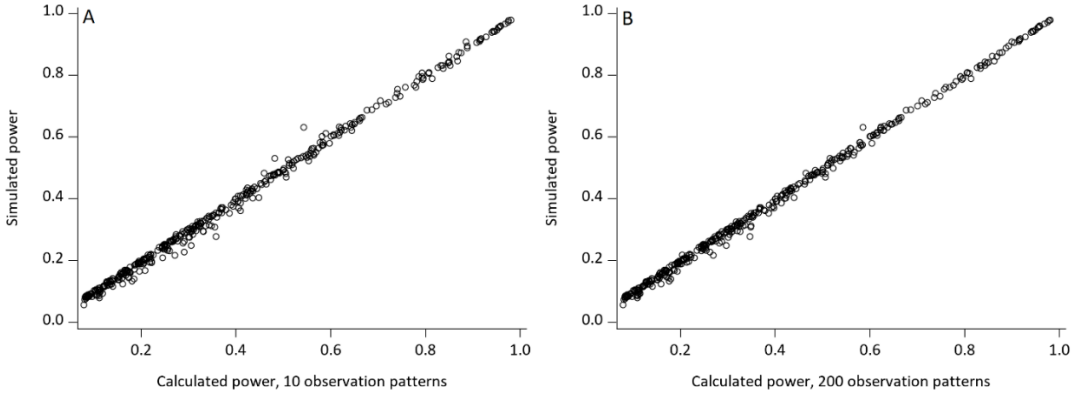


Table 3.2: Calculated vs Empirical power

Number of different observational patterns	$\widehat{\mathrm{E}}(G\|p)$ value	Regression Estimates and R^2 for model: *simulated power = β*calculated power*		Absolute difference, calculated vs simulated power			
		$\hat{\beta}$	R^2	Min	Median	Mean	Max
10 different $\boldsymbol{X}_U$	Overall	0.99139	0.9990	0.000038	0.0044	0.0089	0.090
	G <1	0.99710	0.9998	0.000070	0.0026	0.0037	0.025
	G ≥1	0.98699	0.9984	0.000038	0.0100	0.0142	0.090
200 different $\boldsymbol{X}_U$	Overall	0.99248	0.9994	0.000014	0.0040	0.0073	0.067
	G <1	0.99755	0.9998	0.000074	0.0028	0.0037	0.023
	G ≥1	0.98855	0.9991	0.000014	0.0074	0.0109	0.067

Figure 3.3 shows the high degree of concordance between empirical and calculated power, regardless of the number of $\boldsymbol{X}_U$ used in the power calcluations. Table 3.2 further clarifies the strength of this relationship by showing the results of the regression model $\widetilde{\mathrm{E}}(power|p) = \widehat{\mathrm{E}}(power|p)_{10}$ with K=10 or 200. $\widehat{\mathrm{E}}(power|p)_K$ provided accurate power estimates regardless of the number of $\boldsymbol{X}_U$ used. However, using 200 $\boldsymbol{X}_U$ provided more accurate estimates than using 10, especially when values of $\widehat{\mathrm{E}}(G|p)$ were greater than 1. Importantly, these results suggest that power can be

calculated accurately, and when values of G are less than one, power can be calculated quickly.

IV. Comparison With Other Chapter 2 Methods

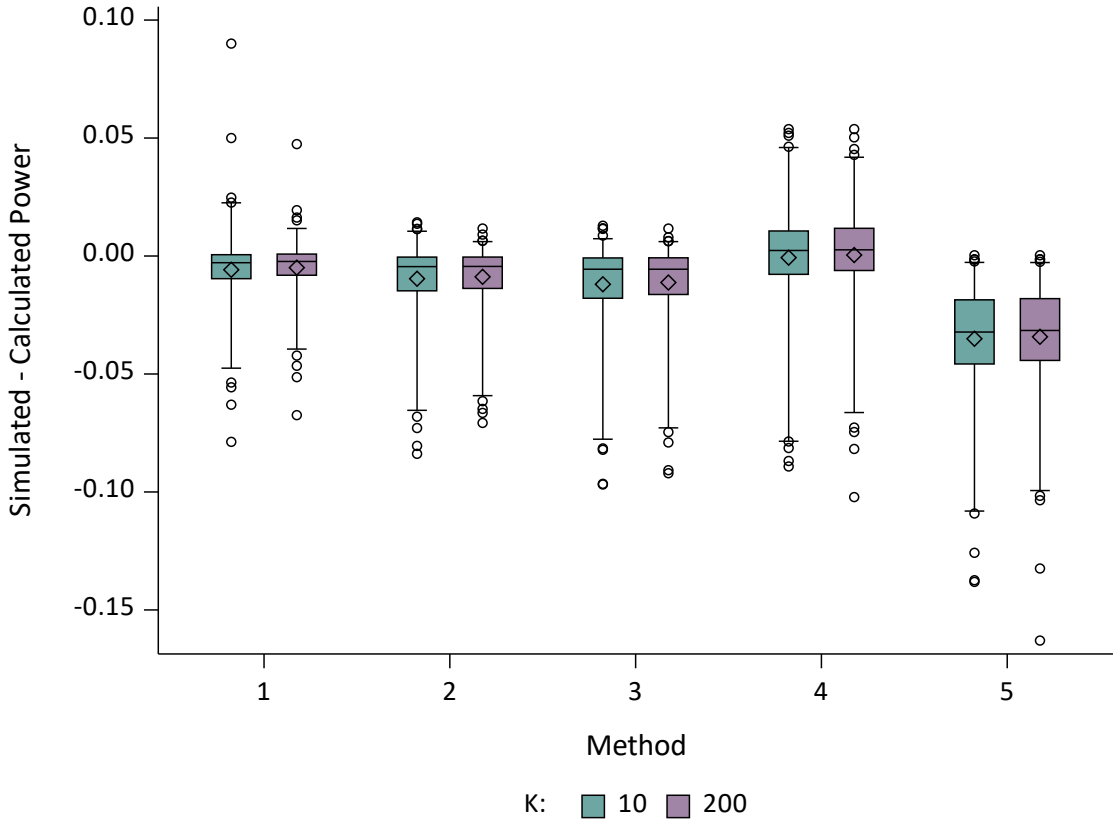
A. Comparing All Methods

Given the complexity of method 1, we again took the opportunity to compare power calculated by this method to that calculated by methods 2-5 described in Chapter 2 section II. Specifically, we averaged power calculated by methods 2-5 over 10 and 200 different $\boldsymbol{X}_U$ for the same 400 designs mentioned in Chapter 3 Section II.B. For each method we calculated

$$DK_j = \widetilde{\mathrm{E}}(power|p) - \widehat{\mathrm{E}}(power|p)_{K_j}$$

where j=1,…,5 represents the Chapter 2 method used to calculated power, and $\widehat{\mathrm{E}}(power|p)_{K_j}$ is the average power calculated by method j calculated over $K \in (10, 200)$ different $\boldsymbol{X}_U$ (all of which were the same for all methods). The performance of all 5 methods is summarized in the box plots of Figure 3.4, with boxes extending from lower 25th to upper 75th percentile and whiskers extending to lower 1st and upper 99th percentile. This figure shows that, regardless of the number of $\boldsymbol{X}_U$ used, the values calculated by method 1 tend to be most consistently close to the simulated value out of all methods with values of $D10_1$ and $D200_1$ both being highly and symmetrically concentrated around zero and with the spread of $D200_1$ being less than of $D10_1$, which is to be anticipated. Conversely, method 5 based on the residual degrees of freedom consistently overestimates power — often by more than 0.05 units, with the number of $\boldsymbol{X}_U$ used somewhat surprisingly having little impact on the overall spread.

Figure 3.4: Distribution of all 400 $D10_j$ and $D200_j$ values by Method j



Additionally, we counted the number of times $\widehat{\mathrm{E}}(power|p)_{K_j}$ calculated by a method was closest to the simulated power. Out of all 400 designs, expected power using 10 different $\boldsymbol{X}_U$ calculated by method 1 was most often closest to simulated power. Specifically, method 1 was closest 181 (45.25%) times, method 4 was closest 119 (29.75%) times, method 3 was closest 81 (20.25%) times, method 2 was closest 15 (3.75%) times, and finally method 5 was closest only 4 (1.00%) times. Results were similar when 200 different $\boldsymbol{X}_U$ were used to calculate expected power with method 1 being slightly more dominant (performing best 198 times). These results, coupled with the distributions observed in Figure 3.4, demonstrate the superior ability of method 1 to calculate expected power for the KR adjusted Wald-test for designs with some anticipated probability of any follow-up observation being missing. Nevertheless, given

the relative simplicity and frequent accuracy of method 4, we believed a more thorough comparison of methods 1 and 4 was warranted.

B. Further Comparison of Methods 1 and 4

To begin with, we wanted to examine the performance of each method when the other performed the best to determine if a pattern emerged suggesting scenarios where one method performed better than the other. For instance, if method 1 performed poorly in the 119 scenarios where method 4 performed best, this would suggest some underlying factor could be causing method 1 to perform poorly. The distribution of the difference between calculated and simulated expected power for each method when either method 1 or method 4 provides the closest value to simulated power is presented in Table 3.3. The results show that method 1 performs accurately and consistently regardless of whether method 1 or method 4 performed best. Conversely, the distribution of $D10_j$ differs substantially depending on whether method 1 or method 4 performs best. As such, there appears to be some factor that is better accommodated by method 1 than by method 4 in calculating expected power for certain scenarios.

Table 3.3: Distribution of $D10_j$ when methods 1 and 4 perform best

			$\boldsymbol{D10_j}$				
Method	**Best Method**	**N**	**Minimum**	**Q25**	**Median**	**Q75**	**Maximum**
1	1	181	0.000038	0.001552	0.003606	0.010151	0.013323
	4	119	0.000344	0.004410	0.007783	0.009911	0.013526
4	1	181	0.000674	0.006541	0.012838	0.017899	0.022694
	4	119	0.000024	0.001669	0.003193	0.006422	0.009421

*Q25 and Q75 represent the 25-th and 75-th percentiles, respectively

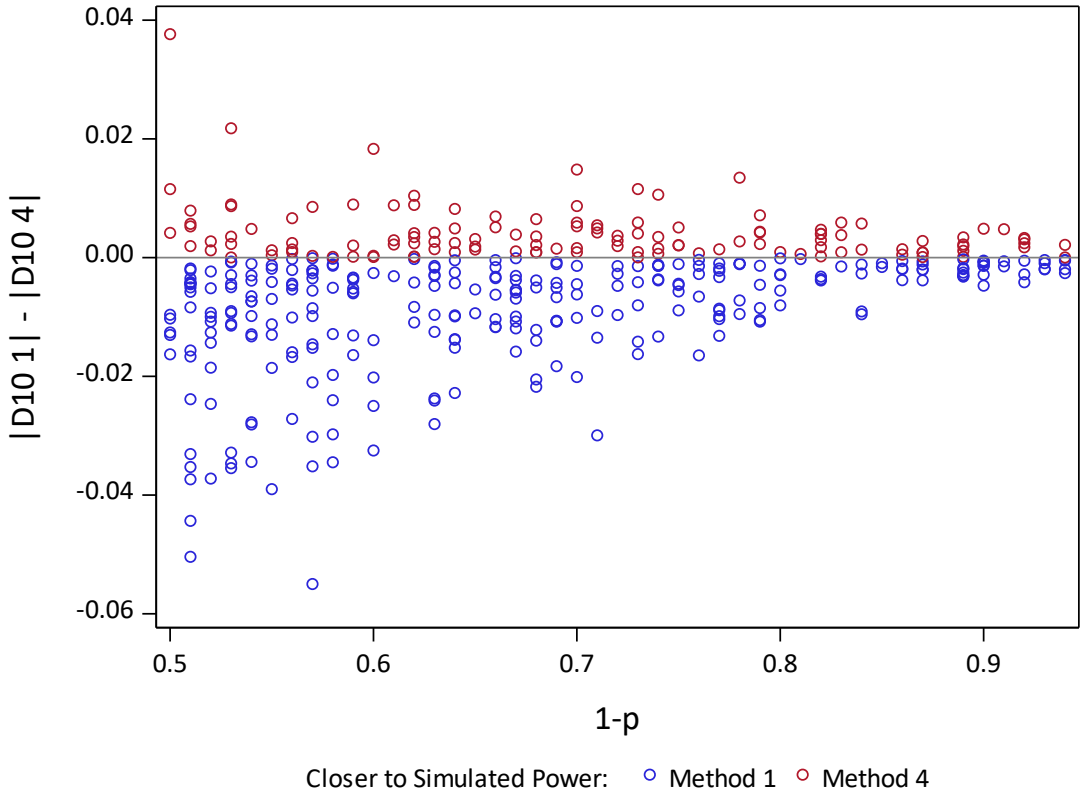
Both methods provide the same value when the distribution of the Wald statistic for the LMM is known exactly, i.e. when data are balanced. The degree of imbalance in

the design then seems a natural place to look for a divergence between the two methods that could explain the reduction in performance of method 4 in certain scenarios. To do so we looked at the difference in difference between simulated and calculated expected power between the two methods. Specifically, we looked at the distribution of $DD10_{1,4}$ defined as:

$$DD10_{1,4} = |D10_1| - |D10_4|$$

such that $DD10_{1,4}$ will be positive when calculated power for method 4 provides a closer value to simulated power and $DD10_{1,4}$ will be negative when calculated power for method 1 provides a closer value to simulated power. Values of $DD10_{1,4}$ were then plotted in Figure 3.5, which shows that as the probability of missing follow-up observations increases, the relative performance of methods 1 and 4 diverges. Moreover, this divergence is driven predominately by a larger error in method 4. Altogether, the results provided in Table 3.3 and Figure 3.5 suggest that both methods perform well when the design of a planned study is relatively balanced, but that method 1 outperforms the simpler method 4 when designs are imbalanced, such as when the probability of a follow-up observation being missing is high.

Figure 3.5: Values of $|D10_1| - |D10_4|$ for all 400 designs examined



V. Comparing newly developed approach to that of Ringham et al.

In the previous sections, we've compared several potential methods to calculate expected power that all obtain expected power by averaging power over multiple different $\boldsymbol{X}_U$. The results indicate that calculating power using method 1 from Chapter 2 Section II tends to provide the best approximation to expected power. In this section, we therefore compared the performance of $\widehat{\mathrm{E}}(power|p)_{10}$ and $\widehat{\mathrm{E}}(power|p)_{200}$ obtained using method 1 to the method described by Ringham et al., which was derived explicitly to provide expected power given some probability of any observation being missing, but does not require averaging over different values of $\boldsymbol{X}_U$ and is therefore computationally efficient.

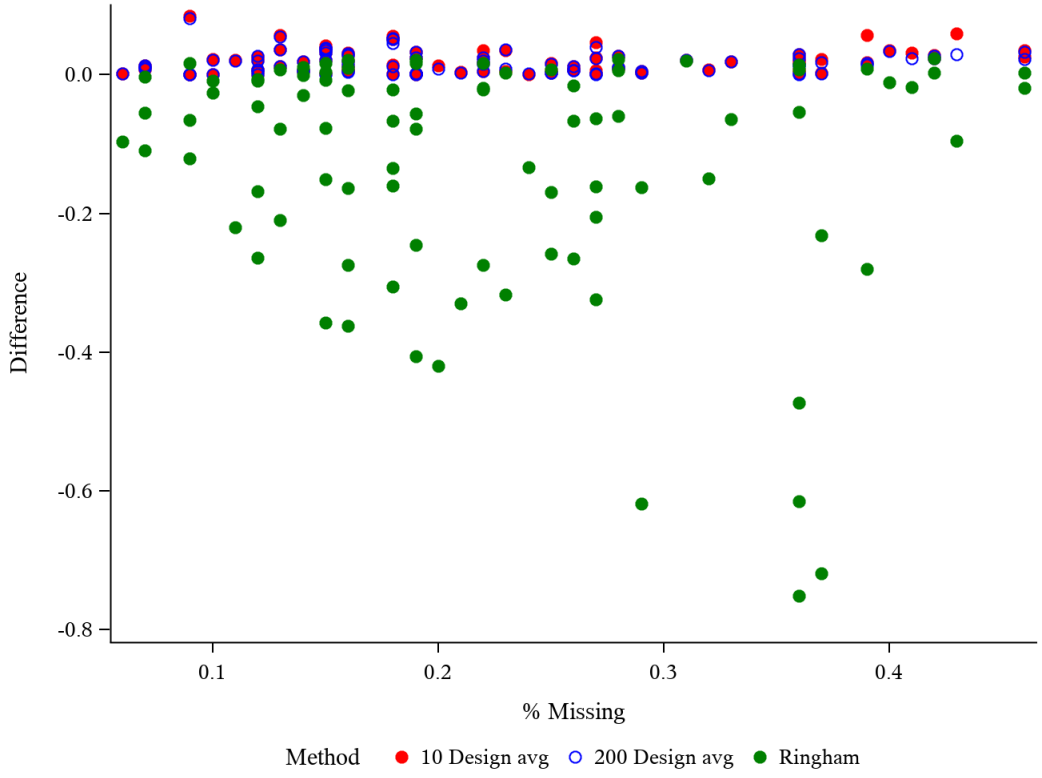
To compare these two methods, 120 designs were generated where again all model standard deviations were less than 100, and the ratio of group by time interaction effect was no more than 3 times the standard deviation of the random slope in order to keep power from being too large. Possible observation time values and their corresponding GLMM within subjects contrasts are provided in Appendix III table A.III.1, with all designs having 10 subjects per treatment group for a total of 20 subjects. For each possible design, $\widehat{\mathrm{E}}(power|p)_{10}$ and $\widehat{\mathrm{E}}(power|p)_{200}$ were generated via method 1, as was the power estimate provided by Ringham, $\widehat{\mathrm{E}}(power|p)_{R}$. Note that while Ringham et al. made no mandate that baseline values be nonmissing, we have done so and therefore modified n_{c} to be $n_{\mathrm{c}} = n(1-p)^{\tau-1}$ instead of $n_{\mathrm{c}} = n(1-p)^{\tau}$. For each design, $\widetilde{\mathrm{E}}(power|p)$ was also calculated from 20000 simulations as the proportion of times the null hypothesis was rejected (again $\alpha = 0.05$).

Results are provided in Figure 3.6, with the anticipated rate of missingness on the X axis and the difference between calculated and empirical power for each of the three methods provided on the Y axis. Results of 32 designs were excluded, because the denominator degrees of freedom calculated by the method of Ringham et al. were between 0 and 0.9, and SAS would not calculate a critical value for this F distribution via the FINV function.

Figure 3.6 shows that all methods are comparable and similar to empirical power for $p \leq 0.1$, which corresponds well with the results originally published by Ringham et al. However, as the rate of missing data increases, the $\widehat{\mathrm{E}}(power|p)_{R}$ vastly underestimates power due to the exponential effect of time in reducing the noncentrality parameter. This point is further clarified in Appendix III Figure A.III.1 showing the

comparatively strong performance of the Ringham et al. method in designs with only 4 observations per subject. This issue would obviously become even more pronounced for studies involving a large number of observations per subject. Conversely, both $\widehat{\mathrm{E}}(power|p)_{10}$ and $\widehat{\mathrm{E}}(power|p)_{200}$ track closely to empirical power, with $\widehat{\mathrm{E}}(power|p)_{200}$ tending to outperform $\mathrm{E}(\widehat{power}|p)_{10}$. As such, these results highlight the need of a method to calculate power that is not overly conservative, and suggests the method we have developed can accurately address this limitation with minimal computational complexity.

Figure 3.6: Comparison of $\widehat{\mathrm{E}}(power|p)_K$ vs Ringham method for *K*=10 and 200



VI. Discussion

In this chapter, we developed a method to calculate power for longitudinal studies with anticipated probability of missing follow-up observations and that plan on analyzing study data using linear mixed models. Specifically, this method utilizes the power calculation methods established in Chapter 2 to calculate power for a number of observation patterns generated from the full non-missing study design according to the anticipated rate of missing data. The average of these power calculations is then used as an estimate of the expected power of the study given the rate of missing data. Using method 1 to calculate power provided the most accurate estimate of expected power compared to methods 2-5. Moreover, method 1 was shown to perform accurately in simulations and outperform the method developed by Ringham et al. when the rate of missing data is greater than 10%, regardless of the number of designs averaged over to estimate expected study power. Lastly, a heuristic quantity G was developed to quickly determine whether a small or large number of designs need to be averaged over in estimating power. However, this method has a few limitations.

First and foremost, this method lacks a closed form solution and requires more computational resources than the method of Ringham et al. Consequently, this method cannot be used to calculate sample size for a study in a non-iterative fashion, but given the small number of different observations typically required for power estimation, we believe required sample size can often be obtained quickly. Second, while this method performed well in calculating power for most studies having G values less than around 80, the number of designs that should be averaged over when the G value is larger could be greater than 200. However, we conjecture such studies are rare, and so considering no

more than 200 different should be sufficient to calculate expected power for most studies. Lastly, this method was only evaluated for tests of the group by time interaction in linear mixed models. While we suspect the method to hold for individual tests of treatment group or time, further research may be warranted before the methods described are extended to the tests of these fixed effects. In the meantime, the results in this chapter further support the accuracy of the power calculation method derived in chapter 2, and the present extension of this method can help investigators accommodate anticipated missingness into study designs and ideally serve as a foundation for sample size calculations in the future, which shall be the focus of the next chapter.

CHAPTER 4
CALCULATING SAMPLE SIZE FOR THE KR ADJUSTED WALD TEST

I. Introduction

Thus far, this dissertation has focused on finding ways to calculate power for longitudinal studies where outcomes will be analyzed via LMM using the KR adjusted Wald test of linear trend. Chapter 2 developed a novel way to calculate power for these studies for a specified observation pattern. Chapter 3 extended these techniques to calculate expected power given a pre-specified rate of missingness among follow-up observations by leveraging heuristic techniques allowing for both accuracy and, in many cases, speed. In this chapter, we turn finally to the ultimate goal of this dissertation: determining the number of subjects required to achieve a desired power to reject the null hypothesis with a given α level.

While Chapter 2 presents an array of literature addressing power calculation for the KR test, comparatively little attention appears to have been lent towards the task of calculating sample size – at least with respect to the KR test directly. As mentioned by Chi et al. (2018), the sample size can be calculated for the KR adjusted test of linear trend in the LMM when data are balanced by utilizing a suite of sample size calculation methods developed for equivalent tests in the GLMM. Calculating power for the KR test of linear trend when designs are unbalanced, for instance with some anticipated rate of missing data, then remains an open problem with likely many possible approaches.

Before detailing our own approach, we will present existing sample size calculation methods for related tests. We will then build upon these methods to develop power calculation methods of our own.

II. Existing Sample Size Calculation Techniques For Related Tests

We frame the following discussion with the task of calculating sample size for the *t*-test . Assuming $x_1 \ldots x_n$ are iid drawn from a $normal(\mu, \sigma^2)$ population, then the quantity

$$\frac{\bar{x} - \mu}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1}$$

and the *t*-test will reject the two-sided hypothesis $H_0: \mu = \mu_0 = 0$ with size α when

$$\left| \frac{\bar{x} - \mu}{\sqrt{\frac{S^2}{n}}} \right| > t_{n-1,1-\frac{\alpha}{2}}$$

If the true value of μ is μ_a, then the sample size required to reject the null hypothesis that $\mu = 0$ at the α level with power =1- β (with β the type II error rate) is given as

$$\left| \frac{\mu_a - 0}{\sqrt{\frac{S^2}{n}}} \right| = t_{n-1,1-\alpha/2} + t_{n-1,1-\beta} \quad \rightarrow \quad n = \frac{\left(t_{n-1,1-\alpha/2} + t_{n-1,1-\beta}\right)\sigma^2}{\mu_a^2}$$

This sample size formula presents an obvious dilemma: the value of n depends on the *t* distribution's degrees of freedom $n - 1$. Instead, the asymptotic equivalence of *t* and z distributions is often invoked with the sample size formula modified instead to

$$n = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)\sigma^2}{\mu_a^2}$$

This formula is now easily applied to the case of simple linear regression and calculating power for the effect given by β_k. Assuming $\hat{\beta}_k \sim normal\left(\beta_k, \sigma^2_{\hat{\beta}_k}\right)$, then

$$n = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)\sigma^2_{\hat{\beta}_k}}{\beta_k^2}$$

Extending this approach to longitudinal designs, however, remains more challenging.

Previously mentioned formulas rely on the value of $\sigma^2_{\hat{\beta}_k}$, which in cases of simple linear regression can, when testing the effect of a single treatment effect, be easily specified in the design stage and is uniquely determined by error variance σ_e^2. In contrast, specification of $\sigma^2_{\hat{\beta}_k}$ in longitudinal studies requires more effort, particularly when the primary effect of interest is the interaction between treatment and time. In such studies, specifying covariance structure and parameters for the outcome **Y** is required, and $\sigma^2_{\hat{\beta}_k}$ can be determined by the kkth element of $\left(\boldsymbol{X}'\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-\mathbf{1}}\boldsymbol{X}'\right)^{-1}$,which depends on both the number of observations and values of time at which observations are taken. Additionally, specifying hypothetical covariance parameter values in complicated covariance structures (such as for random intercept and slope models) may be challenging for investigators unfamiliar with such approaches, although Basagaña and Spiegelman (2012) suggest methods by which these parameters can be derived from questions more intuitive to investigators.

Lu et al. (2008) takes a simple approach to this problem by obtaining the amount φ by which the number of subjects required at the last visit would need to be multiplied so that the variance in treatment effect at the last time point with missing data is the same as what the variance would be without missing data. The sample size for the study is then obtained by multiplying the number required to obtain the desired power for the

treatment effect at the first visit by φ. This approach is convenient in that it does not require specifying many different missing data patterns. However, in this approach data are monotone and time is treated as a categorical variable. Additionally, defining variance of the treatment effect at the final time point is difficult for the KR adjusted Wald-test when data are missing, all of which limits this approach from being easily implemented to the cases we seek to address.

Murray (2007) and Galbraith (2002) note that when data are balanced, sample size for group by time interactions will only depend on two quantities: the vector of timepoints at which subjects are observed, and the ratio of the error variance σ_e^2, and variance of random slope. As such, when data are balanced closed form calculations can be provided for both the number of subjects and number of observations required, with the design requiring the fewest number of observations being the one having only two observations per subject (Galbraith, 2002). However, when data are not balanced, a similar closed form expression for calculating sample size is not possible.

Under the assumption of monotone missingness, Galbraith searches over a grid of values of n subjects and vector of τ observation times given as $\boldsymbol{t}$ (where each previous vector $\boldsymbol{t}$ is nested in the subsequent) to obtain values where the expected power is equal to the desired power. A grid search in this case is necessary because the value of $\sigma^2_{\hat{\beta}_k}$ is given as:

$$\sigma^2_{\hat{\beta}_k} = \left| \left(\sum_{j=1}^{\tau} n_j \boldsymbol{X}'_{(j)} \boldsymbol{\Sigma}^{-1}_{Y_{(j)}} \boldsymbol{X}_{(j)} \right)^{-1} \right|_{44}$$

where, as mentioned in chapter 4: n_j is the number of subjects whose data is missing after time j, and $\boldsymbol{X}_{(j)}$ and $\boldsymbol{\Sigma}_{Y_{(j)}}$ are the design and covariance matrix consisting of patients

dropping out or completing the study after time *j*. In this case, $\sigma^2_{\hat{\beta}_k}$ depends on the value of n_j, and so *n* recursively depends on itself. However, they note that if the expected power E[Power($\boldsymbol{n}$)] is approximated with with Power[E($\boldsymbol{n}$)] as defined in Chapter 3 section II (i.e. E($\boldsymbol{n}$) = $(np_1, \ldots, np_\tau)$), then with p_j being the probability that a patient has non-missing data through time *j* and missing data after, $\sigma^2_{\hat{\beta}_k}$ can instead be given as

$$\sigma^2_{\hat{\beta}_k} = \left| \left(\sum_{j=1}^{\tau} p_j \boldsymbol{X}'_{(j)} \boldsymbol{\Sigma}^{-1}_{Y_{(j)}} \boldsymbol{X}_{(j)} \right)^{-1} \right|_{44}$$

which conveniently permits a closed form calculation for the sample size required for the study. However, since this method tends to overestimate power, it by necessity also underestimates sample size even if often only by a small amount. Additionally, this approximation cannot be easily extended to the case of a general rate of missingness due to the potentially intractable number of possible observation patterns which would need to be summed over.

In summary, no method could be found to calculate power for the KR adjusted Wald test while accommodating some general missingness pattern. Most existing methods focus on monotone missingness and tend to use *z*-approximations to the distribution of the Wald statistic. The method described by Galbraith can easily be shown to calculate power correctly for the Wald test when data are balanced (and therefore non-missing) – the *z*-approximation would provide an excellent initial guess of sample size, and power based on the F distribution could quickly be found through iteration. There are therefore two primary limitations to all current methods in planning for missing data, monotone or general. First, the variance in $\widehat{\boldsymbol{\beta}}$ is obtained from $\boldsymbol{\phi}$ instead of $\boldsymbol{\phi}_A$ used in the

KR adjusted Wald test statistic, and these two quantities will typically not equal each other when data are unbalanced. Second, the variance in $\widehat{\boldsymbol{\Sigma}}_Y$ is not accounted for in the variance of $\widehat{\boldsymbol{\beta}}$.

III. A New Approach to Calculating Sample Size for the KR Adjusted Wald-Test

In this section, we will develop a method to calculate sample size for the KR test that maintains accuracy without being computationally cumbersome. Sample size will then be calculated for the expected power of a design given a probability of any follow-up observation being missing, $\mathrm{E}(power|p)$, as in Galbraith (2002). However, since data are missing in a general (not necessarily monotone) fashion, approximating $\mathrm{E}(power|p)$ with $power[\mathrm{E}(\omega|p)]$ is not often feasible. Instead, $\mathrm{E}(power|p)$ will need to be estimated as in Chapter 3 by averaging calculated power over a sample of possible $\boldsymbol{X}_{\mathrm{U}}$. Sample size can then be iteratively solved for by increasing (or decreasing) the sample size used to generate $\boldsymbol{X}_{\mathrm{F}}$ from which $\boldsymbol{X}_{\mathrm{U}}$ is drawn until the desired value of $\widehat{\mathrm{E}}(power|p)$is achieved. Fortunately, Chapter 3 has shown that $\mathrm{E}(power|p)$ can often be estimated accurately by averaging over only a small number of $\boldsymbol{X}_{\mathrm{U}}$. The remaining challenge with regard to minimizing computational speed is to determine from which sample size an iterative sample size calculation should begin.

A. Specifying the correct effect size

To calculate sample size, we need to specify an "effect size" δ or ratio of $\boldsymbol{\beta}$ to $\sigma^2_{\widehat{\boldsymbol{\beta}}}$. In the linear model (GLM, LMM, etc.) $\sigma^2_{\widehat{\boldsymbol{\beta}}}$ is obtained from $\boldsymbol{\phi}$, or in the KR case from $\boldsymbol{\phi}_{\mathrm{A}}$**.** As shown by Galbraith (2002) and Zhao and Edland (2021), when data are balanced,

a closed form expression exists allowing δ to be expressed as $\delta = n \cdot f(\boldsymbol{t}, \sigma_e^2, \sigma_i^2 or\ \sigma_s^2)$, with $f(\boldsymbol{t}, \sigma_e^2, \sigma_i^2 or\ \sigma_s^2)$ being a function of the time vector of observations $\boldsymbol{t}$, σ_e^2, and either σ_i^2 or σ_s^2 depending on the fixed effect of interest. This closed form expression permits sample size to be factored out and solved for by δ in terms of the variance components of the outcome Y instead of the variance of $\boldsymbol{\beta}$. However, when data are unbalanced, this approach is not feasible. Two problems must then be addressed when specifying for longitudinal studies under the KR framework. First (*i*), sample size must somehow be isolated from effect size δ to avoid a recursive sample size formula. Second (*ii*), $\sigma^2_{\hat{\boldsymbol{\beta}}}$ must be given in terms of $\boldsymbol{\phi}_A$ instead of $\boldsymbol{\phi}$.

(i) Isolating sample size from δ To address the first issue, we look at the balanced case and note that desired sample size n_d can be calculated by inflating the ratio of $\boldsymbol{\beta}$ to $\sigma^2_{\hat{\boldsymbol{\beta}}}$ directly in terms of a given starting sample size n_s instead of having to specify δ in terms of $\boldsymbol{\Sigma}_{\boldsymbol{Y}_i}$. For instance, in the case of a simple t-test using a standard normal approximation we can express n_d as

$$n_d = n_s\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 \left(\frac{\sqrt{\frac{\sigma^2}{n_s}}}{\mu_a}\right)^2 \quad \text{with} \quad \left(\frac{\mu_a}{\sqrt{\frac{\sigma^2}{n_s}}}\right)^2 = \omega|n_s$$

and so equivalently

$$n_d = \frac{n_s\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{\omega|n_s}$$

In other words, this formula tells us what sample size n_d is required so that $\omega = \left(z_{1-\alpha/2} + z_{1-\beta}\right)^2$ when using n_d instead of n_s. Noting that we can equivalently express *z*

as the noncentrality parameter of the *t*-distribution under the alternative hypothesis given n_s subjects, we can then finally express n_d as

$$n_d = \frac{n_s\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{(\omega|n_s)^2}$$

In short, this reformulation allows n_d to be calculated by first obtaining $\omega|n_s$ for some arbitrary sample size n_s and then solving for n_d. We can then use this method to calculate sample size by using any of the expressions for ω given in Chapter 2 Table 2.1. Specifically, we can for an arbitrary sample size n_s calculate $\widehat{\mathrm{E}}(\omega|n_s)$ by averaging $\omega|n_s$ over a number of $\boldsymbol{X}_{\mathrm{U}}$ and then obtain n_d as

$$n_d = \frac{n_s\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{\left[\widehat{\mathrm{E}}(\omega|n_s)\right]^2}$$

This method only applies to tests of individual fixed effects (i.e. only Wald tests with rank one contrasts) and is facilitated by the symmetry of the normal distribution. For instance, a random variable $N{\sim}n\left(z_{1-\alpha} + z_{1-\beta}, 1\right)$ will be greater than $z_{1-\alpha}$ exactly $(1-\beta)\%$ of the time. Conversely, the F-distribution is often nonsymmetric, and so determining the necessary value of noncentrality parameter required to achieve $(1-\beta)\%$ power is not straightforward. Instead, we will use an approximate value obtained by matching cumulants of the Z and F distributions.

These cumulants can be obtained through the 'quantile' function in SAS, however in this dissertation we will use an approximation that explicitly relates the cumulants of Z and F distributions. Specifically, Ferreira (2011) showed that, for the random variable *F* having a central *F* distribution with v_1 numerator and v_2 denominator degrees of freedom, $\mathrm{P}(F < f) \approx \mathrm{P}(Z < z)$ with $Z{\sim}normal(0,1)$ and

$$z = \frac{\sqrt[3]{\frac{(2v_2 + v_1 f/3 + v_1 - 2)f}{2v_2 + 4v_1 f/3}} - \left(1 - \frac{2}{9v_1}\right)}{\sqrt{\frac{2}{9v_1}}} \tag{4.1}$$

provided that v_2 is large and $v_2/v_1 \geq 3$. Letting $z = z_{1-\alpha} + z_{1-\beta}$ then we can solve for f in equation 4.1 (which we can refer to as $f_{(1-\alpha)+(1-\beta)}$) and obtain a new formulation for n_d that accommodates multi-rank contrasts:

$$n_d = ceil\left(ln_s \frac{f_{(1-\alpha)+(1-\beta)}}{\widehat{\mathrm{E}}(\omega|n_s)}\right) \tag{4.2}$$

Where *ceil* is the ceiling function and $f_{(1-\alpha)+(1-\beta)}$ is obtained as:

$$f_{(1-\alpha)+(1-\beta)} = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \tag{4.3}$$

$$a = \frac{v_1}{3}\ , \qquad b = v_1 + 2v_2 - \frac{4}{3}Kv_1 - 2\ , \qquad c = -2Kv_2$$

$$K = \left(1 + z\sqrt{\frac{2}{9v_1}} - \frac{2}{9v_1}\right)^3$$

Additionally, $v_1 = rank(\boldsymbol{C}) = l$ and $v_2 = 10000$ to maintain consistency with the z approximation. As an example, suppose we desire 85% power to reject the null hypothesis of the F test with $\alpha = 0.05$. Assume $v_1 = l = 4$. Here we would have $z = z_{0.95} + z_{0.85} = 1.644854 + 1.036433 = 2.681287$. Using the values in equation 4.3 we obtain $f_{(1-\alpha)+(1-\beta)} = 3.9203138$. Now, we can check that $P(Z < 2.681287) = 0.99633$ and similarly $P\left(F_{4,10000} < 3.92031\right) = 0.99651$, and so there is a high degree of concordance between the F and Z approximation. Thus, n_d can be thought of as the sample size required for the noncentrality parameter in the F distribution to be approximately equivalent to $z_{1-\alpha} + z_{1-\beta}$ in terms of cumulants. While this method does

not provide an exact match, it appears to perform well when the numerator degrees of freedom are 6 or less.

(ii) *Providing $\sigma^2_{\hat{\boldsymbol{\beta}}}$ in terms of* $\boldsymbol{\Phi}_A$ The method described in (i) uses the value of the noncentrality parameter provided in Chapter 2 Table 2.1 for method 1and therefore naturally incorporates the modifications made by KR so that var($\widehat{\boldsymbol{\beta}}$) = $\boldsymbol{\Phi}_A$ instead of $\boldsymbol{\Phi}$. However, the value of $\boldsymbol{\Phi}_A$ will crucially depend on the value of n_s, with the elements of $\mathbf{A}$ tending towards 0 (and therefore $\boldsymbol{\Phi}_A$ towards $\boldsymbol{\Phi}$) as the value of n tends to ∞ given a similar data observation pattern. With the eventual goal being an iterative algorithm that searches for the correct sample size starting at n_{d_s}, we then want to choose a value of n_s that provides and initial value of $\boldsymbol{\Phi}_A$ neither leads to largely overestimated values of $\omega|n_{d_s}$ at small values of n_{d_s}nor to largely underestimated values of $\omega|n_{d_s}$ at large values of n_{d_s}, as both scenarios would lead to a larger number of iterations needing to be performed in the search algorithm.

First, we will define a function: $\varphi(x) = r * ceil(x/r)$, where r is the number of treatment groups in the study. We settle on a value of $n_s = \varphi(50)$, as this value will provide a value of $\mathbf{A}$ (on which $\boldsymbol{\Phi}_A$ depends) that noticeably impacts ω in the presence of substantially unbalanced data, but that does not greatly underestimate ω when $n_{d_s}$is large and would lead to a large number of computations at each iteration of the sample size search. The choice of n_s was not further investigated, as it will only severely impact calculation speed in designs where data are anticipated to be highly unbalanced, e.g. when probability of any follow-up observation being missing is high.

B. Specification of the sample size calculation algorithm

After the design, model parameters, and probability *p* of any follow-up observation being missing have all been specified, the first step of the sample size calculation algorithm is to provide a starting value n_{d_1} from which to obtain the final desired sample size n_{d_f}. This is done by calculating $\widehat{\mathrm{E}}\big(\omega|n_s = \varphi(50)\big)$ as specified in A.*ii* with the number of $\boldsymbol{X}_U$ being given as:

1, if $p = 0$
25, if p > 0 and $\widehat{\mathrm{E}}(G) \leq 1$ (we use 25 instead of 10, as this can still be done quickly)
50, if p > 0 and 1< $\widehat{\mathrm{E}}(G) \leq 2.5$
100, if p > 0 and 2.5< $\widehat{\mathrm{E}}(G) \leq 5$
150, if p > 0 and 5< $\widehat{\mathrm{E}}(G) \leq 7.5$
200, if p > 0 and 7.5< $\widehat{\mathrm{E}}(G) \leq 10$
250, if p > 0 and $\widehat{\mathrm{E}}(G) > 10$

based on the results in chapter 3 suggesting larger values of G indicate a larger number of $\boldsymbol{X}_U$ need to be averaged over in order to provide a stable power estimate (while this stage is concerned with estimating the expected value of ω, the number of $\boldsymbol{X}_U$ determined in this stage will be carried forward into the power calculation phase as well). Then, n_{d_1} is obtained using equation 4.2 with $\widehat{\mathrm{E}}\big(\omega|n_s = \varphi(50)\big)$ used to calculate n_{d_1}. $\widehat{\mathrm{E}}(power|p)$ is then calculated using n_{d_1} as in chapter 3, and if $\widehat{\mathrm{E}}(power|p)$ is equal to the power desired by the investigator, the algorithm moves to step 3. Otherwise, the algorithm continues to the second step.

In the second step, the sample size in each group is increased by 1. While in more complicated designs, the sample size could be increased per group proportionate to the allocation ratio, we have restricted our focus to studies with equal group allocation ratio.

Power is recalculated using the new sample size, and the process is repeated until calculated power is at least equal to desired power.

For the third and final step, since ω could be underestimated for large sample sizes due to ω being calculated over a sample size of 50 and the elements of **A** therefore being more nonzero than would be observed with large sample sizes, the final step of the algorithm reduces the sample size in each group by the group's allocation ratio. Power is then recalculated, and the process is repeated until power is less than the desired power. The last sample size before power is less than desired power is then chosen as n_{d_f}. This final step will typically add little to overall calculation time, as n_{d_1} will typically be less than n_{d_f} and will therefore only involve a single additional iteration for the algorithm.

C. Algorithm Example

In Chapter 2 section III, we calculated power for the omnibus test of any differential effect of time in either of the two rat groups compared to placebo. In that example, power was calculated for a single value of $\boldsymbol{X}_U$ randomly sampled from $\boldsymbol{X}$. We now focus on general probability p of any follow-up observation being missing and ask: "How many rats will be required for each treatment group to achieve an expected power of 90% given anticipated p=0.15"? The model parameters are specified in Chapter 2 section III and are used as inputs to our sample size calculation algorithm.

In step 1, $\boldsymbol{X}$ is initially specified as in Chapter 2 section III with $\varphi(50) = 51$ rats or 17 rats per group. From $\boldsymbol{X}$, 10 different $\boldsymbol{X}_U$ are randomly generated according to p=0.15. For each $\boldsymbol{X}_U$, the noncentrality parameter ω_U and value G_U are calculated. Finally, $\widehat{\mathrm{E}}(\omega|n_s = 51, p = 0.15)$ and $\widehat{\mathrm{E}}(G|n_s = 51, p = 0.15)$ are calculated as the average value

of ω_U and G_U for the 10 $\boldsymbol{X}_U$. Doing so, we obtain $\widehat{\mathrm{E}}(\omega|n_s = 51, p = 0.15) = 32.1509$ and $\widehat{\mathrm{E}}(G|n_s = 51, p = 0.15) = 0.004$. Now, n_{d_1} can be calculated as

$$n_{d_1} = ceil\left(ln_s \frac{f_{(1-\alpha)+(1-\beta)}}{\widehat{\mathrm{E}}(\omega|n_s)}\right)$$

From equation 4.3 using:

$v_1 = rank(\boldsymbol{C}) = 2, \quad v_2 = 1000, \quad z = z_{0.95} + z_{0.9} = 1.64485 + 1.28155 = 2.9264$

we obtain $f_{(1-\alpha)+(1-\beta)}$ = 6.4843 and therefore $n_{d_1} = 21$ (i.e. 7 rats per group).

In step 2, we first calculate expected power obtained from 25 (since G <1) different $\boldsymbol{X}_U$ generated from consisting of $\frac{1}{3}n_{d_1} = 7$ rats per group. Doing so, we obtain $\widehat{\mathrm{E}}(power|n = 21, p = 0.15) = 0.85603$, which is less than the desired value of 0.9, and so the process is repeated using 7 + 1 = 8 rats per group instead (i.e. $n_{d_2} = 24$). Doing so, we obtain $\widehat{\mathrm{E}}(power|n = 24, p = 0.15) = 0.91359$, which is greater than 0.9, and so the algorithm proceeds to step 3.

In step 3, we subtract 1 from the n per group settled upon in step 2 and recalculate expected power. Doing so, we obtain $\widehat{\mathrm{E}}(power|n = 21, p = 0.15) = 0.85603$ which is less than 0.8. Therefore, no further steps are necessary and the algorithm terminates with $n_{d_f} = 24$ or 8 rats per group being the final sample size selected. Empirical results from 20,000 simulated trials likewise support each step of this algorithm with average simulated power using 7 rats per group obtained as 85.56% and from 8 rats per group obtained as 91.02%. This algorithm was performed in SAS 9.4 and took 23 seconds to run.

IV. Simulation Study and Evaluation of Performance

A set of simulations were performed to evaluate the performance of the methods developed to calculate expected power for the KR adjusted Wald-test. Specifically, we first calculated power for designs 1 and 5 specified in Chapter 2 Table 2.1. We compare simulated power to the expected power calculated by the algorithm specified in previous section II.B using each of power calculation methods 1-4 specified in Chapter 2 section II. Note the denominator degrees of freedom were modified such that $m = n(1-p) - l - 1$ with p the expected rather than observed proportion of follow-up observations being missing. Only models with both random intercept and random slope were examined, since Chapter 2 results suggest power and sample size for intercept only models can be calculated via simple methods. Additionally, method 5 was not examined in this simulation study as it was shown in Chapter 3 to substantially overestimate power.

Typically, the desired sample size would be the minimal number of subjects required to achieve the desired expected power. However, the ability to calculate this minimal number may not be the singular criteria by which performance should be evaluated especially when comparing two sample size calculation methods. Consider the scenario comparing two different methods: method 1 determines n=14 and n=15 to provide 84.9% and 91.0% power, respectively, while method 2 determines n=14 and n=15 to provide 87.5% and 92.5% power. Additionally, suppose the true power values for n=14 and n=15 are 85.05% and 91.2%. Method 1 would then incorrectly determine n=15 would be the number required to achieve at least 85% power, despite having more accurate nominal power values. In evaluating these methods, we therefore examine both

the ability of each method to provide the correct sample size as well as the comparison of calculated and simulated power at each sample size.

For designs 1 and 5 in Chapter 2 table 2.1, sample size was calculated to achieve 75%, 80%, 85%, 90%, and 95% expected power for each design with probability of a follow-up observation missing being 0.4, 0.2, and 0.1. For instance, Table 4.1 shows the sample size required to achieve 90% expected power for each method for design 9 assuming each follow-up observation has a 10% chance of being missing. Methods 1, 2, and 4 all calculate the correct sample size, with the nominal expected power associated with method 1 being closest to the empirical power associated with that sample size.

Table 4.1: Sample size calculation for design 1, desired power = 90%

Method	*p*	N per group calculated	Nominal Power	Empirical Power
1	0.10	9	0.9316	0.91584
2	0.10	9	0.9344	0.91584
3	0.10	8	0.9010	0.86699
4	0.10	9	0.9338	0.91584

Expected power was calculated for a total of 15 specifications for each method (power, missing probability *p* combination, so 5*3) for each design. With a total of 4 methods and 2 designs, this means a total of 120 sample size calculations were performed. The results of these calculations are presented in Appendix IV Table A.IV.1, and this table also includes power calculations that were not given as solutions for one method but did appear as solutions for other methods so that power could be compared for each method at each sample size. While the full set of results is available in Appendix IV, we will focus on certain important aspects of these results in this section.

First and foremost, we are interested in the ability of these methods to calculate the correct sample size. Table 4.2 shows the number of times each method correctly

calculated the sample size required such that the simulated (empirical) expected power achieved the desired level. On average, methods 1 and 4 calculated the correct sample size most frequently (both 73% of the time). In all methods, sample size per group was never off by more than 1. Calculations for design 5 were more accurate for all methods, with methods 1 and 4 again performing best (each being correct 87% of the time). We suspect the reason for this difference in performance lies in sample size. For design 1, calculated sample sizes ranged from 6 to 13 per group, whereas for design 5 calculated sample sizes ranged from 10 to 19 per group. Methods 1 and 4 were each only incorrect a total of 3 times when sample sizes were 10 per group or larger, and the difference between nominal and empirical power in such cases was a maximum of 0.0082 (0.9558 vs 0.9476) for method 1 and 0.0098 (0.9574 vs 0.9476) for method 4.

Table 4.2: Sample Size Calculation Performance

		Sample Size Calculation		
Design	**Method**	**Correct**	**Overestimated**	**Underestimated**
1 (Single group by time contrast)	1	9	0	6
	2	9	0	6
	3	4	0	11
	4	9	0	6
5 (Omnibus contrast)	1	13	1	1
	2	12	0	3
	3	5	0	10
	4	13	2	0
Total	1	22	1	7
	2	21	0	9
	3	9	0	21
	4	22	2	6

We also looked at the concordance between nominal and empirical expected power at each of the sample sizes calculated for all methods. Specifically, some methods had different sample sizes, so we made sure that power was calculated using all methods

for any sample size calculated by any method. For instance, suppose methods 1, 2, and 5 determined for 75%, 80%, 85%, 90%, and 95% power a total of 6, 7, 8, 10, and 12 subjects per group were required while method 3 determined 5, 7, 8, 9, and 11 subjects were required. In this scenario, we would calculate power for 5, 6, 7, 8, 9, 10, 11, and 12 subjects per group using each of the power calculation methods and compare the calculated to empirical power. Table 4.3 provides a summary of the absolute deviation between calculated and empirical expected power over all sample sizes for each method, both per design and in total. This table shows that all methods performed well in calculating nominal power for any specific sample size. Performance was poorer in design 1, again perhaps due to the smaller sample sizes, but all methods still performed well. Thus, as shown in Appendix IV Table A.IV.1, while the correct sample size was not always chosen, this can in some cases be due to nominal and empirical power being very similar but on opposite sides of the desired power threshold. Method 1 tended to provide the closest approximation to empirical power followed closely by Method 4, with both having a median difference of less 0.01 from empirical power. In fact, Method 1 achieved the closest approximation to empirical power in 25 (73.5%) of 34 sample sizes (compared to 7 times (20.6%) for method 3 and 2 times (5.9%) for method 6), and for the 9 sample sizes for which another method performed better, the maximum difference between nominal and empirical power for method 1 was 0.0087.

Table 4.3: Comparison of nominal vs empirical power at calculated sample size

			Absolute difference between Calculated and Empirical Power				
Design	**Method**	**Number of Sample Sizes**	**Min**	**Max**	**Median**	**Mean**	**SD**
1 (Single group by time contrast)	1	15	0.000048	0.062738	0.015756	0.022604	0.021254
	2	15	0.007508	0.080843	0.031440	0.034168	0.023165
	3	15	0.009786	0.091593	0.034050	0.041051	0.025604
	4	15	0.006658	0.074113	0.030442	0.031502	0.021667
5 (Omnibus contrast)	1	19	0.000063	0.008664	0.002511	0.002780	0.002093
	2	19	0.000152	0.026222	0.005116	0.007886	0.007768
	3	19	0.005950	0.062212	0.025244	0.025901	0.015850
	4	19	0.000833	0.010948	0.002408	0.003755	0.002952
Total	1	34	0.000048	0.062738	0.003954	0.011526	0.017142
	2	34	0.000152	0.080843	0.010980	0.019481	0.020881
	3	34	0.005950	0.091593	0.027503	0.032585	0.021759
	4	34	0.000833	0.074113	0.006882	0.015996	0.019987

Altogether, these results suggest that all methods are capable of providing accurate sample size calculations, with method 1 being the most accurate. However, method 4 frequently provided the same sample size as method 1 with only a slightly larger difference between nominal and empirical power suggesting this simple method may be a viable option for calculating power. Moreover, these results confirm that sample size can be calculated quickly. Each algorithm provided a sample size calculation in around 20 seconds since G was always less than 0.1 and therefore only 25 $\boldsymbol{X}_U$ were averaged over.

V. Practical Example

We turn again to the study on growth curves of rat body weights described in Chapter 2 Section III.B provided by Wolfinger (1996) and Box (1950). To summarize: rats were provided a control treatment, Thyroxin, or Thiouracil, and the question of interest for the study was whether there is a difference in average trajectory of body weight between either treatment group and the control group.

We can now calculate the sample size required to achieve 75%, 80%, 85%, and 90% power via the algorithm described in X using each power calculation method. In addition to methods 1-4 in Chapter 1, we will also use method 5 (with $m = n\tau(1-p) - rank(\boldsymbol{X}||\boldsymbol{Z})$ with p the expected rather than observed proportion of follow-up observations being missing) to show how using a traditional power calculation based on the residual degrees of freedom (as can be found in Helms 1992) or similar normal approximation to calculate sample size for the KR adjusted Wald-test compares to the sample sizes obtained using methods 1-4. In this example, we will assume each treatment group has the number of rats, and the probability of any follow-up observation being missing will be p = 0.15. We lastly performed a final sample size calculation needed for 90% power assuming β_4 = -2.44 instead of -6.43, as this scenario would likely require a large sample size, in order to compare the performance of all methods in a large sample scenario. Results are provided in Table 4.5.

These results show that methods 1-3 are again able to calculate sample size accurately for a design encountered by researchers. Method 4 tended to slightly underestimate power, while method 5 tended to vastly overestimate power leading to incorrect sample size in some instances. However, when sample sizes were large, all 5 methods performed accurately. This convergence in performance conforms with expectations: the distribution of the KR adjusted Wald-test statistic asymptotically converges with that of the traditional Wald-test in the LMM. Ultimately, these results, coupled with the results in section III and in previous chapters, highlight the need for a power and sample size approach specifically targeting the KR adjusted Wald-test when sample sizes are small, and suggest method 1 is the optimal approach tried of doing so.

Table 4.5: Sample size calculations for rat bodyweight study

Value of β_4	Desired Power	Method	Calculated N per group required	Nominal Expected Power	Empirical Expected Power	Simulations Converging
-6.43	80%	1	7	0.85603	0.85559	19839
		2	7	0.85675	0.85559	19839
		3	7	0.85722	0.85559	19839
		4	7	0.84681	0.85559	19839
		5	6	0.84961	0.78517	19695
	85%	1	7	0.85603	0.85559	19839
		2	7	0.85675	0.85559	19839
		3	7	0.85722	0.85559	19839
		4	8	0.90193	0.91020	19921
		5	7	0.90315	0.85559	19839
	90%	1	8	0.90781	0.91020	19921
		2	8	0.90842	0.91020	19921
		3	8	0.90881	0.91020	19921
		4	8	0.90193	0.91020	19921
		5	7	0.90315	0.85559	19839
-2.44	90%	1	35	0.90152	0.9018	20000
		2	35	0.90165	0.9018	20000
		3	35	0.90174	0.9018	20000
		4	35	0.90052	0.9018	20000
		5	35	0.90856	0.9018	20000

CHAPTER 5
SUMMARY DISCUSSION AND IDEAS FOR FUTURE RESEARCH

I. Summary Discussion

The Kenward Roger adjustment to the Wald test statistic in the LMM was developed to address three primary issues with using the traditional form of the statistic (equation 1.1) when making inferences about fixed effects. First the REML variance estimates are biased. Second, the variance of $\widehat{\boldsymbol{\beta}}$ does not take into account the variance in $\widehat{\boldsymbol{\Sigma}}$ on which $\widehat{\boldsymbol{\beta}}$ depends. Lastly, the distribution of the test statistic is generally unknown except for special cases. Kenward and Roger only defined their adjustment under the null hypothesis, however, and so calculating power for their adjusted statistic is not straightforward. Currently, no methods for calculating power or sample size exist that target the KR adjusted Wald statistic by specifically addressing all three issues for which the KR adjustment was developed. In this dissertation, we addressed this void in methodology by introducing a technique for calculating power and sample size (which in this dissertation we generally referred to as "method 1") that extends to the alternative hypothesis the approach originally used by Kenward and Roger to adjust the Wald statistic under the null.

Additionally, we introduced three other methods (methods 2-4) utilizing simpler methodologies. All four of these methods were compared in their ability to calculate power and sample size, both to each other and to a fifth pre-existing method described in Helms (1992) that simply uses the residual degrees of freedom and traditional value of the noncentrality parameter. Specifically, all methods were compared on their ability to

calculate power for designs with a specific observation pattern, their ability to calculate power for designs with some anticipated probability of any follow-up observation being missing, and their ability to calculate sample size for designs with some anticipated probability of any follow-up observation being missing.

All methods performed well when only a random intercept was used in the model. When both intercept and slope were included as random effects in the model, method 1 consistently outperformed the other four methods examined, especially when the designs became more imbalanced (in this case caused by an increase in missing follow-up data). Outperforming these other four methods clarifies why a sophisticated approach to calculating the KR adjusted Wald test is necessary. Outperforming methods 3-5 shows that the KR adjusted variance of $\widehat{\boldsymbol{\beta}}$ (given as $\boldsymbol{\Phi}_{\mathrm{A}}$) should be accounted for, and in the case of methods 4-5, the adjustment to the denominator degrees of freedom in the F distribution should be accounted for in any method calculating power or sample size for this test. Outperforming method 2 suggests a more subtle issue that must be accounted for, as method 2 accommodates both the adjusted variance of $\widehat{\boldsymbol{\beta}}$ and the adjustment to the denominator degrees of freedom.

Specifically, the KR method obtains denominator degrees of freedom and scale factor by matching to the expected value and variance of the adjusted Wald statistic. Using $\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}$ as the estimated variance of in $\widehat{\boldsymbol{\beta}}$ the KR adjusted Wald statistic W_A we have:

$$\mathrm{E}(W_A) = \mathrm{E}\left(\mathrm{E}\left(W_A|\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\right)\right)$$

However, method 2 utilizes the traditional relationship between the noncentrality parameter and F statistic in the linear model (i.e. $\omega = lF$). The accuracy of this relationship implicitly relies on the assumption that $\mathrm{E}(W_A) = \mathrm{E}\left(W_A|\widehat{\boldsymbol{\Phi}}_{\mathrm{A}} = \boldsymbol{\Phi}_{\mathrm{A}}\right)$. When

data are unbalanced, then the Taylor expansion in Appendix II suggest that $\mathrm{E}\left(\mathrm{E}\left(W_A|\widehat{\boldsymbol{\phi}}_{\mathrm{A}}\right)\right) < \mathrm{E}\left(W_A|\widehat{\boldsymbol{\phi}}_{\mathrm{A}} = \boldsymbol{\phi}_{\mathrm{A}}\right)$ (although there is a remainder term in this expansion that makes this statement difficult to prove definitively). Additionally, since we have

$$\mathrm{E}(\lambda W_A) = \mathrm{E}(F) = \frac{m(l+\omega)}{l(m-2)}$$

then in unbalanced designs ω obtained from $\mathrm{E}\left(\mathrm{E}\left(W_A|\widehat{\boldsymbol{\phi}}_{\mathrm{A}}\right)\right)$ must be less than if ω were obtained from $\mathrm{E}\left(W_A|\widehat{\boldsymbol{\phi}}_{\mathrm{A}} = \boldsymbol{\phi}_{\mathrm{A}}\right)$ provided degrees of freedom are the same, which is attested to by the fact that in all calculations examined in this dissertation the noncentrality parameter obtained using method 1 was smaller than that obtained using method 2. Therefore, the performance of method 1 over method 2 (although often marginal) suggests that method 1 better captures the reduction in noncentrality parameter due to the hierarchical relationship of $\widehat{\boldsymbol{\phi}}_{\mathrm{A}}$ and W_A in unbalanced designs. Ultimately, the relative performance of these methods highlights the importance of calculating power and sample size using methodology similar to that of Kenward and Roger and suggests that method 1 successfully performs this task.

Power and sample size calculations accommodating some anticipated probability p of any follow-up data being missing rely on averaging calculated power over a number of different observed design matrices $\boldsymbol{X}_U$ generated by deleting follow-up observations from the full potential design matrix $\boldsymbol{X}_F$ according to p. As such, these calculations have the potential to be computationally intensive. To limit the computational demands of this method, we developed a heuristic value G that can be used to determine when only a few number of $\boldsymbol{X}_U$ need to be averaged over in order to provide a consistent estimate of the expected power (or sample size needed to obtain such power) of the study. Specifically,

when values of G are low, we found only a few $\boldsymbol{X}_U$ need to be considered in calculating expected power. Moreover, these low values of G are more likely when power is high and p is low, which will be typical for most sample size calculations of interest to researchers. As such, these results suggest that sample size calculations performed by iteratively searching for a minimum sample size achieving desired expected power can typically be conducted quickly with few computational demands.

Aiding in the speed of such a sample size algorithm, we developed an approach to calculating an initial starting sample size from which to search that utilizes the relationship between the noncentrally parameter and the Z (standard normal) distribution. This approach relies on the approximation between F and Z distributions that, while therefore not perfectly accurate, often provides a starting sample size within one or two values of the final determined sample size and easily accommodates the general form of the Wald statistic in the linear model.

Ultimately, this dissertation not only highlights the need for methods accommodating the KR adjusted Wald statistic in power and sample size calculations, but provides an accurate method of doing so that is computationally practical. Specifically, method 1 addresses the same issues targeted by Kenward and Roger using similar methodology, and upon examination this new method provides reliably accurate power and sample size calculations, both objectively and compared to other methods, for the KR adjusted Wald statistic.

II. Limitations and Future Directions

While the methods developed in this dissertation (and method 1 in particular) have been demonstrated to accurately calculate power for the KR adjusted Wald test in the LMM, these methods are not without their limitations in applicability, nor is this dissertation without limits in scope.

Method 1 has only been developed to calculate power for models treating time as a continuous variable in the design matrix $\boldsymbol{X}$. Notably, in designs where $\boldsymbol{X}$ has a dummy variable for each time-point, the obtained solution for the noncentrality parameter will not match the value in known exact cases. The method 1 solution for the noncentrality parameter can easily be modified to obtain the correct value for both models by multiplying ω by $1 - \frac{1}{m} + \frac{l^2}{mW}$, where W is the number of within subjects factors (in this case number of time points). We have not examined this adjustment further, however, as this type of model restricts the benefits of the mixed model by requiring balanced data. Such an adjustment may nonetheless be necessary if this method is to be applied to designs incorporating random cluster effects in future research.

Another limitation with method 1 appears in the bias of the remainder term in the Taylor expansion. Specifically, when we have in balanced data $\mathrm{E}\left(\boldsymbol{\beta}'\boldsymbol{C}\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}\boldsymbol{C}'\boldsymbol{\beta}\right) = \boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta} - \boldsymbol{\beta}'\boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta}\boldsymbol{\beta} + O(n^{-1})$ (Appendix II section B). As the proportion of follow-up observations missing becomes more extreme, the bias in this value obtained by the Taylor Expansion increases. In such cases this bias will cause the numerator in the solution of the noncentrality parameter to shrink and denominator to inflate by an excessive amount, which will result in a noncentrality parameter substantially smaller than the correct value. However, this issue is relegated to extremely

unbalanced designs, for instance cases where 60 or 70 percent of follow-up observations are missing, and will thus be of little practical concern when calculating power for most studies.

The limitations in the scope of this dissertation also provide avenues for future research. For instance, this dissertation only examined linear mixed models with either random intercept or random intercept and slope. The KR adjustment, however, can be applied to more general covariance matrices. In the original adjustment developed by Kenward and Roger in 1997, a term appears in the expression for $\widehat{\boldsymbol{\Phi}}_A$ that becomes zero for linear covariance structures (i.e. structures where the second derivative of the covariance matrix is $\mathbf{0}$ with respect to all estimated parameters, as in the random intercept and slope models) and is not included in this dissertation or the original proc mixed and GLIMMIX procedure in SAS. Kenward and Roger (2009) noted that dropping this term in nonlinear covariance structures (for instance AR(1) structures in the GLM) could introduce additional sources of bias, and they provided an adjustment to their method that extends to such nonlinear covariances (and has since been implemented in proc Mixed/GLIMMIX as an additional option). Therefore, two additional areas of research could be explored: first whether method 1 as currently exists adequately calculates power for this adjusted method in nonlinear covariance models, and secondly whether a parallel adjustment to method 1 is feasible.

Another natural direction for future research would be the extension of method 1 to models that do not assume normality of the outcome. Proc GLIMMIX in SAS already permits the KR adjustment for a variety of assumed conditional outcome distributions, including Binomial and Poisson (provided a pseudo likelihood approach is used for

estimation). Therefore, future research extending method 1 to such generalized linear mixed model approach has potentially wide applicability for a wide class of models.

In the meantime, the methods examined in this dissertation, and in particular method 1, allows for accurate calculation of power and sample size for the KR adjusted Wald test in relatively small sample longitudinal studies with normally distributed outcomes. Moreover, these methods allow researchers to accurately plan for a range of missing data in the study design phase and thus allowing for power and sample size calculations that are neither overly conservative nor optimistic.

REFERENCES

Alnosaier, W. S. (2007). *Kenward-Roger approximate F test for fixed effects in mixed linear models* (Unpublished doctoral dissertation). Oregon State University.

Basagaña, X., & Spiegelman, D. (2012). *The Design of Observational Longitudinal Studies*. ms, Digital. https://cdn1.sph.harvard.edu/wp-content/uploads/sites/271/2012/08/optitxs-The-Design-of-Observational-Longitudinal-Studies.pdf

Box, G. E. (1950). Problems in the analysis of growth and wear curves. *Biometrics*, *6*(4), 362. doi:10.2307/3001781

Catellier, D. J., and Muller, K. E. (2000). Tests for Gaussian repeated measures with missing data in small samples. *Statistics in Medicine, 19*(8), 1101-1114. doi:10.1002/(sici)1097-0258(20000430)19:83.0.co;2-h

Chi, Y., Glueck, D. H., and Muller, K. E. (2018). Power and Sample Size for Fixed Effects Inference in Reversible Linear Mixed Models. *The American Statistician, 73*(4), 350-359. doi:10.1080/00031305.2017.1415972

Ferreira, D.F. (2011). A normal approximation to the F distribution. *Revista Brasileira de Biometria*, 29(2), 222–228

Fisher, R. A. (1919). XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh, 52*(2), 399-433. doi:10.1017/s0080456800012163

Fitzmaurice, G., and Molenberghs, G. (2008). Advances in longitudinal data analysis. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods Longitudinal Data Analysis,* 3-27. doi:10.1201/9781420011579.pt1

Galbraith, S., Stat, M., & Marschner, I. C. (2002). Guidelines for the design of clinical trials with longitudinal outcomes. *Controlled Clinical Trials*, *23*(3), 257–273. doi:10.1016/s0197-2456(02)00205-2

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of Profile Data. *Psychometrika*, *24*(2), 95–112. doi:10.1007/bf02289823

Gupta, A. (2006). Hotelling's T-Squared Statistic. *Encyclopedia of Environmetrics*. doi:10.1002/9780470057339.vah016

Gurka, M. J. (2006). Selecting the Best Linear Mixed Model Under REML. *The American Statistician, 60*(1), 19-26. doi:10.1198/000313006x90396

Hedeker, D., Gibbons, R. D., and Waternaux, C. (1999). Sample Size Estimation for Longitudinal Designs with Attrition: Comparing Time-Related Contrasts Between Two Groups. *Journal of Educational and Behavioral Statistics, 24*(1), 70-93. doi:10.3102/10769986024001070

Helms, R. W. (1992). Intentionally incomplete longitudinal designs: I. Methodology and comparison of some full span designs. *Statistics in Medicine*, *11*(14-15), 1889–1913. doi:10.1002/sim.4780111411

Hotelling, H. (1931). The Generalization of Student's Ratio. *The Annals of Mathematical Statistics, 2*(3), 360-378. doi:10.1214/aoms/1177732979

Hotelling, Harold. A Generalized T Test and Measure of Multivariate Dispersion. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 23--41, University of California Press, Berkeley, Calif., 1951

Johnstone, I. M., and Nadler, B. (2017). Roy's largest root test under rank-one alternatives. *Biometrika*. doi:10.1093/biomet/asw060

Josey, K. P., Ringham, B. M., Barón, A. E., Schenkman, M., Sauder, K. A., Muller, K. E., Dabelea, D., & Glueck, D. H. (2021). Power for balanced linear mixed models with complex missing data processes. *Communications in Statistics - Theory and Methods*, 1–19. doi:10.1080/03610926.2021.1909732

Kackar, A. N. and Harville, D. A. (1984) Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association*, 79, 853-862. doi:10.1080/01621459.1984.10477102

Kenward, M. G., and Roger, J. H. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics, 53*(3), 983. doi:10.2307/2533558

Kenward, M. G., & Roger, J. H. (2009). An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics & Data Analysis*, *53*(7), 2583–2595. doi:10.1016/j.csda.2008.12.013

Kreidler, S. M., and Glueck, D. H. (2014). *Calculating power for the general linear multivariate model and the general linear mixed model* (Unpublished doctoral dissertation). University of Colorado Anschutz Medical Campus.

Laird, N. M., and Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics, 38*(4), 963. doi:10.2307/2529876

Lawley, D. N. (1938). A Generalization of Fisher's z Test. *Biometrika, 30*(1/2), 180. doi:10.2307/2332232

Lu, K., Luo, X., & Chen, P.-Y. (2008). Sample Size Estimation for repeated measures analysis in randomized clinical trials with missing data. *The International Journal of Biostatistics*, 4(1). doi:10.2202/1557-4679.1098

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate analysis*. San Diego, CA: Academic Press.

McKeon, J. J. (1974). F approximations to the distribution of Hotelling's T20. *Biometrika, 61*(2), 381-383. doi:10.1093/biomet/61.2.381

Muller, K. E., and Peterson, B. L. (1984). Practical methods for computing power in testing the multivariate general linear hypothesis. *Computational Statistics & Data Analysis, 2*(2), 143-158. doi:10.1016/0167-9473(84)90002-1

Murray, D. M., Blitstein, J. L., Hannan, P. J., Baker, W. L., & Lytle, L. A. (2007). Sizing a trial to alter the trajectory of health behaviours: Methods, parameter estimates, and their application. *Statistics in Medicine*, 26(11), 2297–2316. doi:10.1002/sim.2714

O'Brien R.G., Shieh G. (1992) Pragmatic, unifying algorithm gives power probabilities for common F tests of the multivariate general linear hypothesis. In paper presented at the Annual Joint Statistical Meetings of the American Statistical Association, Boston, Massachusetts.

PASS 14 Power Analysis and Sample Size Software. 2015. NCSS, LLC. Kaysville, Utah, USA, ncss.com/software/pass.

Pillai, K. C., and Samson, P. (1959). On Hotelling's Generalization of T 2. *Biometrika, 46*(1/2), 160. doi:10.2307/2332818

Petersen, K. B. and Pedersen, M. S. (2012). The Matrix Cookbook. *Technical University of Denmark* (Version 20121115)

Ringham, B. M., Kreidler, S. M., Muller, K. E., and Glueck, D. H. (2015). Multivariate test power approximations for balanced linear mixed models in studies with missing data. *Statistics in Medicine*, *35*(17), 2921–2937. doi:10.1002/sim.6811

Simpson, S. L., Edwards, L. J., Muller, K. E., Sen, P. K., and Styner, M. A. (2010). A linear exponent AR(1) family of correlation structures. *Statistics in Medicine, 29*(17), 1825-1838. doi:10.1002/sim.3928

Singull, M., and Koski, T. (2012). On the Distribution of Matrix Quadratic Forms. *Communications in Statistics - Theory and Methods, 41*(18), 3403-3415. doi:10.1080/03610926.2011.563009

Stigler, S. M. (2007). The Epic Story of Maximum Likelihood. *Statistical Science, 22*(4), 598-620. doi:10.1214/07-sts249

Tang, Y. (2017). Closed-form REML estimators and sample size determination for mixed effects models for repeated measures under monotone missingness. *Statistics in Medicine, 36*(13), 2135-2147. doi:10.1002/sim.7270

Tu, X. M., Zhang, J., Kowalski, J., Shults, J., Feng, C., Sun, W., & Tang, W. (2007). Power analyses for longitudinal study designs with missing data. *Statistics in Medicine*, *26*(15), 2958–2981. doi:10.1002/sim.2773

Verbeke, G., & Lesaffre, E. (1999). The effect of drop-out on the efficiency of longitudinal experiments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *48*(3), 363–375. doi:10.1111/1467-9876.00158

Wang, C., Hall, C. B., and Kim, M. (2012). A comparison of power analysis methods for evaluating effects of a predictor on slopes in longitudinal designs with missing data. *Statistical Methods in Medical Research*, *24*(6), 1009–1029. doi:10.1177/0962280212437452

Wolfinger, R. D. (1996). Heterogeneous variance: Covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics*, *1*(2), 205. doi:10.2307/1400366

Zhao, Y., & Edland, S. D. (2021). Power formulas for mixed effects models with random slope and intercept comparing rate of change across groups. *The International Journal of Biostatistics*. doi:10.1515/ijb-2020-0107

APPENDIX I

CHAPTER 1 LIKELIHOOD MAXIMIZATION

Proof 1: Maximum Likelihood Estimates for the General Linear Multivariate Model

Before we begin maximizing the likelihood function, we will define a useful operation.

Definition 1: $\frac{\partial}{\partial \boldsymbol{A}}\boldsymbol{A}$

For matrix $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$, $\frac{\partial}{\partial \mathbf{A}}\mathbf{A} = \mathbf{D}_{\mathbf{A}} = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mn} \end{bmatrix}$

where $\begin{cases} d_{ij} = 1 \ \ if \ a_{ij} \in \boldsymbol{a} \\ d_{ij} = 0 \ if \ a_{ij} \notin \boldsymbol{a} \end{cases}$

and $\boldsymbol{a}$ is the set of all elements of $\mathbf{A}$ which are allowed to vary when taking the derivative. For instance, if $\boldsymbol{a} = [a_{11}, a_{24}]$, then $d_{11} = 1$, $d_{24} = 1$, and all other $d_{ij} = 0$. This concept will be necessary to show $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Sigma}}$ maximize the likelihood function regardless of which combination of elements of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are allowed to vary.

Maximizing the likelihood function

(i) We'll maximize $l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$ w.r.t. $\boldsymbol{\beta}$ by first taking $\frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$. Note that

(a) $\frac{\partial}{\partial \boldsymbol{\beta}}[(\boldsymbol{Y} - \boldsymbol{X\beta})\boldsymbol{\Sigma}^{-1}] = \frac{\partial}{\partial \boldsymbol{\beta}}[\boldsymbol{Y\Sigma}^{-1} - \boldsymbol{X\beta\Sigma}^{-1}] = \mathbf{0}_{(nxp)} - \boldsymbol{XD}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1}$

(b) $\frac{\partial}{\partial \boldsymbol{\beta}}[(\boldsymbol{Y} - \boldsymbol{X\beta})'] = \frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{Y}) - \frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}'\boldsymbol{X}') = \mathbf{0}_{(nxp)} - \boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'$

(c) $\frac{\partial}{\partial \boldsymbol{\beta}}[(\boldsymbol{Y} - \boldsymbol{X\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta})'] = (\boldsymbol{Y} - \boldsymbol{X\beta})\boldsymbol{\Sigma}^{-1}\frac{\partial}{\partial \boldsymbol{\beta}}[(\boldsymbol{Y} - \boldsymbol{X\beta})'] + \frac{\partial}{\partial \boldsymbol{\beta}}[(\boldsymbol{Y} - \boldsymbol{X\beta})\boldsymbol{\Sigma}^{-1}](\boldsymbol{Y} - \boldsymbol{X\beta})'$

$= \boldsymbol{X\beta\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}' - \boldsymbol{Y\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}' + \boldsymbol{XD}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}'\boldsymbol{X}' - \boldsymbol{XD}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}'$

Now, using cyclic permutation under the trace operator and the fact that $\frac{\partial}{\partial \boldsymbol{X}} tr(\boldsymbol{A}) = tr\left(\frac{\partial}{\partial \boldsymbol{X}}\boldsymbol{A}\right)$ we have

$$\frac{\partial}{\partial \boldsymbol{\beta}} tr[(\boldsymbol{Y} - \boldsymbol{X\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta})'] = tr\left[\boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X\beta\Sigma}^{-1} - \boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y\Sigma}^{-1} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{XD}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1} - \boldsymbol{YXD}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1}\right]$$

$$= tr\left[\left(\boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X\beta} - \boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{XD}_{\boldsymbol{\beta}} - \boldsymbol{YXD}_{\boldsymbol{\beta}}\right)\boldsymbol{\Sigma}^{-1}\right]$$

$$= tr[\boldsymbol{K}\boldsymbol{\Sigma}^{-1}]$$

This function will obviously equal zero when $\boldsymbol{K}\boldsymbol{\Sigma}^{-1} = \mathbf{O}_{pxp}$ which will occur when $\boldsymbol{X}\boldsymbol{\beta}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}' = \boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'$ since then so too will $\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}'\boldsymbol{X}' = \boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}'$. Solving we have:

$$\boldsymbol{X}\boldsymbol{\beta}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}' = \boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'$$

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y}$$

$$\boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y}$$

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}'\boldsymbol{Y}$$

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{Y}$$

For the second derivative test we have

$$\frac{\partial^2}{\partial^2\boldsymbol{\beta}} tr[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'] = tr\left[\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}' + \boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\right]$$

$$= tr\left[\left(\boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}} + \boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}}\right)\boldsymbol{\Sigma}^{-1}\right]$$

Note for $\boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}} = \boldsymbol{A}$ we have $a_{ii} = \sum_{u=1}^{n}\sum_{v=1}^{q} x_{uv}x_{uv} > 0$ for all $i = 1, \dots, \tau$ when all elements of $\boldsymbol{\beta}$ are allowed to vary and $0 < a_{ii} \leq \sum_{u=1}^{n}\sum_{v=1}^{q} x_{uv}x_{uv}$ for all $i = 1, \dots, \tau$ otherwise, and since $\boldsymbol{\Sigma}^{-1}$ is positive semidefinite with all eigenvalues, $\lambda_i(\boldsymbol{\Sigma}^{-1})$, nonnegative and at least one nonzero, we have

$$\frac{\partial^2}{\partial^2\boldsymbol{\beta}} tr[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'] = \sum_{i=1}^{\tau} a_{ii}\,\lambda_i(\boldsymbol{\Sigma}^{-1}) > 0$$

Therefore, $-\frac{1}{2}\frac{\partial^2}{\partial^2\boldsymbol{\beta}} tr[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'] < 0$ and so the MLE for $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{Y}$$

(ii) We'll maximize $l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$ w.r.t. $\boldsymbol{\Sigma}$ by first taking $\frac{\partial}{\partial \boldsymbol{\Sigma}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$. Note that $\frac{\partial}{\partial \boldsymbol{A}} log|\boldsymbol{A}| = tr\left(\boldsymbol{A}^{-1}\frac{\partial \boldsymbol{A}}{\partial \boldsymbol{A}}\right)$ (Petersen and Pedersen, 2012, expression 43) so

$$l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) = -\frac{n\tau}{2} log(2\pi) - \frac{n}{2} log|\boldsymbol{\Sigma}| - \frac{1}{2} tr[(\boldsymbol{Y} - \boldsymbol{X\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta})']$$

$$\frac{\partial}{\partial \boldsymbol{\Sigma}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) = -\frac{n}{2} tr\left[(\boldsymbol{\Sigma})^{-1}\frac{\partial}{\partial \boldsymbol{\Sigma}}(\boldsymbol{\Sigma})\right] - \frac{1}{2} tr\left[\frac{\partial}{\partial \boldsymbol{\Sigma}}(\boldsymbol{\Sigma}^{-1})(\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})\right]$$

$$= -\frac{n}{2} tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}}] + \frac{1}{2} tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})]$$

Solving for zero we have

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}} = \frac{1}{n}[\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})]$$

$$\boldsymbol{D}_{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{D}_{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})$$

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})$$

For the second partial derivative test we have

(a) $\frac{\partial^2}{\partial^2 \boldsymbol{\Sigma}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) = \frac{1}{2} tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}) + tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})]$

which evaluated at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma} = \frac{1}{n}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)$ is a function of the traces of powers (i.e. squared, cubed, etc.) of $\left[\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)\right]^{-1}$, a symmetric positive semi-definite matrix of rank>0, and is therefore greater than 0.

(b) $\frac{\partial^2}{\partial^2 \boldsymbol{\beta}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$ evaluated at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma} = \frac{1}{n}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)$ is also greater than 0 since again $\left[\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)\right]^{-1}$ is positive semidefinite, and so the proof is almost identical to the second derivative test given in **(i)** but with this new quantity instead of $\boldsymbol{\Sigma}^{-1}$

(c) $\frac{\partial^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\Sigma}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$ evaluated at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma} = \frac{1}{n}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)$

$$= -tr\left[\left(\boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}} - \boldsymbol{Y}\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\right]$$

$$= 0$$

because from (i) we have $\left(\left(\boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}} - \boldsymbol{Y}\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}}\right)\right) = \boldsymbol{0}$

when $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$

Therefore, $\left(\frac{\partial^2}{\partial^2\boldsymbol{\Sigma}}\right)\left(\frac{\partial^2}{\partial^2\boldsymbol{\beta}}\right) - \left(\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\Sigma}}\right)^2 > 0$ evaluated at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma} = \frac{1}{n}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)$. Thus, $l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$ achieves bivariate maximum at

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\,,\ \widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right).$$

Proof 2: Maximum Likelihood Estimates for the Balanced General Linear Model

(i) We'll maximizing $l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$ by first taking $\frac{\partial}{\partial\boldsymbol{\beta}}l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$. Note that

(a) $\frac{\partial}{\partial\boldsymbol{\beta}}[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}] = \frac{\partial}{\partial\boldsymbol{\beta}}[(\boldsymbol{Y}' - \boldsymbol{\beta}'\boldsymbol{X}')\boldsymbol{\Sigma}^{-1}] = -\boldsymbol{1}_{1xr}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}$

(b) $\frac{\partial}{\partial\boldsymbol{\beta}}[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})] = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}\frac{\partial}{\partial\boldsymbol{\beta}}[\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}] + \frac{\partial}{\partial\boldsymbol{\beta}}[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}](\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$

$$= -(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}} - \boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

$$= -\boldsymbol{Y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}} - \boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} + \boldsymbol{D}_{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\beta}$$

Setting $\frac{\partial}{\partial\boldsymbol{\beta}}l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) = 0$ and solving we have

$$\boldsymbol{Y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X} + \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} = \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X} + \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\beta}$$

$$\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} = \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\beta} \quad \text{(since then also } \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} = \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\beta}\text{)}$$

$$(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\beta}$$

$$(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} = \widehat{\boldsymbol{\beta}}$$

(ii) We'll next continue maximizing $l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$ by first taking $\frac{\partial}{\partial\boldsymbol{\Sigma}}l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$.

Note that $\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$ is scalar and so:

$$\tfrac{1}{2}(\boldsymbol{Y}-\boldsymbol{X\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{X\beta}) = tr\left[\tfrac{1}{2}(\boldsymbol{Y}-\boldsymbol{X\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{X\beta})\right] = tr\left[\tfrac{1}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{X\beta})(\boldsymbol{Y}-\boldsymbol{X\beta})'\right]$$

Now,

$$\frac{\partial}{\partial \boldsymbol{\Sigma}} l(\boldsymbol{\beta},\boldsymbol{\Sigma}|\boldsymbol{Y}) = -\frac{n}{2}\frac{\partial}{\partial \boldsymbol{\Sigma_i}}(log|\boldsymbol{\Sigma_i}|) - \frac{1}{2}\frac{\partial}{\partial \boldsymbol{\Sigma}}\left(tr\left[\sum_{i=1}^{n}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})'\right]\right)$$

$$= -\frac{n}{2}tr\left[(\boldsymbol{\Sigma_i})^{-1}\frac{\partial}{\partial \boldsymbol{\Sigma_i}}(\boldsymbol{\Sigma_i})\right] - \frac{1}{2}tr\left[\sum_{i=1}^{n}\frac{\partial}{\partial \boldsymbol{\Sigma}}(\boldsymbol{\Sigma}_i^{-1})(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})'\right]$$

$$= -\frac{n}{2}tr\left[\boldsymbol{\Sigma}_i^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\right] + \frac{1}{2}tr\left[\sum_{i=1}^{n}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})'\right]$$

Setting $\frac{\partial}{\partial \boldsymbol{\Sigma}} l(\boldsymbol{\beta},\boldsymbol{\Sigma}|\boldsymbol{Y}) = 0$ and solving we have,

$$\frac{n}{2}tr\left[\boldsymbol{\Sigma}_i^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\right] = \frac{1}{2}tr\left[\sum_{i=1}^{n}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})'\right]$$

$$\frac{n}{2}tr[\boldsymbol{I}_{\tau x\tau}] = \frac{1}{2}tr\left[\sum_{i=1}^{n}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})'\right]$$

Note in balanced data $\boldsymbol{\Sigma_i}$ is the same for all subjects, and so can factor out $\boldsymbol{\Sigma}_i^{-1}$to write

$$\frac{n}{2}tr[\boldsymbol{I}_{\tau x\tau}] = \frac{1}{2}tr\left[\boldsymbol{\Sigma}_i^{-1}\sum_{i=1}^{n}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})'\right]$$

$$n\boldsymbol{I}_{\tau x\tau} = \boldsymbol{\Sigma}_i^{-1}\sum_{i=1}^{n}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})'$$

$$\boldsymbol{\Sigma}_i = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})'$$

$$\widehat{\boldsymbol{\Sigma}}_i = \frac{1}{n}(\boldsymbol{Y}-\boldsymbol{X\beta})(\boldsymbol{Y}-\boldsymbol{X\beta})'$$

(iii) We'll now use the second partial derivative test to prove bivariate maximization

We found zeros in (i) and (ii) at $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X'\Sigma^{-1}X})^{-1}\boldsymbol{X'\Sigma^{-1}Y}$ and $\widehat{\boldsymbol{\Sigma}}_i = \frac{1}{n}(\boldsymbol{Y}-\boldsymbol{X\beta})(\boldsymbol{Y}-\boldsymbol{X\beta})'$

Note that $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Sigma}}_i$ are zeros regardless of values of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$. Thus $\frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$ and $\frac{\partial}{\partial \boldsymbol{\Sigma}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$ are both zero when evaluated at $\boldsymbol{\Sigma_i} = \widehat{\boldsymbol{\Sigma}}_i$ and $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$, respectively.

Next,

$$\frac{\partial^2}{\partial^2 \boldsymbol{\beta}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) = -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\beta}}\left\{-\boldsymbol{Y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}} - \boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} + \boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\beta}\right\}$$

$$= -\frac{1}{2}\left(\boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}} + \boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}}\right)$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{r} d_i d_j a_{ij} \text{ , for } \boldsymbol{A} = \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}$$

Note, $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}$ is symmetric and through the spectral theorem can be written as $\boldsymbol{G}'\boldsymbol{\Lambda}\boldsymbol{G}$, with $\boldsymbol{\Lambda}$ being a diagonal matrix of non-negative eigenvalues of $\boldsymbol{A}$ (since $\boldsymbol{\Sigma}^{-1}$ is positive definite, so is $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}$). So $\sum_{i=1}^{r}\sum_{j=1}^{r} d_i d_j a_{ij} = \sum_{j=1}^{r}\left[d_j \Lambda_{jj}\left(\sum_{i=1}^{r} d_i g_{ij}\right)^2\right] > 0$ and so, when evaluated at $\boldsymbol{\Sigma_i} = \widehat{\boldsymbol{\Sigma}}_i$ and $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$, we have

$$\frac{\partial^2}{\partial^2 \boldsymbol{\beta}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) = -\frac{1}{2}\left(\boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}} + \boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}}\right) = -\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} d_i d_j a_{ij} \quad < 0$$

Next,

$$\frac{\partial^2}{\partial^2 \boldsymbol{\Sigma}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$$

$$= \frac{\partial}{\partial \boldsymbol{\Sigma}}\left(-\frac{n}{2} tr\left[\boldsymbol{\Sigma}_i^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\right] + \frac{1}{2} tr\left[\sum_{i=1}^{n} \boldsymbol{\Sigma}_i^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\right]\right)$$

$$= \frac{n}{2} tr\left[\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\right)\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\right)\right] - \frac{2}{2} tr\left[\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\right)\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\right)\boldsymbol{\Sigma}_i^{-1}\sum_{i=1}^{n}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\right]$$

$$= \frac{n}{2} tr\left[\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\right)\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\right)\right] - \frac{2}{2} tr\left[\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\right)\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\right)\boldsymbol{\Sigma}_i^{-1} n\widehat{\boldsymbol{\Sigma}}_i\right]$$

Evaluated at $\boldsymbol{\Sigma_i} = \widehat{\boldsymbol{\Sigma}}_i = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'$ and $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}$,

$$= \frac{n}{2} tr\left[\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{D}_{\boldsymbol{\Sigma}_i})\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{D}_{\boldsymbol{\Sigma}_i})\right] - \frac{2n}{2} tr\left[\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{D}_{\boldsymbol{\Sigma}_i})\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{D}_{\boldsymbol{\Sigma}_i})\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\boldsymbol{\Sigma}}_i\right]$$

$$= -\frac{n}{2} tr\left[\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{D}_{\boldsymbol{\Sigma}_i})\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{D}_{\boldsymbol{\Sigma}_i})\right]$$

Note, when all elements of $\boldsymbol{\Sigma_i}$ are allowed to vary, $tr\left[\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{D}_{\boldsymbol{\Sigma}_i})\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{D}_{\boldsymbol{\Sigma}_i})\right] =$ $tr\left[\widehat{\boldsymbol{\Sigma}}_i^{-1}(\mathbf{1}_{\tau x\tau})\widehat{\boldsymbol{\Sigma}}_i^{-1}(\mathbf{1}_{\tau x\tau})\right]$ which is the square of the sum of all elements of $\widehat{\boldsymbol{\Sigma}}_i^{-1}$ and is therefore positive. When some elements are held fixed, $tr\left[\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{D}_{\boldsymbol{\Sigma}_i})\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{D}_{\boldsymbol{\Sigma}_i})\right]$ will obviously still be greater than zero albeit less than the square of the sum of all elements of $\widehat{\boldsymbol{\Sigma}}_i^{-1}$. Thus, when evaluated at $\boldsymbol{\Sigma_i} = \widehat{\boldsymbol{\Sigma}}_i$ and $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ we have

$$\frac{\partial^2}{\partial^2 \boldsymbol{\Sigma}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) = -\frac{n}{2} tr\left[\widehat{\boldsymbol{\Sigma}}_i^{-1}(\mathbf{1}_{pxp})\widehat{\boldsymbol{\Sigma}}_i^{-1}(\mathbf{1}_{pxp})\right] < 0$$

Finally,

$$\frac{\partial^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\Sigma}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) = \frac{\partial}{\partial \boldsymbol{\Sigma}}\left(\frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})\right)$$

$$= \frac{\partial}{\partial \boldsymbol{\Sigma}}\left(-\boldsymbol{Y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}} - \boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} + \boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\beta}\right)$$

$$= \boldsymbol{Y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}} - \frac{\partial}{\partial \boldsymbol{\Sigma}}\left(\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}}\right) + \boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} -$$

$$\frac{\partial}{\partial \boldsymbol{\Sigma}}\left(\boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\beta}\right)$$

Note when $\boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}$, we have

$$\frac{\partial}{\partial \boldsymbol{\Sigma}}\left(\boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\beta}\right) = \frac{\partial}{\partial \boldsymbol{\Sigma}}\left(\boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right)$$

$$= -\boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} + \boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} -$$

$$\boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}$$

$$= -\boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}$$

Thus, at $\boldsymbol{\Sigma_i} = \widehat{\boldsymbol{\Sigma}}_i$ and $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ we have

$$\frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\Sigma}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) = \boldsymbol{Y}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}} - \boldsymbol{Y}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}\boldsymbol{D}_{\boldsymbol{\beta}} + \boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Y} - \boldsymbol{D}'_{\boldsymbol{\beta}}\boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{D}_{\boldsymbol{\Sigma}_i}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Y}$$

$$= 0$$

In summary, at $\boldsymbol{\Sigma_i} = \widehat{\boldsymbol{\Sigma}}_i$ and $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ we have

$$\frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) = 0$$

$$\frac{\partial}{\partial \boldsymbol{\Sigma}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) = 0$$

$$\frac{\partial^2}{\partial^2 \boldsymbol{\beta}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) < 0$$

$$\frac{\partial^2}{\partial^2 \boldsymbol{\Sigma}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) < 0$$

$$\left[\frac{\partial^2}{\partial^2 \boldsymbol{\beta}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})\right]\left[\frac{\partial^2}{\partial^2 \boldsymbol{\Sigma}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})\right] - \left[\frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\Sigma}} l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})\right]^2 > 0$$

Therefore $l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$ achieves bivariate maximum at

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} \text{ and } \widehat{\boldsymbol{\Sigma}}_i = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'$$

APPENDIX II

CHAPTER 2 DEFINITIONS AND DERIVATIONS

A. Definitions

Some of these terms have already been presented in the body of the text, but have been repeated here for a consolidated list of terms.

$\boldsymbol{\theta} = \boldsymbol{C}(\boldsymbol{C}'\boldsymbol{\phi}\boldsymbol{C})^{-1}\boldsymbol{C}'$ with $\boldsymbol{\phi} = \left(\boldsymbol{X}'\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\boldsymbol{X}'\right)^{-1}$

$$\mathbf{P}_i = -\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}$$

$$\mathbf{Q}_{ij} = \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_j}\boldsymbol{\Sigma}^{-1}$$

$$\boldsymbol{M}_{ij} = \frac{\partial^2 \boldsymbol{\theta}}{\partial \sigma_i \partial \sigma_j}$$

$w_{ij} = ij^{th}$ entry of the inverse of the expected information matrix

$$\mathbf{A} = 2\tilde{\boldsymbol{\Lambda}} = 2\boldsymbol{\phi}\sum_{i=1}^{r}\sum_{j=1}^{r}\left[w_{ij}\left(\mathbf{Q}_{ij} - \mathbf{P}_i\boldsymbol{\phi}\mathbf{P}_j\right)\right]\boldsymbol{\phi} \quad \left(\text{note } \widehat{\boldsymbol{\Lambda}} \text{ is obtained using } \widehat{\boldsymbol{\Sigma}} \text{ in place of } \boldsymbol{\Sigma}\right)$$

$$\mathrm{A}_2 = \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij} tr\left(\boldsymbol{\theta}\boldsymbol{\phi}\mathbf{P}_i\boldsymbol{\phi}\boldsymbol{\theta}\boldsymbol{\phi}\mathbf{P}_j\boldsymbol{\phi}\right)$$

$$\mathrm{A}_3 = \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\boldsymbol{\beta}'\boldsymbol{M}_{ij}\boldsymbol{\beta}$$

B. Taylor expansion of Expected Value of KR adjusted Wald Statistic

$$\mathrm{E}(\mathrm{W}) = \mathrm{E}\left(\mathrm{E}\left(F|\widehat{\boldsymbol{\Sigma}}\right)\right)$$

$$= \mathrm{E}\left(tr\left[\left(\boldsymbol{C}'\widehat{\boldsymbol{\phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}(\boldsymbol{C}'(\boldsymbol{\phi}+\boldsymbol{\Lambda})\boldsymbol{C})\right]\right) + \mathrm{E}\left((\boldsymbol{\beta}-\boldsymbol{\beta}_H)'\boldsymbol{C}\left(\boldsymbol{C}'\widehat{\boldsymbol{\phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}\boldsymbol{C}'(\boldsymbol{\beta}-\boldsymbol{\beta}_H)\right)$$

From Alnosaier (2007) we have:

$$\mathrm{E}\left(tr\left[\left(\boldsymbol{C}'\widehat{\boldsymbol{\phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}(\boldsymbol{C}'(\boldsymbol{\phi}+\boldsymbol{\Lambda})\boldsymbol{C})\right]\right) = 1 + \frac{\mathrm{A}_2}{l} + O\left(n^{-\frac{3}{2}}\right)$$

To derive $\mathrm{E}\left((\boldsymbol{\beta}-\boldsymbol{\beta}_H)'\boldsymbol{C}\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}\boldsymbol{C}'(\boldsymbol{\beta}-\boldsymbol{\beta}_H)\right)$ first note from lemma 3.1.2.1 from Alnosaier (2007) we have:

$$\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1} = \left[\boldsymbol{I}-\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}\boldsymbol{C}\right)^{-1}\left(\boldsymbol{C}'\widehat{\mathbf{A}}\boldsymbol{C}\right)+O_p(n^{-2})\right]\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}\boldsymbol{C}\right)^{-1}$$

$$= \left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}\boldsymbol{C}\right)^{-1}-\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}\boldsymbol{C}\right)^{-1}\left(\boldsymbol{C}'\widehat{\mathbf{A}}\boldsymbol{C}\right)\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}\boldsymbol{C}\right)^{-1}+O_p(n^{-2})\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}\boldsymbol{C}\right)^{-1}$$

And so, noting $\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}\boldsymbol{C}\right)^{-1}$ is $O_p(n)$,

$$\mathrm{E}\left[\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}\right] = \mathrm{E}\left[\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}\boldsymbol{C}\right)^{-1}\right]-\mathrm{E}\left[\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}\boldsymbol{C}\right)^{-1}\left(\boldsymbol{C}'\widehat{\mathbf{A}}\boldsymbol{C}\right)\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}\boldsymbol{C}\right)^{-1}\right]+\mathrm{E}\left[O_p(n^{-1})\right]$$

Now, letting $\widehat{\boldsymbol{\theta}}=\boldsymbol{C}\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}\boldsymbol{C}\right)^{-1}\boldsymbol{C}'$ we then have

$$\mathrm{E}\left(\boldsymbol{\beta}'\boldsymbol{C}\left(\boldsymbol{C}'\widehat{\boldsymbol{\Phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-1}\boldsymbol{C}'\boldsymbol{\beta}\right) = \mathrm{E}\left[\boldsymbol{\beta}'\widehat{\boldsymbol{\theta}}\boldsymbol{\beta}\right]-\mathrm{E}\left[\boldsymbol{\beta}'\widehat{\boldsymbol{\theta}}\widehat{\mathbf{A}}\widehat{\boldsymbol{\theta}}\boldsymbol{\beta}\right]+O(n^{-1})$$

We can now obtain $\mathrm{E}\left[\boldsymbol{\beta}'\widehat{\boldsymbol{\theta}}\boldsymbol{\beta}\right]$ and $\mathrm{E}\left[\boldsymbol{\beta}'\widehat{\boldsymbol{\theta}}\widehat{\mathbf{A}}\widehat{\boldsymbol{\theta}}\boldsymbol{\beta}\right]$ using a Taylor expansion about $\boldsymbol{\sigma}$.

First, we'll establish bounds of certain values which shall appear in the Taylor expansion derived quantities of E(F). From Alnosaier (2007) we have:

(1) $\boldsymbol{\Phi}, \frac{\partial\boldsymbol{\Phi}}{\partial\sigma_i}, \frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\sigma_j}$ are all $O(n^{-1})$

(2) $\boldsymbol{\theta}, \frac{\partial\boldsymbol{\theta}}{\partial\sigma_i}, \frac{\partial^2\boldsymbol{\theta}}{\partial\sigma_i\sigma_j}$ are all $O(n)$

(3) $\boldsymbol{\Lambda}=\widetilde{\boldsymbol{\Lambda}}+O\left(n^{-5/2}\right)$, with $\widetilde{\boldsymbol{\Lambda}}, \frac{\partial\widetilde{\boldsymbol{\Lambda}}}{\partial\sigma_i}, \frac{\partial^2\widetilde{\boldsymbol{\Lambda}}}{\partial\sigma_i\sigma_j}$ being $O(n^{-2})$

(4) $\mathrm{E}(\hat{\sigma}_i-\sigma_i)=0+O(n^{-2})$ for linear covariance structures (i.e. of the form $\frac{\partial^2\boldsymbol{\Sigma}}{\partial\sigma_i\partial\sigma_j}=\mathbf{0}$)

(5) $Cov\left(\hat{\sigma}_i,\hat{\sigma}_j\right)=w_{ij}+O(n^{-2})$

(6) w_{ij} is $O(n^{-1})$, with w_{ij} being the ij^{th} element of the inverse of the expected information matrix

(7) $\mathrm{E}\left[(\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j)(\hat{\sigma}_k - \sigma_k)\right]$ is $O(n^{-2})$

Additionally we have

(8) $\mathrm{E}\left[(\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j)\right] = w_{ij} + O(n^{-2})$ for linear covariance structures

We can now perform the Taylor Expansion to obtain $\mathrm{E}\left[\boldsymbol{\beta}'\widehat{\boldsymbol{\theta}}\boldsymbol{\beta}\right]$ and $\mathrm{E}\left[\boldsymbol{\beta}'\widehat{\boldsymbol{\theta}}\widehat{\mathbf{A}}\widehat{\boldsymbol{\theta}}\boldsymbol{\beta}\right]$

(i) Deriving $\mathrm{E}\left[\boldsymbol{\beta}'\widehat{\boldsymbol{\theta}}\boldsymbol{\beta}\right]$

Via Taylor expansion we have:

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta} + \sum_{i=1}^{r}(\hat{\sigma}_i - \sigma_i)\frac{\partial \boldsymbol{\theta}}{\partial \sigma_i} + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}(\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j)\frac{\partial^2 \boldsymbol{\theta}}{\partial \sigma_i \partial \sigma_j} + O(n^{-1})$$

Note

$$\frac{\partial \boldsymbol{\theta}}{\partial \sigma_i} = -\mathbf{C}(\mathbf{C}'\boldsymbol{\phi}\mathbf{C})^{-1}\mathbf{C}'\frac{\partial \boldsymbol{\phi}}{\partial \sigma_j}\mathbf{C}(\mathbf{C}'\boldsymbol{\phi}\mathbf{C})^{-1}\mathbf{C}'$$

$$\frac{\partial^2 \boldsymbol{\theta}}{\partial \sigma_i \partial \sigma_j} = \boldsymbol{\theta}(\boldsymbol{\phi}\mathbf{P}_i\boldsymbol{\phi})\boldsymbol{\theta}(\boldsymbol{\phi}\mathbf{P}_j\boldsymbol{\phi})\boldsymbol{\theta} + \boldsymbol{\theta}(\boldsymbol{\phi}\mathbf{P}_j\boldsymbol{\phi})\boldsymbol{\theta}(\boldsymbol{\phi}\mathbf{P}_i\boldsymbol{\phi})\boldsymbol{\theta}$$

$$- \boldsymbol{\theta}\boldsymbol{\phi}\mathbf{P}_i\boldsymbol{\phi}\mathbf{P}_j\boldsymbol{\phi}\boldsymbol{\theta} - \boldsymbol{\theta}\boldsymbol{\phi}\mathbf{P}_j\boldsymbol{\phi}\mathbf{P}_i\boldsymbol{\phi}\boldsymbol{\theta} + \boldsymbol{\theta}\boldsymbol{\phi}\mathbf{Q}_{ij}\boldsymbol{\phi}\boldsymbol{\theta}$$

$$= \boldsymbol{M}_{ij}$$

Therefore,

$$\mathrm{E}\left[\widehat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta} + \sum_{i=1}^{r}\mathrm{E}(\hat{\sigma}_i - \sigma_i)\frac{\partial \boldsymbol{\theta}}{\partial \sigma_i} + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}\mathrm{E}\left[(\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j)\right]\frac{\partial^2 \boldsymbol{\theta}}{\partial \sigma_i \partial \sigma_j} + O(n^{-1})$$

$$= \boldsymbol{\theta} + O(n^{-1}) + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\boldsymbol{M}_{ij} + O(n^{-1}) + O(n^{-1})$$

$$= \boldsymbol{\theta} + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\boldsymbol{M}_{ij} + O(n^{-1})$$

And consequently

$$\mathrm{E}\left[\boldsymbol{\beta}'\widehat{\boldsymbol{\theta}}\boldsymbol{\beta}\right] = \boldsymbol{\beta}'\mathrm{E}\left[\widehat{\boldsymbol{\theta}}\right]\boldsymbol{\beta} = \boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta} + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\boldsymbol{\beta}'\boldsymbol{M}_{ij}\boldsymbol{\beta} + O(n^{-1})$$

(ii) Deriving $\mathrm{E}\left[\boldsymbol{\beta}'\widehat{\boldsymbol{\theta}}\widehat{\mathbf{A}}\widehat{\boldsymbol{\theta}}\boldsymbol{\beta}\right]$

First, note $\mathbf{A} = 2\widetilde{\boldsymbol{\Lambda}}$ and is therefore $O(n^{-2})$. Next,

$$\frac{\partial \boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta}}{\partial \sigma_i} = \boldsymbol{\theta}\mathbf{A}\frac{\partial \boldsymbol{\theta}}{\partial \sigma_i} + \boldsymbol{\theta}\frac{\partial \mathbf{A}}{\partial \sigma_i}\boldsymbol{\theta} + \frac{\partial \boldsymbol{\theta}}{\partial \sigma_i}\mathbf{A}\boldsymbol{\theta}$$

which is $O(1)$ since all of its terms are are $O(n)O(n^{-2})O(n) = O(1)$.

$$\frac{\partial^2 \boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta}}{\partial \sigma_i \partial \sigma_j} = \boldsymbol{\theta}\mathbf{A}\frac{\partial^2 \boldsymbol{\theta}}{\partial \sigma_i \partial \sigma_j} + \frac{\partial^2 \boldsymbol{\theta}}{\partial \sigma_i \partial \sigma_j}\mathbf{A}\boldsymbol{\theta} + \frac{\partial \boldsymbol{\theta}\mathbf{A}}{\partial \sigma_j}\frac{\partial \boldsymbol{\theta}}{\partial \sigma_i} + \frac{\partial \boldsymbol{\theta}}{\partial \sigma_i}\frac{\partial \mathbf{A}\boldsymbol{\theta}}{\partial \sigma_j} + \boldsymbol{\theta}\frac{\partial \mathbf{A}}{\partial \sigma_i}\frac{\partial \boldsymbol{\theta}}{\partial \sigma_j} + \frac{\partial \mathbf{A}}{\partial \sigma_j}\frac{\partial \boldsymbol{\theta}}{\partial \sigma_i}\boldsymbol{\theta}$$

$$+ \boldsymbol{\theta}\frac{\partial^2 \mathbf{A}}{\partial \sigma_i \partial \sigma_j}\boldsymbol{\theta}$$

which is again $O(1)$ since all of its terms are $O(n)O(n^{-2})O(n) = O(1)$.

Now via Taylor expansion we have:

$$\widehat{\boldsymbol{\theta}}\widehat{\mathbf{A}}\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta} + \sum_{i=1}^{r}(\hat{\sigma}_i - \sigma_i)\frac{\partial \boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta}}{\partial \sigma_i} + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}(\hat{\sigma}_i - \sigma_i)\left(\hat{\sigma}_j - \sigma_j\right)\frac{\partial^2 \boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta}}{\partial \sigma_i \partial \sigma_j} + O(n^{-2})$$

And so

$$\mathrm{E}\left[\widehat{\boldsymbol{\theta}}\widehat{\mathbf{A}}\widehat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta} + O(n^{-1}) \quad \text{and therefore} \quad \mathrm{E}\left[\boldsymbol{\beta}'\widehat{\boldsymbol{\theta}}\widehat{\mathbf{A}}\widehat{\boldsymbol{\theta}}\boldsymbol{\beta}\right] = \boldsymbol{\beta}'\boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta}\boldsymbol{\beta} + O(n^{-1})$$

(iii) Deriving E(W)

Putting everything together from Alnosaier and parts (i) and (ii) we have

$$\mathrm{E(W)} = \mathrm{E}\left(\mathrm{E}\left(F|\widehat{\boldsymbol{\Sigma}}\right)\right)$$

$$= \mathrm{E}\left(tr\left[\left(\boldsymbol{C}'\widehat{\boldsymbol{\phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-\mathbf{1}}(\boldsymbol{C}'(\boldsymbol{\phi}+\boldsymbol{\Lambda})\boldsymbol{C})\right]\right) + \mathrm{E}\left((\boldsymbol{\beta}-\boldsymbol{\beta}_H)'\boldsymbol{C}\left(\boldsymbol{C}'\widehat{\boldsymbol{\phi}}_{\mathrm{A}}\boldsymbol{C}\right)^{-\mathbf{1}}\boldsymbol{C}'(\boldsymbol{\beta}-\boldsymbol{\beta}_H)\right)$$

$$= 1 + \frac{\mathrm{A}_2}{l} + O\left(n^{-\frac{3}{2}}\right) + \boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta} + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\boldsymbol{\beta}'\boldsymbol{M}_{ij}\boldsymbol{\beta} + O(n^{-1}) + \boldsymbol{\beta}'\boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta}\boldsymbol{\beta}$$

$$+ O(n^{-2})$$

$$= 1 + \frac{\mathrm{A}_2}{l} + \boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta} + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\boldsymbol{\beta}'\boldsymbol{M}_{ij}\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta}\boldsymbol{\beta} + O(n^{-1})$$

Letting

$$\mathrm{A}_3 = \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\boldsymbol{\beta}'\boldsymbol{M}_{ij}\boldsymbol{\beta}$$

we can then simply write

$$\mathrm{E(W)} = 1 + \frac{\mathrm{A}_2}{l} + \boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta} + \mathrm{A}_3 + \boldsymbol{\beta}'\boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta}\boldsymbol{\beta} + O(n^{-1})$$

C. Adjusting solution of ω to match correct value in exact cases

(i) Obtaining value of $\mathrm{A}_3 + \boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta} - \boldsymbol{\beta}'\boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta}\boldsymbol{\beta}$ in balanced linear mixed models

We'll start by assuming all observations are independent, so $\boldsymbol{\Sigma}_{\boldsymbol{Y}_i} = \sigma^2\mathbf{I}_{pxp}$. We can now evaluate the terms in the expressions.

(1) $\boldsymbol{\phi} = \left(\boldsymbol{X}'\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-\mathbf{1}}\boldsymbol{X}\right)^{-1} = \left(\boldsymbol{X}'\frac{1}{\sigma^2}\mathbf{I}\boldsymbol{X}\right)^{-1} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$

(2) $\boldsymbol{\theta} = \mathbf{C}(\mathbf{C}'\boldsymbol{\phi}\mathbf{C})^{-1}\mathbf{C}' = \mathbf{C}(\mathbf{C}'\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\mathbf{C})^{-1}\mathbf{C}' = \frac{1}{\sigma^2}\mathbf{C}(\mathbf{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\mathbf{C})^{-1}\mathbf{C}'$

(3) $\mathbf{P}_1 = -\boldsymbol{X}'\boldsymbol{\Sigma}^{-\mathbf{1}}\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma^2}\boldsymbol{\Sigma}^{-\mathbf{1}}\boldsymbol{X} = -\boldsymbol{X}'\frac{1}{\sigma^2}\mathbf{I}\frac{1}{\sigma^2}\boldsymbol{X} = -\left(\frac{1}{\sigma^2}\right)^2\boldsymbol{X}'\boldsymbol{X}$

(4) $\mathbf{Q}_{11} = -\boldsymbol{X}'\boldsymbol{\Sigma}^{-\mathbf{1}}\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma^2}\boldsymbol{\Sigma}^{-\mathbf{1}}\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma^2}\boldsymbol{\Sigma}^{-\mathbf{1}}\boldsymbol{X} = \left(\frac{1}{\sigma^2}\right)^3\boldsymbol{X}'\boldsymbol{X}$

(5) $\mathbf{Q}_{11} - \mathbf{P}_1\boldsymbol{\phi}\mathbf{P}_1 = \left(\frac{1}{\sigma^2}\right)^3 \boldsymbol{X}'\boldsymbol{X} - \left(\frac{1}{\sigma^2}\right)^3 \boldsymbol{X}'\boldsymbol{X} = \mathbf{0}$

(6) $\mathbf{A} = 2\widetilde{\boldsymbol{\Lambda}} = 2\boldsymbol{\phi}\sum_{i=1}^{r}\sum_{j=1}^{r}\left[w_{ij}\left(\mathbf{Q}_{ij} - \mathbf{P}_i\boldsymbol{\phi}\mathbf{P}_j\right)\right]\boldsymbol{\phi} = 2w_{11}\boldsymbol{\Phi}(\mathbf{Q}_{11} - \mathbf{P}_1\boldsymbol{\phi}\mathbf{P}_1) = \mathbf{0}$

(7) $w_{11} = \mathrm{CRLB}(\widehat{\sigma}^2) = \frac{2\sigma^4}{np-q}$

(8) $2\boldsymbol{\theta}(\boldsymbol{\phi}\mathbf{P}_1\boldsymbol{\phi})\boldsymbol{\theta}(\boldsymbol{\phi}\mathbf{P}_1\boldsymbol{\phi})\boldsymbol{\theta} = \frac{2}{\sigma^6}\mathbf{C}(\mathbf{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\mathbf{C})^{-1}\mathbf{C}'$

(9) $\boldsymbol{\theta}\boldsymbol{\phi}(\mathbf{2}\mathbf{P}_1\boldsymbol{\phi}\mathbf{P}_1 - \mathbf{2}\mathbf{Q}_{11})\boldsymbol{\phi}\boldsymbol{\theta} = 0$

(10) $M_{11} = 2\boldsymbol{\theta}(\boldsymbol{\phi}\mathbf{P}_1\boldsymbol{\phi})\boldsymbol{\theta}(\boldsymbol{\phi}\mathbf{P}_1\boldsymbol{\phi})\boldsymbol{\theta} - \boldsymbol{\theta}\boldsymbol{\phi}(\mathbf{2}\mathbf{P}_1\boldsymbol{\phi}\mathbf{P}_1 - \mathbf{2}\mathbf{Q}_{11})\boldsymbol{\phi}\boldsymbol{\theta} = \frac{2}{\sigma^6}\mathbf{C}(\mathbf{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\mathbf{C})^{-1}\mathbf{C}'$

So:

$$w_{11}M_{11} = \frac{2\sigma^4}{np-q}\frac{2}{\sigma^6}\mathbf{C}(\mathbf{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\mathbf{C})^{-1}\mathbf{C}' = \frac{4}{np-q}\frac{1}{\sigma^2}\mathbf{C}(\mathbf{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\mathbf{C})^{-1}\mathbf{C}'$$

$$A_3 = \frac{1}{2}\boldsymbol{\beta}'w_{11}M_{11}\boldsymbol{\beta} = \frac{2}{np-q}\frac{1}{\sigma^2}\boldsymbol{\beta}'\mathbf{C}(\mathbf{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\mathbf{C})^{-1}\mathbf{C}'\boldsymbol{\beta}$$

Therefore, we have:

$$A_3 + \boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta} - \boldsymbol{\beta}'\boldsymbol{\theta}\mathbf{A}\boldsymbol{\theta}\boldsymbol{\beta} = A_3 + \boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta} = \left[\frac{2}{(np-q)\sigma^2} + \frac{1}{\sigma^2}\right]\boldsymbol{\beta}'\mathbf{C}(\mathbf{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\mathbf{C})^{-1}\mathbf{C}'\boldsymbol{\beta}$$

$$= \frac{2+np-q}{np-q}\boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta}$$

$$= \frac{m+2}{m}\boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta}$$

The value of *m* is obtained exactly by the KR method for ANOVA and HLT tests of linear trend with rank 1 contrast, and so this solution should, and upon calculation does, also generalize to the HLT test of linear trend.

(ii) Obtaining exact solution

Our goal is to adjust $\mathrm{E}_a(\mathrm{W})$ into $\mathrm{E}_a^*(\mathrm{W})$ in a way that parallels the adjustment of $\mathrm{E}_0(\mathrm{W})$ into $\mathrm{E}_0^*(\mathrm{W})$ made by Kenward and roger. In the original KR paper, in order to obtain the correct values for *l* and *m*, the value of

$$\mathrm{E}_0(\mathrm{W}) = \frac{1}{l}(l + \mathrm{A}_2) = 1 + \frac{\mathrm{A}_2}{l}$$

is instead modified to $\mathrm{E}_0^*(\mathrm{W})$ so that

$$\mathrm{E}_0^*(\mathrm{W}) = \left(1 - \frac{\mathrm{A}_2}{l}\right)^{-1}$$

Formally we can express this as:

$$\mathrm{E}_0(\mathrm{W}) = 1 + \frac{\mathrm{A}_2}{l} = a + b - 0 = a + b - c$$

$$\mathrm{E}_0^*(\mathrm{W}) = \frac{1}{1 - \frac{\mathrm{A}_2}{l}} = \frac{(a-c)^2}{a - \mathrm{b} + \mathrm{c}}$$

Now we have:

$$\mathrm{E}_a(\mathrm{W}) = \frac{1}{l}(\boldsymbol{\beta'\theta\beta} + A_3 - \boldsymbol{\beta'\theta A\theta\beta}) = \frac{1}{l}(a + b - c)$$

Noting that in the solution to the noncentrality parameter we multiply by l we can focus just on the quantity $\boldsymbol{\beta'\theta\beta} + A_3 - \boldsymbol{\beta'\theta A\theta\beta} = a + b - c$. Employing the form of the adjustment obtained by Kenward and Roger we then have the similar form:

$$\mathrm{E}_a^*(\mathrm{W}) = \frac{1}{l}\frac{(a-c)^2}{a - \mathrm{b} + \mathrm{c}} = \frac{1}{l}\frac{(\boldsymbol{\beta'\theta\beta} - \boldsymbol{\beta'\theta A\theta\beta})^2}{(\boldsymbol{\beta'\theta\beta} - A_3 + \boldsymbol{\beta'\theta A\theta\beta})}$$

And solving for $\mathrm{E}_a^*(\mathrm{W})$ as in part (i) we have:

$$\mathrm{E}_a^*(\mathrm{W}) = \frac{1}{l}\frac{(\boldsymbol{\beta'\theta\beta})^{\mathbf{2}}}{\frac{m-2}{m}\boldsymbol{\beta'\theta\beta}} = \frac{1}{l}\frac{m}{m-2}\boldsymbol{\beta'\theta\beta}$$

Therefore:

$$\omega = l\frac{\mathrm{E}_a^*(\mathrm{W})}{\mathrm{E}_0^*(\mathrm{W})} = \frac{\frac{m}{m-2}\boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta}}{\frac{m}{m-2}} = \boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta}$$

Now obviously to obtain the desired result of $\omega = \lambda\boldsymbol{\beta}'\boldsymbol{\theta}\boldsymbol{\beta}$ we simply multiply the solution to the noncentrality parameter by λ and therefore obtain

$$\omega = l\lambda\frac{\mathrm{E}_a^*(\mathrm{W})}{\mathrm{E}_0^*(\mathrm{W})}$$

APPENDIX III

CHAPTER 3 DERIVATIONS AND SUPPLEMENTAL INFORMATION

Calculating *dPower*

We use the limit definition of a derivative:

$$f'(x) = \lim_{h \to \infty} \frac{f(x+h) - f(x)}{h}$$

Let the power function given some value of the noncentrality parameter ω be given as $\psi(\omega)$, then we let h=0.000001 and calculate approximate *dPower* as:

$$dPower = \frac{\psi(\omega + h) - \psi(\omega)}{h} = \frac{\psi(\omega + 0.000001) - \psi(\omega)}{0.000001}$$

This approximation provides accurate calculation of $dPower$ when the slope of the power function around ω is not extremely steep, which is true for values of power for which researchers are interested.

Supplemental tables and figures

Table A.III.1: Observation times and GLMM contrasts for comparing to Ringham method

	Contrasts for GLMM Equivalent Hypothesis	
LMM $\boldsymbol{t}'$	Between subjects, $\boldsymbol{C}'$	Within subjects, $\boldsymbol{U}'$
[1 2 3 4]	[0 1]	[-3 -1 1 3]
[3 6 9 12 15]	[0 1]	[-12 -6 0 6 12]
[1 4 9 13 18 21]	[0 1]	[-20 -14 -4 4 14 20]
[2 5 6 9]	[0 1]	[-7 -1 1 7]
[1 3 7 10 13 17 19]	[0 1]	[-18 -14 -6 0 6 14 18]

Figure A.III.1: Performance of power methods when design has four time points

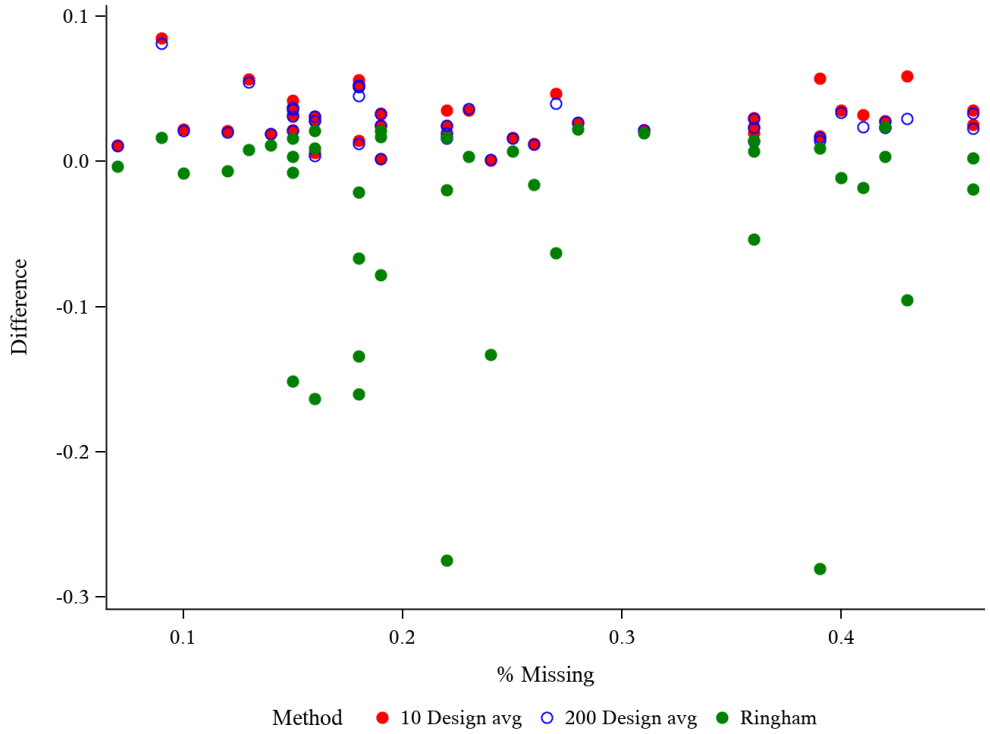
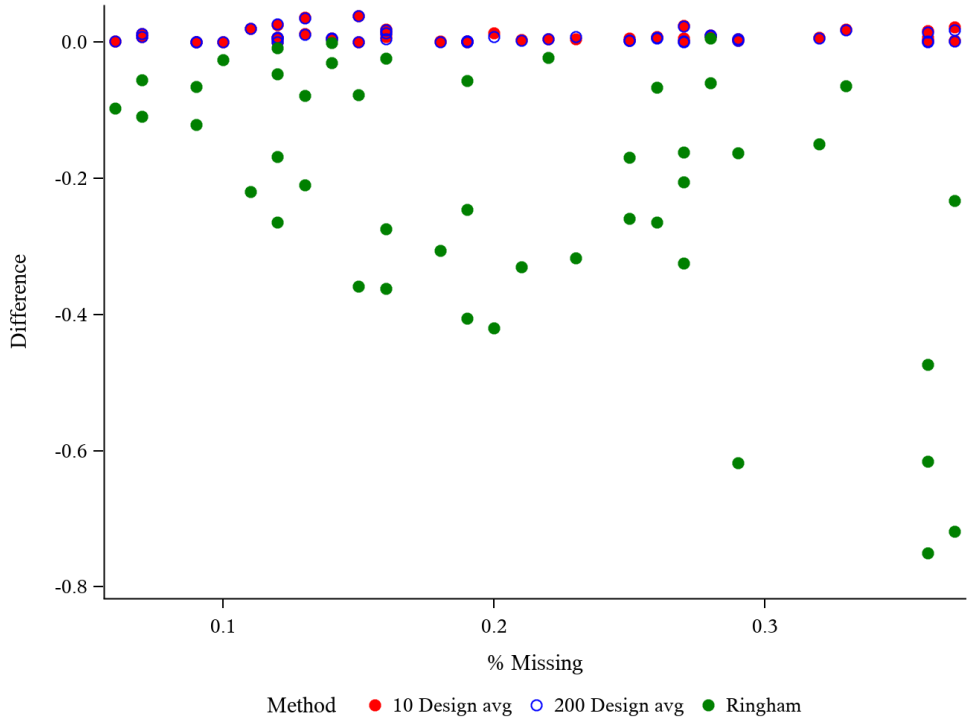


Figure A.III.2: Performance of power methods when design has more than four time points



APPENDIX IV

CHAPTER 4 SUPPLEMENTAL INFORMATION AND PROGRAMS

Supplemental Tables

Table A.IV.1: Results of Chapter 4 Simulation Study

Design	Method	1-*p*	Desired Power	N per group	Nominal Power	Empirical Power	Sample Size Calculation
1	1	0.6	0.75	8	0.75909	0.743952	Underestimate
1	1	0.6	0.8	9	0.82005	0.812538	Correct
1	1	0.6	0.85	10	0.87001	0.870058	Correct
1	1	0.6	0.9	11	0.90427	0.90717	Correct
1	1	0.6	0.95	13	0.95224	0.952842	Correct
1	1	0.8	NA	6	0.7209	0.658547	
1	1	0.8	0.75	7	0.80764	0.76397	Correct
1	1	0.8	0.8	7	0.80764	0.76397	Underestimate
1	1	0.8	0.85	8	0.8671	0.832648	Underestimate
1	1	0.8	0.9	9	0.91008	0.89368	Underestimate
1	1	0.8	0.95	11	0.95894	0.954674	Correct
1	1	0.9	0.75	6	0.7652	0.702462	Underestimate
1	1	0.9	0.8	7	0.84187	0.803974	Correct
1	1	0.9	0.85	8	0.89414	0.86699	Correct
1	1	0.9	0.9	9	0.9316	0.915844	Correct
1	1	0.9	0.95	10	0.95578	0.947608	Underestimate
1	2	0.6	0.75	8	0.79351	0.743952	Underestimate
1	2	0.6	0.8	9	0.84732	0.812538	Correct
1	2	0.6	0.85	10	0.89058	0.870058	Correct
1	2	0.6	0.9	11	0.91963	0.90717	Correct
1	2	0.6	0.95	13	0.96035	0.952842	Correct
1	2	0.8	NA	6	0.73939	0.658547	
1	2	0.8	0.75	7	0.82075	0.76397	Correct
1	2	0.8	0.8	7	0.82075	0.76397	Underestimate
1	2	0.8	0.85	8	0.87665	0.832648	Underestimate
1	2	0.8	0.9	9	0.91679	0.89368	Underestimate
1	2	0.8	0.95	11	0.9624	0.954674	Correct
1	2	0.9	0.75	6	0.77369	0.702462	Underestimate
1	2	0.9	0.8	7	0.84788	0.803974	Correct
1	2	0.9	0.85	8	0.89843	0.86699	Correct
1	2	0.9	0.9	9	0.9344	0.915844	Correct
1	2	0.9	0.95	10	0.9577	0.947608	Underestimate
1	3	0.6	0.75	8	0.81277	0.743952	Underestimate
1	3	0.6	0.8	8	0.81277	0.743952	Correct
1	3	0.6	0.85	9	0.86252	0.812538	Underestimate
1	3	0.6	0.9	10	0.902	0.870058	Underestimate

1	3	0.6	NA	11	0.9282	0.90717	
1	3	0.6	0.95	13	0.96484	0.952842	Correct
1	3	0.8	0.75	6	0.75014	0.658547	Underestimate
1	3	0.8	0.8	7	0.82845	0.76397	Underestimate
1	3	0.8	0.85	8	0.8823	0.832648	Underestimate
1	3	0.8	0.9	9	0.92077	0.89368	Underestimate
1	3	0.8	0.95	11	0.96446	0.954674	Correct
1	3	0.9	0.75	6	0.77873	0.702462	Underestimate
1	3	0.9	0.8	7	0.8515	0.803974	Correct
1	3	0.9	0.85	7	0.8515	0.803974	Underestimate
1	3	0.9	0.9	8	0.90104	0.86699	Underestimate
1	3	0.9	NA	9	0.93612	0.915844	
1	3	0.9	0.95	10	0.95888	0.947608	Underestimate
1	4	0.6	0.75	8	0.78658	0.743952	Underestimate
1	4	0.6	0.8	9	0.84298	0.812538	Correct
1	4	0.6	0.85	10	0.88801	0.870058	Correct
1	4	0.6	0.9	11	0.91736	0.90717	Correct
1	4	0.6	0.95	13	0.9595	0.952842	Correct
1	4	0.8	NA	6	0.73266	0.658547	
1	4	0.8	0.75	7	0.81607	0.76397	Correct
1	4	0.8	0.8	7	0.81607	0.76397	Underestimate
1	4	0.8	0.85	8	0.87345	0.832648	Underestimate
1	4	0.8	0.9	9	0.91475	0.89368	Underestimate
1	4	0.8	0.95	11	0.96178	0.954674	Correct
1	4	0.9	0.75	6	0.77122	0.702462	Underestimate
1	4	0.9	0.8	7	0.84639	0.803974	Correct
1	4	0.9	0.85	8	0.8975	0.86699	Correct
1	4	0.9	0.9	9	0.93383	0.915844	Correct
1	4	0.9	0.95	10	0.95741	0.947608	Underestimate
5	1	0.6	NA	11	0.69771	0.693458	
5	1	0.6	0.75	12	0.75515	0.752985	Correct
5	1	0.6	0.8	13	0.8024	0.797011	Underestimate
5	1	0.6	NA	14	0.84464	0.845952	
5	1	0.6	0.85	15	0.8752	0.876355	Correct
5	1	0.6	0.9	16	0.90496	0.902386	Correct
5	1	0.6	NA	18	0.94133	0.943501	
5	1	0.6	0.95	19	0.95476	0.954823	Correct
5	1	0.8	NA	10	0.73783	0.738354	
5	1	0.8	0.75	11	0.79532	0.800145	Correct
5	1	0.8	0.8	12	0.84551	0.847017	Overestimate
5	1	0.8	0.85	13	0.88321	0.883914	Correct

5	1	0.8	0.9	14	0.91322	0.916338	Correct
5	1	0.8	0.95	16	0.95228	0.955042	Correct
5	1	0.9	0.75	10	0.77584	0.776826	Correct
5	1	0.9	0.8	11	0.82956	0.838224	Correct
5	1	0.9	0.85	12	0.87391	0.878398	Correct
5	1	0.9	0.9	13	0.90731	0.910967	Correct
5	1	0.9	0.95	15	0.95169	0.954201	Correct
5	2	0.6	NA	11	0.71968	0.693458	
5	2	0.6	0.75	12	0.7743	0.752985	Correct
5	2	0.6	0.8	13	0.81884	0.797011	Underestimate
5	2	0.6	0.85	14	0.85785	0.845952	Underestimate
5	2	0.6	NA	15	0.88656	0.876355	
5	2	0.6	0.9	16	0.91365	0.902386	Correct
5	2	0.6	NA	18	0.94731	0.943501	
5	2	0.6	0.95	19	0.95951	0.954823	Correct
5	2	0.8	NA	10	0.74905	0.738354	
5	2	0.8	0.75	11	0.80483	0.800145	Correct
5	2	0.8	0.8	11	0.80483	0.800145	Correct
5	2	0.8	0.85	12	0.85274	0.847017	Underestimate
5	2	0.8	NA	13	0.88903	0.883914	
5	2	0.8	0.9	14	0.91763	0.916338	Correct
5	2	0.8	0.95	16	0.95489	0.955042	Correct
5	2	0.9	0.75	10	0.78241	0.776826	Correct
5	2	0.9	0.8	11	0.83496	0.838224	Correct
5	2	0.9	0.85	12	0.87787	0.878398	Correct
5	2	0.9	0.9	13	0.9103	0.910967	Correct
5	2	0.9	0.95	15	0.9533	0.954201	Correct
5	3	0.6	0.75	11	0.75567	0.693458	Underestimate
5	3	0.6	0.8	12	0.80542	0.752985	Underestimate
5	3	0.6	NA	13	0.84481	0.797011	
5	3	0.6	0.85	14	0.87926	0.845952	Underestimate
5	3	0.6	0.9	15	0.90427	0.876355	Underestimate
5	3	0.6	NA	16	0.92763	0.902386	
5	3	0.6	0.95	18	0.95641	0.943501	Underestimate
5	3	0.6	NA	19	0.96666	0.954823	
5	3	0.8	0.75	10	0.77918	0.738354	Underestimate
5	3	0.8	0.8	11	0.8297	0.800145	Correct
5	3	0.8	0.85	12	0.87254	0.847017	Underestimate
5	3	0.8	0.9	13	0.90459	0.883914	Underestimate
5	3	0.8	NA	14	0.92968	0.916338	
5	3	0.8	0.95	16	0.96193	0.955042	Correct

5	3	0.9	0.75	10	0.80804	0.776826	Correct
5	3	0.9	0.8	10	0.80804	0.776826	Underestimate
5	3	0.9	0.85	11	0.8556	0.838224	Underestimate
5	3	0.9	NA	12	0.89389	0.878398	
5	3	0.9	0.9	13	0.92259	0.910967	Correct
5	3	0.9	0.95	15	0.96015	0.954201	Correct
5	4	0.6	NA	11	0.68251	0.693458	
5	4	0.6	NA	12	0.74416	0.752985	
5	4	0.6	0.75	13	0.79474	0.797011	Overestimate
5	4	0.6	0.8	14	0.83804	0.845952	Correct
5	4	0.6	0.85	15	0.87112	0.876355	Correct
5	4	0.6	0.9	16	0.90144	0.902386	Correct
5	4	0.6	NA	18	0.94012	0.943501	
5	4	0.6	0.95	19	0.95399	0.954823	Correct
5	4	0.8	NA	10	0.73481	0.738354	
5	4	0.8	0.75	11	0.79416	0.800145	Correct
5	4	0.8	0.8	12	0.84462	0.847017	Overestimate
5	4	0.8	0.85	13	0.88308	0.883914	Correct
5	4	0.8	0.9	14	0.91324	0.916338	Correct
5	4	0.8	0.95	16	0.9527	0.955042	Correct
5	4	0.9	0.75	10	0.77799	0.776826	Correct
5	4	0.9	0.8	11	0.83218	0.838224	Correct
5	4	0.9	0.85	12	0.87599	0.878398	Correct
5	4	0.9	0.9	13	0.90914	0.910967	Correct
5	4	0.9	0.95	15	0.95286	0.954201	Correct

Blue values not included in Table 4.3 averages as they are duplicates

SAS Code

The following SAS Code on the following pages calculates sample size required to achieve 90% power using Method 1 and β_4 = -6.43 in Table 4.5. SAS program editor formatting does not carry over well to a paper document with defined margins. Thus, for a more readable file, copy and past the program into SAS.

```
/*Specify Design*/
******************;
******************;
%let ncpuse=ncp1; *Other options: ncp2, ncp3, ncp4, ncp5;
%let ddfuse=m1; *Other options: m2, m3, m4, m5;
%let xbar1=57.7; *Hypothesized population mean of treatment group 1
(investigative treatment);
%let xbar2=52.8800; *Hypothesized population mean of treatment group 2 (control
treatment);
%let xbar3=51.8; *Hypothesized population mean of treatment group 3 (comparator
treatment);
%let btime1=20.05; *Hypothesized effect of Time in group 1; *Use 24.04 for
large sample example;
%let btime2=26.48; *Hypothesized effect of Time in group 2;
%let btime3=27.5714; *Hypothesized effect of Time in group 3;
%let g1ratio=1; *Number of observations of which we will take the mean for
treatment group;
%let g2ratio=1; *Number of observations of which we will take the mean for
control group;
%let g3ratio=1; *Number of observations of which we will take the mean for
comparator group;
%let missparm=0.85; *1-probability of any follow-up being missing;
%let errdev=4.342303536; *Hypothesized within subjects error standard
deviation;
%let ranistd=5.62418883; *Hypothesized standard deviation of Random Intercept;
%let ransstd=3.888238676; *Hypothesized standard deviation of Random Slope (set
to 0 if want intercept only);
%let correlation=-0.114792316; *Hypothesized correlation between random
intercept and random slope (set to 0 if want intercept only);
%let timevec=0,1,2,3,4; *Vector of time values for each subject when no
observations missing;
%let alpha=0.05; *Desired Type-I Error Rate;
%let despower=0.9; *Desired Power;
%let seed1=607824;
%let seed2=52451;
%let seed3=7983831;
%let ran1seed1=8695815;
%let ran1seed2=64638;
%let ran2seed1=1436208;
%let ran2seed2=3378521;
%let ran3seed1=145532;
%let ran3seed2=7530156;
%let misseed1=647211;
%let misseed2=71617929;
%let misseed3=4278484;

data work.MacroVals;
       missparm=&missparm;
       n1=ceil(((&g1ratio/(&g1ratio+&g2ratio+&g3ratio))*50));
       n2=ceil(((&g2ratio/(&g1ratio+&g2ratio+&g3ratio))*50));
       n3=ceil(((&g3ratio/(&g1ratio+&g2ratio+&g3ratio))*50));
       if missparm=1 then reps=1;
       else reps=10;
       call symput('n1', n1);
       call symput('n2', n2);
       call symput('n3', n3);
       call symput('reps', reps);
run;



/*
```

```
*/
data work.Treatment;
       seed=&misseed1;
       reps=&reps;
       do rep=1 to reps;
        do n=1 to &n1;
        ID=compress(put(n,best8.)||"A");
               do Time=min(&timevec) to max(&timevec);
                     z=ranuni(seed); /*Uses uniform(0,1) random variable "z" to
assign missing values if z>missparm, modified later to force baseline as
nonmissing*/
                     if  z>&missparm then u=0;
                     else if z<=&missparm then u=1;
                     output;
               end;
         end;
        end;
        drop seed z;
run;

proc sql;
       create table work.Treatobs as
             select rep, ID, Time, case Time when . then 1 else 1 end as
intercept,
                           case Time when . then 1 else 1 end as Treatment,
                      case Time when . then 0 else 0 end as Comparator, u
             from work.Treatment
             where Time in (&timevec);
quit;

data work.Placebo;
       seed=&misseed2;
       reps=&reps;
   do rep=1 to reps;
        do n=1 to &n2;
        ID=compress(put(n,best8.)||"B");
               do time=min(&timevec) to max(&timevec);
                     z=ranuni(seed); /*Uses uniform(0,1) random variable "z" to
assign missing values if z>missparm, modified later to force baseline as
nonmissing*/
                     if  z>&missparm then u=0;
                     else if z<=&missparm then u=1;
                     output;
               end;
         end;
        end;
        drop seed z;
run;

proc sql;
       create table work.Plbobs as
             select rep, ID, Time, case Time when . then 1 else 1 end as
intercept,
                           case Time when . then 0 else 0 end as Treatment,
                           case Time when . then 0 else 0 end as Comparator, u
             from work.Placebo
             where Time in (&timevec);
quit;

data work.Comparator;
       seed=&misseed3;
       reps=&reps;
/*
```

```
*/

   do rep=1 to reps;
        do n=1 to &n3;
        ID=compress(put(n,best8.)||"C");
               do time=min(&timevec) to max(&timevec);
                     z=ranuni(seed); /*Uses uniform(0,1) random variable "z" to
assign missing values if z>missparm, modified later to force baseline as
nonmissing*/
                     if  z>&missparm then u=0;
                     else if z<=&missparm then u=1;
                     output;
               end;
         end;
        end;
        drop seed z;
run;

proc sql;
       create table work.Cmptorobs as
              select rep, ID, Time, case Time when . then 1 else 1 end as
intercept,
                           case Time when . then 0 else 0 end as Treatment,
                           case Time when . then 1 else 1 end as Comparator, u
              from work.Comparator
              where Time in (&timevec);
quit;

data work.MixedData_a; *Combine placebo and treatment designs into one dataset;
       set work.Treatobs work.Plbobs work.Cmptorobs;
run;

proc sql; *Will create variable "M" to serve as missing data indicator and make
sure no baseline values are missing;
       create table work.MixedData_b as
              select rep, ID, intercept, Treatment, Comparator, Time,
Treatment*Time as TreatTime, Comparator*Time as CompTime,
                           case when u=0 and time ne min(&timevec) then . else
1 end as M
              from work.MixedData_a;
quit;

proc sql;
       create table work.MixedData as
              select rep, ID, intercept, Treatment, Comparator, Time, TreatTime,
CompTime
              from work.MixedData_b
              where M ne .;
quit;

proc sort data=work.MixedData;
       by rep id time;
run;
quit;

%macro Mij(i=&covparmi,j=&covparmj); /*This macro calculates the second partial
derivatives of THETA with respect to covariance parameters i and j*/
       (O*PHI*P&i*PHI*O*PHI*P&j*PHI*O+O*PHI*P&j*PHI*O*PHI*P&i*PHI*O-
O*PHI*(P&i*PHI*P&j+P&j*PHI*P&i-Q&i&j-Q&j&i)*PHI*O)
%mend Mij;
/*
```

```
*/
data work.power_b;
       COL1=9999;
run;

%macro power(rep=&sim);
proc sql;
       create table work.mixeddata_&rep as
              select *
              from work.MixedData
              where rep=&rep;
quit;
proc sort data=work.mixeddata_&rep;
       by rep id time;
run;
quit;
data work.seqids_&rep;
       set work.MixedData_&rep;
       by id;
       retain order 0;
              if first.id then order=order+1;
run;
ods select all;
proc iml;
Contrasts={0 0 0 0 1 0, 0 0 0 0 0 1}`;
C=Contrasts;
L=trace(ginv(C)*C);
q=L+1; *+1 if to account for intercept;
ivar=&ranistd;
svar=&ransstd;
corr=&correlation;
errvar=&errdev*&errdev;
cov=(ivar*({1 0, 0 0})+svar*({0 0, 0 1}))*(corr*({0 1, 1 0})+I(2))*(ivar*({1 0,
0 0})+svar*({0 0, 0 1}));
bint=&xbar2; *group 2 will be used as reference group;
btreat=&xbar1-&xbar2; *treatment effect;
bcompare=&xbar3-&xbar2;
btime=&btime2; *time effect in reference group 2;
btreattime=&btime1-&btime2; *group*time interaction;
bcomptime=&btime3-&btime2;
Beta=(bint||btreat||bcompare||btime||btreattime||bcomptime)`;
B=Beta;
n=&n1+&n2+&n3;
use work.seqids_&rep;;
       do i=1 to n;
         read all var {intercept Treatment Comparator Time TreatTime CompTime}
where(order=i) into X;
         read all var {intercept Time} where(order=i) into Z;
         S=z*cov*z`+I((nrow(X)))@errvar;
         if i=1 then V=(X`*inv(S)*X);
     else V=V+(X`*inv(S)*X);
end;
PHI=inv(V);
use work.seqids_&rep;;
       do i=1 to n;
         read all var {intercept Treatment Comparator Time TreatTime CompTime}
where(order=i) into X;
         read all var {intercept Time} where(order=i) into Z;
         S=z*cov*z`+I((nrow(X)))@errvar;
         d1=z*({1 0, 0 0})*z`;
         d2=z*({0 0, 0 1})*z`;
         d3=I(nrow(X));
/*
```

```
*/
        d4=z*({0 1, 1 0})*z`;
        if i=1 then P1=-X`*inv(S)*d1*inv(S)*X; else P1=P1-
X`*inv(S)*d1*inv(S)*X;
        if i=1 then P2=-X`*inv(S)*d2*inv(S)*X; else P2=P2-
X`*inv(S)*d2*inv(S)*X;
        if i=1 then P3=-X`*inv(S)*d3*inv(S)*X; else P3=P3-
X`*inv(S)*d3*inv(S)*X;
        if i=1 then P4=-X`*inv(S)*d4*inv(S)*X; else P4=P4-
X`*inv(S)*d4*inv(S)*X;
        if i=1 then Q11=X`*inv(S)*d1*inv(S)*d1*inv(S)*X; else
Q11=Q11+X`*inv(S)*d1*inv(S)*d1*inv(S)*X;
        if i=1 then Q12=X`*inv(S)*d1*inv(S)*d2*inv(S)*X; else
Q12=Q12+X`*inv(S)*d1*inv(S)*d2*inv(S)*X;
        if i=1 then Q13=X`*inv(S)*d1*inv(S)*d3*inv(S)*X; else
Q13=Q13+X`*inv(S)*d1*inv(S)*d3*inv(S)*X;
        if i=1 then Q14=X`*inv(S)*d1*inv(S)*d4*inv(S)*X; else
Q14=Q14+X`*inv(S)*d1*inv(S)*d4*inv(S)*X;
        if i=1 then Q21=X`*inv(S)*d2*inv(S)*d1*inv(S)*X; else
Q21=Q21+X`*inv(S)*d2*inv(S)*d1*inv(S)*X;
        if i=1 then Q22=X`*inv(S)*d2*inv(S)*d2*inv(S)*X; else
Q22=Q22+X`*inv(S)*d2*inv(S)*d2*inv(S)*X;
        if i=1 then Q23=X`*inv(S)*d2*inv(S)*d3*inv(S)*X; else
Q23=Q23+X`*inv(S)*d2*inv(S)*d3*inv(S)*X;
        if i=1 then Q24=X`*inv(S)*d2*inv(S)*d4*inv(S)*X; else
Q24=Q24+X`*inv(S)*d2*inv(S)*d4*inv(S)*X;
        if i=1 then Q31=X`*inv(S)*d3*inv(S)*d1*inv(S)*X; else
Q31=Q31+X`*inv(S)*d3*inv(S)*d1*inv(S)*X;
        if i=1 then Q32=X`*inv(S)*d3*inv(S)*d2*inv(S)*X; else
Q32=Q32+X`*inv(S)*d3*inv(S)*d2*inv(S)*X;
        if i=1 then Q33=X`*inv(S)*d3*inv(S)*d3*inv(S)*X; else
Q33=Q33+X`*inv(S)*d3*inv(S)*d3*inv(S)*X;
        if i=1 then Q34=X`*inv(S)*d3*inv(S)*d4*inv(S)*X; else
Q34=Q34+X`*inv(S)*d3*inv(S)*d4*inv(S)*X;
        if i=1 then Q41=X`*inv(S)*d4*inv(S)*d1*inv(S)*X; else
Q41=Q41+X`*inv(S)*d4*inv(S)*d1*inv(S)*X;
        if i=1 then Q42=X`*inv(S)*d4*inv(S)*d2*inv(S)*X; else
Q42=Q42+X`*inv(S)*d4*inv(S)*d2*inv(S)*X;
        if i=1 then Q43=X`*inv(S)*d4*inv(S)*d3*inv(S)*X; else
Q43=Q43+X`*inv(S)*d4*inv(S)*d3*inv(S)*X;
        if i=1 then Q44=X`*inv(S)*d4*inv(S)*d4*inv(S)*X; else
Q44=Q44+X`*inv(S)*d4*inv(S)*d4*inv(S)*X;
*Information Matrices;
        if i=1 then I11a=trace(inv(S)*d1*inv(S)*d1); else
I11a=trace(I11a)+trace(inv(S)*d1*inv(S)*d1);
        if i=1 then I12a=trace(inv(S)*d1*inv(S)*d2); else
I12a=trace(I12a)+trace(inv(S)*d1*inv(S)*d2);
        if i=1 then I13a=trace(inv(S)*d1*inv(S)*d3); else
I13a=trace(I13a)+trace(inv(S)*d1*inv(S)*d3);
        if i=1 then I14a=trace(inv(S)*d1*inv(S)*d4); else
I14a=trace(I14a)+trace(inv(S)*d1*inv(S)*d4);
        if i=1 then I21a=trace(inv(S)*d2*inv(S)*d1); else
I21a=trace(I21a)+trace(inv(S)*d2*inv(S)*d1);
        if i=1 then I22a=trace(inv(S)*d2*inv(S)*d2); else
I22a=trace(I22a)+trace(inv(S)*d2*inv(S)*d2);
        if i=1 then I23a=trace(inv(S)*d2*inv(S)*d3); else
I23a=trace(I23a)+trace(inv(S)*d2*inv(S)*d3);
        if i=1 then I24a=trace(inv(S)*d2*inv(S)*d4); else
I24a=trace(I24a)+trace(inv(S)*d2*inv(S)*d4);
        if i=1 then I31a=trace(inv(S)*d3*inv(S)*d1); else
I31a=trace(I31a)+trace(inv(S)*d3*inv(S)*d1);
/*
```

```
*/
        if i=1 then I32a=trace(inv(S)*d3*inv(S)*d2); else
I32a=trace(I32a)+trace(inv(S)*d3*inv(S)*d2);
        if i=1 then I33a=trace(inv(S)*d3*inv(S)*d3); else
I33a=trace(I33a)+trace(inv(S)*d3*inv(S)*d3);
        if i=1 then I34a=trace(inv(S)*d3*inv(S)*d4); else
I34a=trace(I34a)+trace(inv(S)*d3*inv(S)*d4);
        if i=1 then I41a=trace(inv(S)*d4*inv(S)*d1); else
I41a=trace(I41a)+trace(inv(S)*d4*inv(S)*d1);
        if i=1 then I42a=trace(inv(S)*d4*inv(S)*d2); else
I42a=trace(I42a)+trace(inv(S)*d4*inv(S)*d2);
        if i=1 then I43a=trace(inv(S)*d4*inv(S)*d3); else
I43a=trace(I43a)+trace(inv(S)*d4*inv(S)*d3);
        if i=1 then I44a=trace(inv(S)*d4*inv(S)*d4); else
I44a=trace(I44a)+trace(inv(S)*d4*inv(S)*d4);
end;
I11=((0.5*trace(I11a)-0.5*trace(2*PHI*Q11-PHI*P1*PHI*P1)))*({1 0 0 0, 0 0 0 0,
0 0 0 0, 0 0 0 0});
I12=((0.5*trace(I12a)-0.5*trace(2*PHI*Q12-PHI*P1*PHI*P2)))*({0 1 0 0, 0 0 0 0,
0 0 0 0, 0 0 0 0});
I13=((0.5*trace(I13a)-0.5*trace(2*PHI*Q13-PHI*P1*PHI*P3)))*({0 0 1 0, 0 0 0 0,
0 0 0 0, 0 0 0 0});
I14=((0.5*trace(I14a)-0.5*trace(2*PHI*Q14-PHI*P1*PHI*P4)))*({0 0 0 1, 0 0 0 0,
0 0 0 0, 0 0 0 0});
I21=((0.5*trace(I21a)-0.5*trace(2*PHI*Q21-PHI*P2*PHI*P1)))*({0 0 0 0, 1 0 0 0,
0 0 0 0, 0 0 0 0});
I22=((0.5*trace(I22a)-0.5*trace(2*PHI*Q22-PHI*P2*PHI*P2)))*({0 0 0 0, 0 1 0 0,
0 0 0 0, 0 0 0 0});
I23=((0.5*trace(I23a)-0.5*trace(2*PHI*Q23-PHI*P2*PHI*P3)))*({0 0 0 0, 0 0 1 0,
0 0 0 0, 0 0 0 0});
I24=((0.5*trace(I24a)-0.5*trace(2*PHI*Q24-PHI*P2*PHI*P4)))*({0 0 0 0, 0 0 0 1,
0 0 0 0, 0 0 0 0});
I31=((0.5*trace(I31a)-0.5*trace(2*PHI*Q31-PHI*P3*PHI*P1)))*({0 0 0 0, 0 0 0 0,
1 0 0 0, 0 0 0 0});
I32=((0.5*trace(I32a)-0.5*trace(2*PHI*Q32-PHI*P3*PHI*P2)))*({0 0 0 0, 0 0 0 0,
0 1 0 0, 0 0 0 0});
I33=((0.5*trace(I33a)-0.5*trace(2*PHI*Q33-PHI*P3*PHI*P3)))*({0 0 0 0, 0 0 0 0,
0 0 1 0, 0 0 0 0});
I34=((0.5*trace(I34a)-0.5*trace(2*PHI*Q34-PHI*P3*PHI*P4)))*({0 0 0 0, 0 0 0 0,
0 0 0 1, 0 0 0 0});
I41=((0.5*trace(I41a)-0.5*trace(2*PHI*Q41-PHI*P4*PHI*P1)))*({0 0 0 0, 0 0 0 0,
0 0 0 0, 1 0 0 0});
I42=((0.5*trace(I42a)-0.5*trace(2*PHI*Q42-PHI*P4*PHI*P2)))*({0 0 0 0, 0 0 0 0,
0 0 0 0, 0 1 0 0});
I43=((0.5*trace(I43a)-0.5*trace(2*PHI*Q43-PHI*P4*PHI*P3)))*({0 0 0 0, 0 0 0 0,
0 0 0 0, 0 0 1 0});
I44=((0.5*trace(I44a)-0.5*trace(2*PHI*Q44-PHI*P4*PHI*P4)))*({0 0 0 0, 0 0 0 0,
0 0 0 0, 0 0 0 1});
I=ginv(I11+I12+I13+I14+I21+I22+I23+I24+I31+I32+I33+I34+I41+I42+I43+I44);
*Inverse of Expected Information Matrix;
*W Terms;
w11=({1 0 0 0})*I*({1 0 0 0})`;
w12=({1 0 0 0})*I*({0 1 0 0})`;
w13=({1 0 0 0})*I*({0 0 1 0})`;
w14=({1 0 0 0})*I*({0 0 0 1})`;
w21=({0 1 0 0})*I*({1 0 0 0})`;
w22=({0 1 0 0})*I*({0 1 0 0})`;
w23=({0 1 0 0})*I*({0 0 1 0})`;
w24=({0 1 0 0})*I*({0 0 0 1})`;
w31=({0 0 1 0})*I*({1 0 0 0})`;
w32=({0 0 1 0})*I*({0 1 0 0})`;
/*
```

```
*/
w33=({0 0 1 0})*I*({0 0 1 0})`;
w34=({0 0 1 0})*I*({0 0 0 1})`;
w41=({0 0 0 1})*I*({1 0 0 0})`;
w42=({0 0 0 1})*I*({0 1 0 0})`;
w43=({0 0 0 1})*I*({0 0 1 0})`;
w44=({0 0 0 1})*I*({0 0 0 1})`;
*Sum Terms;
c11=w11*(Q11-(P1*PHI*P1));
c12=w12*(Q12-(P1*PHI*P2));
c13=w13*(Q13-(P1*PHI*P3));
c14=w14*(Q14-(P1*PHI*P4));
c21=w21*(Q21-(P2*PHI*P1));
c22=w22*(Q22-(P2*PHI*P2));
c23=w23*(Q23-(P2*PHI*P3));
c24=w24*(Q24-(P2*PHI*P4));
c31=w31*(Q31-(P3*PHI*P1));
c32=w32*(Q32-(P3*PHI*P2));
c33=w33*(Q33-(P3*PHI*P3));
c34=w34*(Q34-(P3*PHI*P4));
c41=w41*(Q41-(P4*PHI*P1));
c42=w42*(Q42-(P4*PHI*P2));
c43=w43*(Q43-(P4*PHI*P3));
c44=w44*(Q44-(P4*PHI*P4));
sum=c11+c12+c13+c14+c21+c22+c23+c24+c31+c32+c33+c34+c41+c42+c43+c44;
PHIA=PHI+2*PHI*(sum)*PHI;
Astar=2*PHI*(sum)*PHI;
*Theta denoted as O in this code;
      O=C*ginv(C`*PHI*C)*C`;
*Traces;
*A1;
ta11=w11*trace(O*PHI*P1*PHI)*trace(O*PHI*P1*PHI);
ta12=w12*trace(O*PHI*P1*PHI)*trace(O*PHI*P2*PHI);
ta13=w13*trace(O*PHI*P1*PHI)*trace(O*PHI*P3*PHI);
ta14=w14*trace(O*PHI*P1*PHI)*trace(O*PHI*P4*PHI);
ta21=w21*trace(O*PHI*P2*PHI)*trace(O*PHI*P1*PHI);
ta22=w22*trace(O*PHI*P2*PHI)*trace(O*PHI*P2*PHI);
ta23=w23*trace(O*PHI*P2*PHI)*trace(O*PHI*P3*PHI);
ta24=w24*trace(O*PHI*P2*PHI)*trace(O*PHI*P4*PHI);
ta31=w31*trace(O*PHI*P3*PHI)*trace(O*PHI*P1*PHI);
ta32=w32*trace(O*PHI*P3*PHI)*trace(O*PHI*P2*PHI);
ta33=w33*trace(O*PHI*P3*PHI)*trace(O*PHI*P3*PHI);
ta34=w34*trace(O*PHI*P3*PHI)*trace(O*PHI*P4*PHI);
ta41=w41*trace(O*PHI*P4*PHI)*trace(O*PHI*P1*PHI);
ta42=w42*trace(O*PHI*P4*PHI)*trace(O*PHI*P2*PHI);
ta43=w43*trace(O*PHI*P4*PHI)*trace(O*PHI*P3*PHI);
ta44=w44*trace(O*PHI*P4*PHI)*trace(O*PHI*P4*PHI);
A1=ta11+ta12+ta13+ta14+ta21+ta22+ta23+ta24+ta31+ta32+ta33+ta34+ta41+ta42+ta43+t
a44;
*A2;
tb11=w11*trace(O*PHI*P1*PHI*O*PHI*P1*PHI);
tb12=w12*trace(O*PHI*P1*PHI*O*PHI*P2*PHI);
tb13=w13*trace(O*PHI*P1*PHI*O*PHI*P3*PHI);
tb14=w14*trace(O*PHI*P1*PHI*O*PHI*P4*PHI);
tb21=w21*trace(O*PHI*P2*PHI*O*PHI*P1*PHI);
tb22=w22*trace(O*PHI*P2*PHI*O*PHI*P2*PHI);
tb23=w23*trace(O*PHI*P2*PHI*O*PHI*P3*PHI);
tb24=w24*trace(O*PHI*P2*PHI*O*PHI*P4*PHI);
tb31=w31*trace(O*PHI*P3*PHI*O*PHI*P1*PHI);
tb32=w32*trace(O*PHI*P3*PHI*O*PHI*P2*PHI);
tb33=w33*trace(O*PHI*P3*PHI*O*PHI*P3*PHI);
/*
```

```
*/
tb34=w34*trace(O*PHI*P3*PHI*O*PHI*P4*PHI);
tb41=w41*trace(O*PHI*P4*PHI*O*PHI*P1*PHI);
tb42=w42*trace(O*PHI*P4*PHI*O*PHI*P2*PHI);
tb43=w43*trace(O*PHI*P4*PHI*O*PHI*P3*PHI);
tb44=w44*trace(O*PHI*P4*PHI*O*PHI*P4*PHI);
A2=tb11+tb12+tb13+tb14+tb21+tb22+tb23+tb24+tb31+tb32+tb33+tb34+tb41+tb42+tb43+t
b44;
*Mij component of E(F_NonCentral) terms;
M11=%Mij(i=1,j=1);
M12=%Mij(i=1,j=2);
M13=%Mij(i=1,j=3);
M14=%Mij(i=1,j=4);
M21=%Mij(i=2,j=1);
M22=%Mij(i=2,j=2);
M23=%Mij(i=2,j=3);
M24=%Mij(i=2,j=4);
M31=%Mij(i=3,j=1);
M32=%Mij(i=3,j=2);
M33=%Mij(i=3,j=3);
M34=%Mij(i=3,j=4);
M41=%Mij(i=4,j=1);
M42=%Mij(i=4,j=2);
M43=%Mij(i=4,j=3);
M44=%Mij(i=4,j=4);
*A3 component of E(F) and Var(F);
tc11=w11*B`*M11*B;
tc12=w12*B`*M12*B;
tc13=w13*B`*M13*B;
tc14=w14*B`*M14*B;
tc21=w21*B`*M21*B;
tc22=w22*B`*M22*B;
tc23=w23*B`*M23*B;
tc24=w24*B`*M24*B;
tc31=w31*B`*M31*B;
tc32=w32*B`*M32*B;
tc33=w33*B`*M33*B;
tc34=w34*B`*M34*B;
tc41=w41*B`*M41*B;
tc42=w42*B`*M42*B;
tc43=w43*B`*M43*B;
tc44=w44*B`*M44*B;
A3=0.5*(tc11+tc12+tc13+tc14+tc21+tc22+tc23+tc24+tc31+tc32+tc33+tc34+tc41+tc42+t
c43+tc44);
L=trace(ginv(C)*C);
*DF components;
E=inv(1-(A2/L));
B=(1/(2*L))*(A1+(6*A2));
gg=((L+1)*A1-(L+4)*A2)/((L+2)*A2);
dd1=gg/(3*L+2*(1-gg));
dd2=(L-gg)/(3*L+2*(1-gg));
dd3=(L-gg+2)/(3*L+2*(1-gg));
V=(2/L)*((1+(dd1*B))/((1-(dd2*B))*(1-(dd2*B))*(1-(dd3*B))));
rr=V/(2*E*E);
B=Beta;
*Degrees of Freedom M and Scale Factor Lambda;
m=4+((L+2)/((L*rr)-1));
Lambda=m/(E*(m-2));
*Expected Value;
E_null=(1/L)*(L+A2);
/*
```

```
*/
E_alt=E_null + (1/L)*(B`*C*ginv(C`*PHI*C)*C`*B+A3)-
(1/L)*(B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B);
ncp=lambda*(B`*C*ginv(C`*PHIA*C)*C`*B);
rat1=(lambda*E_null)/(m/(m-2));
ncpform=B`*C*ginv(C`*PHI*C)*C`*B;
E0=ginv(1-(A2/L)); *KR adjustment to E_null;
EA=(1/L)*ginv(ncpform-
A3+B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B)*((ncpform-
B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B)*(ncpform-
B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B));
ncpnew=L*Lambda*(EA/E0);
ncp1=ncpnew;
ncp2=ncp;
ncp3=ncpform;
ncp4=ncpform;
ncp5=ncpform;
ncpuse=&ncpuse;
alpha=&alpha;
alpha2=1-&alpha;
zscorea=quantile("Normal", alpha2, 0, 1);
despower=&despower;
zscoreb=quantile("Normal", despower, 0, 1);
zscore=zscorea+zscoreb;
QFC32=(abs(zscore)*sqrt(2/(9*L))+(1-(2/(9*L))))**3; *K in dissertation
quadratic formula;
QFa2=L/3; *a in dissertation quadratic formula;
QFb2=L+2*10000-(4/3)*QFC32*L-2; *b in dissertation quadratic formula;
QFc2=-2*QFC32*10000; *c in dissertation quadratic formula;
Fscore=(-QFb2+sqrt(QFb2*QFb2-4*QFa2*QFc2))/(2*QFa2);
multfactor=((Fscore))*(1/ncpuse)*L;
scalera=B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B;
L=trace(ginv(C)*C);
*DF components;
E=inv(1-(A2/L));
B=(1/(2*L))*(A1+(6*A2));
gg=((L+1)*A1-(L+4)*A2)/((L+2)*A2);
dd1=gg/(3*L+2*(1-gg));
dd2=(L-gg)/(3*L+2*(1-gg));
dd3=(L-gg+2)/(3*L+2*(1-gg));
V=(2/L)*((1+(dd1*B))/((1-(dd2*B))*(1-(dd2*B))*(1-(dd3*B))));
rr=V/(2*E*E);
*Degrees of Freedom M and Scale Factor Lambda;
m=4+((L+2)/((L*rr)-1));
Lambda=m/(E*(m-2));
*Power;
alpha=0.05;
alval=1-0.05;
timevec=({&timevec});
m1=m;
m2=m;
m3=m;
m4=(&n1+&n2+&n3)*(((nrow(timevec)-1)*&missparm+1)/nrow(timevec))-L-1;
m5=(&n1+&n2+&n3)*(((nrow(timevec)-
1)*&missparm+1)/nrow(timevec))*(nrow(timevec))-6;
muse=&ddfuse;
critval=finv(alval, L, muse); *So p(Wald<w) = p((1/lambda)*F < f);
power=1-probf(critval, L, muse, ncpuse);
ncp_dev=ncpnew+0.000001;
power_dev=((1-probf(critval, L, m, ncp_dev))-(1-probf(critval, L, m,
ncpnew)))/0.000001;
G=1000*scalera*power_dev;
/*
```

```
*/
R2=1-inv(1+(1/L)*Lambda*ncp*(L/m));
outvec=power||multfactor||G;
create work.power_a from outvec;
append from outvec;
quit;

data work.power_b;
       set work.power_b work.power_a;
run;
%mend power;

%macro average;
       %do designs=1 %to &reps;
              %power(rep=&designs);
       %end;
%mend average;

%average; run;

proc sql;
       create table work.power_c as
              select COL1 as Power, COL2 as multfactor, Col3 as G
              from work.power_b
              where COL1 ne 9999;
quit;

proc sql;
       create table work.AveragePower as
              select mean(Power) as Power, mean(multfactor) as multfactor,
mean(G) as G
              from work.power_c;
quit;


proc sql;
       create table work.macrostarta as
              select Power, multfactor, G, ceil(&n1*multfactor) as n1,
ceil(&n2*multfactor) as n2, ceil(&n3*multfactor) as n3,

                                          case when &missparm*1=1 then 1

                                                        when &missparm*1<1 and
G<=1 then 25

                                                        when &missparm*1<1 and
1<G<=2.5 then 50

                                                        when &missparm*1<1 and
2.5<G<=5 then 100

                                                        when &missparm*1<1 and
5<G<=7.5 then 150

                                                        when &missparm*1<1 and
7.5<G<=10 then 200
/*
```

```
*/

                                                        when &missparm*1<1 and
G>10 then 250


                                                        else . end as reps
             from work.AveragePower;
quit;

data work.MacroValsglobal;
       set work.macrostarta;
       call symput('reps', reps);
run;

data work.MacroVals;
       set work.macrostarta;
       call symput('n1', n1);
       call symput('n2', n2);
       call symput('n3', n3);
run;

%let power=0;

%macro m(n1=&q1,n2=&q2,n3=&q3);
%do %while(&power<&despower);
data work.Treatment;
       seed=&misseed1+1;
       reps=&reps;
       do rep=1 to reps;
        do n=1 to &n1;
        ID=compress(put(n,best8.)||"A");
               do Time=min(&timevec) to max(&timevec);
                     z=ranuni(seed); /*Uses uniform(0,1) random variable "z" to
assign missing values if z>missparm, modified later to force baseline as
nonmissing*/
                     if  z>&missparm then u=0;
                     else if z<=&missparm then u=1;
                     output;
               end;
         end;
        end;
        drop seed z;
run;

proc sql;
       create table work.Treatobs as
              select rep, ID, Time, case Time when . then 1 else 1 end as
intercept,
                                 case Time when . then 1 else 1 end as Treatment,
                           case Time when . then 0 else 0 end as Comparator, u
              from work.Treatment
              where Time in (&timevec);
quit;

data work.Placebo;
       seed=&misseed2+1;
       reps=&reps;
   do rep=1 to reps;
        do n=1 to &n2;
        ID=compress(put(n,best8.)||"B");
/*
```

```
*/
              do time=min(&timevec) to max(&timevec);
                    z=ranuni(seed); /*Uses uniform(0,1) random variable "z" to
assign missing values if z>missparm, modified later to force baseline as
nonmissing*/
                    if  z>&missparm then u=0;
                    else if z<=&missparm then u=1;
                    output;
              end;
         end;
        end;
        drop seed z;
run;

proc sql;
       create table work.Plbobs as
             select rep, ID, Time, case Time when . then 1 else 1 end as
intercept,
                           case Time when . then 0 else 0 end as Treatment,
                           case Time when . then 0 else 0 end as Comparator, u
             from work.Placebo
             where Time in (&timevec);
quit;

data work.Comparator;
       seed=&misseed3+1;
       reps=&reps;
   do rep=1 to reps;
        do n=1 to &n3;
        ID=compress(put(n,best8.)||"C");
              do time=min(&timevec) to max(&timevec);
                    z=ranuni(seed); /*Uses uniform(0,1) random variable "z" to
assign missing values if z>missparm, modified later to force baseline as
nonmissing*/
                    if  z>&missparm then u=0;
                    else if z<=&missparm then u=1;
                    output;
              end;
         end;
        end;
        drop seed z;
run;

proc sql;
       create table work.Cmptorobs as
             select rep, ID, Time, case Time when . then 1 else 1 end as
intercept,
                           case Time when . then 0 else 0 end as Treatment,
                           case Time when . then 1 else 1 end as Comparator, u
             from work.Comparator
             where Time in (&timevec);
quit;

data work.MixedData_a; *Combine placebo and treatment designs into one dataset;
       set work.Treatobs work.Plbobs work.Cmptorobs;
run;

proc sql; *Will create variable "M" to serve as missing data indicator and make
sure no baseline values are missing;
       create table work.MixedData_b as
             select rep, ID, intercept, Treatment, Comparator, Time,
/*
```

```
*/Treatment*Time as TreatTime, Comparator*Time as CompTime,
                              case when u=0 and time ne min(&timevec) then . else
1 end as M
              from work.MixedData_a;
quit;

proc sql;
       create table work.MixedData as
              select rep, ID, intercept, Treatment, Comparator, Time, TreatTime,
CompTime
              from work.MixedData_b
              where M ne .;
quit;

proc sort data=work.MixedData;
       by rep id time;
run;
quit;

data work.power_b;
       COL1=9999;
run;

%macro power(rep=&sim);
proc sql;
       create table work.mixeddata_&rep as
              select *
              from work.MixedData
              where rep=&rep;
quit;
proc sort data=work.mixeddata_&rep;
       by rep id time;
run;
quit;
data work.seqids_&rep;
       set work.MixedData_&rep;
       by id;
       retain order 0;
              if first.id then order=order+1;
run;
ods select all;
proc iml;
Contrasts=({0 0 0 0 1 0, 0 0 0 0 0 1})`;
C=Contrasts;
L=trace(ginv(C)*C);
q=L+1; *+1 if to account for intercept;
ivar=&ranistd;
svar=&ransstd;
corr=&correlation;
errvar=&errdev*&errdev;
cov=(ivar*({1 0, 0 0})+svar*({0 0, 0 1}))*(corr*({0 1, 1 0})+I(2))*(ivar*({1 0,
0 0})+svar*({0 0, 0 1}));
bint=&xbar2;
btreat=&xbar1-&xbar2; *treatment effect;
bcompare=&xbar3-&xbar2;
btime=&btime2; *time effect in reference group 2;
btreattime=&btime1-&btime2; *group*time interaction;
bcomptime=&btime3-&btime2;
Beta=(bint||btreat||bcompare||btime||btreattime||bcomptime)`;
B=Beta;
n=&n1+&n2+&n3;
/*
```

```
*/
use work.seqids_&rep;
       do i=1 to n;
         read all var {intercept Treatment Comparator Time TreatTime CompTime}
where(order=i) into X;
         read all var {intercept Time} where(order=i) into Z;
         S=z*cov*z`+I((nrow(X)))@errvar;
         if i=1 then V=(X`*inv(S)*X);
     else V=V+(X`*inv(S)*X);
end;
PHI=inv(V);
use work.seqids_&rep;;
       do i=1 to n;
         read all var {intercept Treatment Comparator Time TreatTime CompTime}
where(order=i) into X;
         read all var {intercept Time} where(order=i) into Z;
         S=z*cov*z`+I((nrow(X)))@errvar;
         d1=z*({1 0, 0 0})*z`;
         d2=z*({0 0, 0 1})*z`;
         d3=I(nrow(X));
         d4=z*({0 1, 1 0})*z`;
         if i=1 then P1=-X`*inv(S)*d1*inv(S)*X; else P1=P1-
X`*inv(S)*d1*inv(S)*X;
         if i=1 then P2=-X`*inv(S)*d2*inv(S)*X; else P2=P2-
X`*inv(S)*d2*inv(S)*X;
         if i=1 then P3=-X`*inv(S)*d3*inv(S)*X; else P3=P3-
X`*inv(S)*d3*inv(S)*X;
         if i=1 then P4=-X`*inv(S)*d4*inv(S)*X; else P4=P4-
X`*inv(S)*d4*inv(S)*X;
         if i=1 then Q11=X`*inv(S)*d1*inv(S)*d1*inv(S)*X; else
Q11=Q11+X`*inv(S)*d1*inv(S)*d1*inv(S)*X;
         if i=1 then Q12=X`*inv(S)*d1*inv(S)*d2*inv(S)*X; else
Q12=Q12+X`*inv(S)*d1*inv(S)*d2*inv(S)*X;
         if i=1 then Q13=X`*inv(S)*d1*inv(S)*d3*inv(S)*X; else
Q13=Q13+X`*inv(S)*d1*inv(S)*d3*inv(S)*X;
         if i=1 then Q14=X`*inv(S)*d1*inv(S)*d4*inv(S)*X; else
Q14=Q14+X`*inv(S)*d1*inv(S)*d4*inv(S)*X;
         if i=1 then Q21=X`*inv(S)*d2*inv(S)*d1*inv(S)*X; else
Q21=Q21+X`*inv(S)*d2*inv(S)*d1*inv(S)*X;
         if i=1 then Q22=X`*inv(S)*d2*inv(S)*d2*inv(S)*X; else
Q22=Q22+X`*inv(S)*d2*inv(S)*d2*inv(S)*X;
         if i=1 then Q23=X`*inv(S)*d2*inv(S)*d3*inv(S)*X; else
Q23=Q23+X`*inv(S)*d2*inv(S)*d3*inv(S)*X;
         if i=1 then Q24=X`*inv(S)*d2*inv(S)*d4*inv(S)*X; else
Q24=Q24+X`*inv(S)*d2*inv(S)*d4*inv(S)*X;
         if i=1 then Q31=X`*inv(S)*d3*inv(S)*d1*inv(S)*X; else
Q31=Q31+X`*inv(S)*d3*inv(S)*d1*inv(S)*X;
         if i=1 then Q32=X`*inv(S)*d3*inv(S)*d2*inv(S)*X; else
Q32=Q32+X`*inv(S)*d3*inv(S)*d2*inv(S)*X;
         if i=1 then Q33=X`*inv(S)*d3*inv(S)*d3*inv(S)*X; else
Q33=Q33+X`*inv(S)*d3*inv(S)*d3*inv(S)*X;
         if i=1 then Q34=X`*inv(S)*d3*inv(S)*d4*inv(S)*X; else
Q34=Q34+X`*inv(S)*d3*inv(S)*d4*inv(S)*X;
         if i=1 then Q41=X`*inv(S)*d4*inv(S)*d1*inv(S)*X; else
Q41=Q41+X`*inv(S)*d4*inv(S)*d1*inv(S)*X;
         if i=1 then Q42=X`*inv(S)*d4*inv(S)*d2*inv(S)*X; else
Q42=Q42+X`*inv(S)*d4*inv(S)*d2*inv(S)*X;
         if i=1 then Q43=X`*inv(S)*d4*inv(S)*d3*inv(S)*X; else
Q43=Q43+X`*inv(S)*d4*inv(S)*d3*inv(S)*X;
         if i=1 then Q44=X`*inv(S)*d4*inv(S)*d4*inv(S)*X; else
Q44=Q44+X`*inv(S)*d4*inv(S)*d4*inv(S)*X;
/*
```

```
*/
*Information Matrices;
         if i=1 then I11a=trace(inv(S)*d1*inv(S)*d1); else
I11a=trace(I11a)+trace(inv(S)*d1*inv(S)*d1);
         if i=1 then I12a=trace(inv(S)*d1*inv(S)*d2); else
I12a=trace(I12a)+trace(inv(S)*d1*inv(S)*d2);
         if i=1 then I13a=trace(inv(S)*d1*inv(S)*d3); else
I13a=trace(I13a)+trace(inv(S)*d1*inv(S)*d3);
         if i=1 then I14a=trace(inv(S)*d1*inv(S)*d4); else
I14a=trace(I14a)+trace(inv(S)*d1*inv(S)*d4);
         if i=1 then I21a=trace(inv(S)*d2*inv(S)*d1); else
I21a=trace(I21a)+trace(inv(S)*d2*inv(S)*d1);
         if i=1 then I22a=trace(inv(S)*d2*inv(S)*d2); else
I22a=trace(I22a)+trace(inv(S)*d2*inv(S)*d2);
         if i=1 then I23a=trace(inv(S)*d2*inv(S)*d3); else
I23a=trace(I23a)+trace(inv(S)*d2*inv(S)*d3);
         if i=1 then I24a=trace(inv(S)*d2*inv(S)*d4); else
I24a=trace(I24a)+trace(inv(S)*d2*inv(S)*d4);
         if i=1 then I31a=trace(inv(S)*d3*inv(S)*d1); else
I31a=trace(I31a)+trace(inv(S)*d3*inv(S)*d1);
         if i=1 then I32a=trace(inv(S)*d3*inv(S)*d2); else
I32a=trace(I32a)+trace(inv(S)*d3*inv(S)*d2);
         if i=1 then I33a=trace(inv(S)*d3*inv(S)*d3); else
I33a=trace(I33a)+trace(inv(S)*d3*inv(S)*d3);
         if i=1 then I34a=trace(inv(S)*d3*inv(S)*d4); else
I34a=trace(I34a)+trace(inv(S)*d3*inv(S)*d4);
         if i=1 then I41a=trace(inv(S)*d4*inv(S)*d1); else
I41a=trace(I41a)+trace(inv(S)*d4*inv(S)*d1);
         if i=1 then I42a=trace(inv(S)*d4*inv(S)*d2); else
I42a=trace(I42a)+trace(inv(S)*d4*inv(S)*d2);
         if i=1 then I43a=trace(inv(S)*d4*inv(S)*d3); else
I43a=trace(I43a)+trace(inv(S)*d4*inv(S)*d3);
         if i=1 then I44a=trace(inv(S)*d4*inv(S)*d4); else
I44a=trace(I44a)+trace(inv(S)*d4*inv(S)*d4);
end;
I11=((0.5*trace(I11a)-0.5*trace(2*PHI*Q11-PHI*P1*PHI*P1)))*({1 0 0 0, 0 0 0 0,
0 0 0 0, 0 0 0 0});
I12=((0.5*trace(I12a)-0.5*trace(2*PHI*Q12-PHI*P1*PHI*P2)))*({0 1 0 0, 0 0 0 0,
0 0 0 0, 0 0 0 0});
I13=((0.5*trace(I13a)-0.5*trace(2*PHI*Q13-PHI*P1*PHI*P3)))*({0 0 1 0, 0 0 0 0,
0 0 0 0, 0 0 0 0});
I14=((0.5*trace(I14a)-0.5*trace(2*PHI*Q14-PHI*P1*PHI*P4)))*({0 0 0 1, 0 0 0 0,
0 0 0 0, 0 0 0 0});
I21=((0.5*trace(I21a)-0.5*trace(2*PHI*Q21-PHI*P2*PHI*P1)))*({0 0 0 0, 1 0 0 0,
0 0 0 0, 0 0 0 0});
I22=((0.5*trace(I22a)-0.5*trace(2*PHI*Q22-PHI*P2*PHI*P2)))*({0 0 0 0, 0 1 0 0,
0 0 0 0, 0 0 0 0});
I23=((0.5*trace(I23a)-0.5*trace(2*PHI*Q23-PHI*P2*PHI*P3)))*({0 0 0 0, 0 0 1 0,
0 0 0 0, 0 0 0 0});
I24=((0.5*trace(I24a)-0.5*trace(2*PHI*Q24-PHI*P2*PHI*P4)))*({0 0 0 0, 0 0 0 1,
0 0 0 0, 0 0 0 0});
I31=((0.5*trace(I31a)-0.5*trace(2*PHI*Q31-PHI*P3*PHI*P1)))*({0 0 0 0, 0 0 0 0,
1 0 0 0, 0 0 0 0});
I32=((0.5*trace(I32a)-0.5*trace(2*PHI*Q32-PHI*P3*PHI*P2)))*({0 0 0 0, 0 0 0 0,
0 1 0 0, 0 0 0 0});
I33=((0.5*trace(I33a)-0.5*trace(2*PHI*Q33-PHI*P3*PHI*P3)))*({0 0 0 0, 0 0 0 0,
0 0 1 0, 0 0 0 0});
I34=((0.5*trace(I34a)-0.5*trace(2*PHI*Q34-PHI*P3*PHI*P4)))*({0 0 0 0, 0 0 0 0,
0 0 0 1, 0 0 0 0});
I41=((0.5*trace(I41a)-0.5*trace(2*PHI*Q41-PHI*P4*PHI*P1)))*({0 0 0 0, 0 0 0 0,
0 0 0 0, 1 0 0 0});
/*
```

```
*/
I42=((0.5*trace(I42a)-0.5*trace(2*PHI*Q42-PHI*P4*PHI*P2)))*({0 0 0 0, 0 0 0 0,
0 0 0 0, 0 1 0 0});
I43=((0.5*trace(I43a)-0.5*trace(2*PHI*Q43-PHI*P4*PHI*P3)))*({0 0 0 0, 0 0 0 0,
0 0 0 0, 0 0 1 0});
I44=((0.5*trace(I44a)-0.5*trace(2*PHI*Q44-PHI*P4*PHI*P4)))*({0 0 0 0, 0 0 0 0,
0 0 0 0, 0 0 0 1});
I=ginv(I11+I12+I13+I14+I21+I22+I23+I24+I31+I32+I33+I34+I41+I42+I43+I44);
*Inverse of Expected Information Matrix;
*W Terms;
w11=({1 0 0 0})*I*({1 0 0 0})`;
w12=({1 0 0 0})*I*({0 1 0 0})`;
w13=({1 0 0 0})*I*({0 0 1 0})`;
w14=({1 0 0 0})*I*({0 0 0 1})`;
w21=({0 1 0 0})*I*({1 0 0 0})`;
w22=({0 1 0 0})*I*({0 1 0 0})`;
w23=({0 1 0 0})*I*({0 0 1 0})`;
w24=({0 1 0 0})*I*({0 0 0 1})`;
w31=({0 0 1 0})*I*({1 0 0 0})`;
w32=({0 0 1 0})*I*({0 1 0 0})`;
w33=({0 0 1 0})*I*({0 0 1 0})`;
w34=({0 0 1 0})*I*({0 0 0 1})`;
w41=({0 0 0 1})*I*({1 0 0 0})`;
w42=({0 0 0 1})*I*({0 1 0 0})`;
w43=({0 0 0 1})*I*({0 0 1 0})`;
w44=({0 0 0 1})*I*({0 0 0 1})`;
*Sum Terms;
c11=w11*(Q11-(P1*PHI*P1));
c12=w12*(Q12-(P1*PHI*P2));
c13=w13*(Q13-(P1*PHI*P3));
c14=w14*(Q14-(P1*PHI*P4));
c21=w21*(Q21-(P2*PHI*P1));
c22=w22*(Q22-(P2*PHI*P2));
c23=w23*(Q23-(P2*PHI*P3));
c24=w24*(Q24-(P2*PHI*P4));
c31=w31*(Q31-(P3*PHI*P1));
c32=w32*(Q32-(P3*PHI*P2));
c33=w33*(Q33-(P3*PHI*P3));
c34=w34*(Q34-(P3*PHI*P4));
c41=w41*(Q41-(P4*PHI*P1));
c42=w42*(Q42-(P4*PHI*P2));
c43=w43*(Q43-(P4*PHI*P3));
c44=w44*(Q44-(P4*PHI*P4));
sum=c11+c12+c13+c14+c21+c22+c23+c24+c31+c32+c33+c34+c41+c42+c43+c44;
PHIA=PHI+2*PHI*(sum)*PHI;
Astar=2*PHI*(sum)*PHI;
*Theta denoted as O in this code;
      O=C*ginv(C`*PHI*C)*C`;
*Traces;
*A1;
ta11=w11*trace(O*PHI*P1*PHI)*trace(O*PHI*P1*PHI);
ta12=w12*trace(O*PHI*P1*PHI)*trace(O*PHI*P2*PHI);
ta13=w13*trace(O*PHI*P1*PHI)*trace(O*PHI*P3*PHI);
ta14=w14*trace(O*PHI*P1*PHI)*trace(O*PHI*P4*PHI);
ta21=w21*trace(O*PHI*P2*PHI)*trace(O*PHI*P1*PHI);
ta22=w22*trace(O*PHI*P2*PHI)*trace(O*PHI*P2*PHI);
ta23=w23*trace(O*PHI*P2*PHI)*trace(O*PHI*P3*PHI);
ta24=w24*trace(O*PHI*P2*PHI)*trace(O*PHI*P4*PHI);
ta31=w31*trace(O*PHI*P3*PHI)*trace(O*PHI*P1*PHI);
ta32=w32*trace(O*PHI*P3*PHI)*trace(O*PHI*P2*PHI);
ta33=w33*trace(O*PHI*P3*PHI)*trace(O*PHI*P3*PHI);
/*
```

```
*/
ta34=w34*trace(O*PHI*P3*PHI)*trace(O*PHI*P4*PHI);
ta41=w41*trace(O*PHI*P4*PHI)*trace(O*PHI*P1*PHI);
ta42=w42*trace(O*PHI*P4*PHI)*trace(O*PHI*P2*PHI);
ta43=w43*trace(O*PHI*P4*PHI)*trace(O*PHI*P3*PHI);
ta44=w44*trace(O*PHI*P4*PHI)*trace(O*PHI*P4*PHI);
A1=ta11+ta12+ta13+ta14+ta21+ta22+ta23+ta24+ta31+ta32+ta33+ta34+ta41+ta42+ta43+t
a44;
*A2;
tb11=w11*trace(O*PHI*P1*PHI*O*PHI*P1*PHI);
tb12=w12*trace(O*PHI*P1*PHI*O*PHI*P2*PHI);
tb13=w13*trace(O*PHI*P1*PHI*O*PHI*P3*PHI);
tb14=w14*trace(O*PHI*P1*PHI*O*PHI*P4*PHI);
tb21=w21*trace(O*PHI*P2*PHI*O*PHI*P1*PHI);
tb22=w22*trace(O*PHI*P2*PHI*O*PHI*P2*PHI);
tb23=w23*trace(O*PHI*P2*PHI*O*PHI*P3*PHI);
tb24=w24*trace(O*PHI*P2*PHI*O*PHI*P4*PHI);
tb31=w31*trace(O*PHI*P3*PHI*O*PHI*P1*PHI);
tb32=w32*trace(O*PHI*P3*PHI*O*PHI*P2*PHI);
tb33=w33*trace(O*PHI*P3*PHI*O*PHI*P3*PHI);
tb34=w34*trace(O*PHI*P3*PHI*O*PHI*P4*PHI);
tb41=w41*trace(O*PHI*P4*PHI*O*PHI*P1*PHI);
tb42=w42*trace(O*PHI*P4*PHI*O*PHI*P2*PHI);
tb43=w43*trace(O*PHI*P4*PHI*O*PHI*P3*PHI);
tb44=w44*trace(O*PHI*P4*PHI*O*PHI*P4*PHI);
A2=tb11+tb12+tb13+tb14+tb21+tb22+tb23+tb24+tb31+tb32+tb33+tb34+tb41+tb42+tb43+t
b44;
*Mij component of E(F_NonCentral) terms;
M11=%Mij(i=1,j=1);
M12=%Mij(i=1,j=2);
M13=%Mij(i=1,j=3);
M14=%Mij(i=1,j=4);
M21=%Mij(i=2,j=1);
M22=%Mij(i=2,j=2);
M23=%Mij(i=2,j=3);
M24=%Mij(i=2,j=4);
M31=%Mij(i=3,j=1);
M32=%Mij(i=3,j=2);
M33=%Mij(i=3,j=3);
M34=%Mij(i=3,j=4);
M41=%Mij(i=4,j=1);
M42=%Mij(i=4,j=2);
M43=%Mij(i=4,j=3);
M44=%Mij(i=4,j=4);
*A3 component of E(F) and Var(F);
tc11=w11*B`*M11*B;
tc12=w12*B`*M12*B;
tc13=w13*B`*M13*B;
tc14=w14*B`*M14*B;
tc21=w21*B`*M21*B;
tc22=w22*B`*M22*B;
tc23=w23*B`*M23*B;
tc24=w24*B`*M24*B;
tc31=w31*B`*M31*B;
tc32=w32*B`*M32*B;
tc33=w33*B`*M33*B;
tc34=w34*B`*M34*B;
tc41=w41*B`*M41*B;
tc42=w42*B`*M42*B;
tc43=w43*B`*M43*B;
tc44=w44*B`*M44*B;
/*
```

```
*/
A3=0.5*(tc11+tc12+tc13+tc14+tc21+tc22+tc23+tc24+tc31+tc32+tc33+tc34+tc41+tc42+t
c43+tc44);
L=trace(ginv(C)*C);
*DF components;
E=inv(1-(A2/L));
B=(1/(2*L))*(A1+(6*A2));
gg=((L+1)*A1-(L+4)*A2)/((L+2)*A2);
dd1=gg/(3*L+2*(1-gg));
dd2=(L-gg)/(3*L+2*(1-gg));
dd3=(L-gg+2)/(3*L+2*(1-gg));
V=(2/L)*((1+(dd1*B))/((1-(dd2*B))*(1-(dd2*B))*(1-(dd3*B))));
rr=V/(2*E*E);
B=Beta;
*Degrees of Freedom M and Scale Factor Lambda;
m=4+((L+2)/((L*rr)-1));
Lambda=m/(E*(m-2));
*Expected Value;
E_null=(1/L)*(L+A2);
E_alt=E_null + (1/L)*(B`*C*ginv(C`*PHI*C)*C`*B+A3)-
(1/L)*(B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B);
ncp=lambda*(B`*C*ginv(C`*PHIA*C)*C`*B);
rat1=(lambda*E_null)/(m/(m-2));
ncpform=B`*C*ginv(C`*PHI*C)*C`*B;
E0=ginv(1-(A2/L)); *KR adjustment to E_null;
EA=(1/L)*ginv(ncpform-
A3+B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B)*((ncpform-
B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B)*(ncpform-
B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B));
ncpnew=L*Lambda*(EA/E0);
ncp1=ncpnew;
ncp2=ncp;
ncp3=ncpform;
ncp4=ncpform;
ncp5=ncpform;
ncpuse=&ncpuse;
scalera=B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B;
L=trace(ginv(C)*C);
*DF components;
E=inv(1-(A2/L));
B=(1/(2*L))*(A1+(6*A2));
gg=((L+1)*A1-(L+4)*A2)/((L+2)*A2);
dd1=gg/(3*L+2*(1-gg));
dd2=(L-gg)/(3*L+2*(1-gg));
dd3=(L-gg+2)/(3*L+2*(1-gg));
V=(2/L)*((1+(dd1*B))/((1-(dd2*B))*(1-(dd2*B))*(1-(dd3*B))));
rr=V/(2*E*E);
*Degrees of Freedom M and Scale Factor Lambda;
m=4+((L+2)/((L*rr)-1));
Lambda=m/(E*(m-2));
*Power;
alpha=&alpha;
alval=1-alpha;
timevec=({&timevec});
m1=m;
m2=m;
m3=m;
m4=(&n1+&n2+&n3)*(((nrow(timevec)-1)*&missparm+1)/nrow(timevec))-L-1;
m5=(&n1+&n2+&n3)*(((nrow(timevec)-
1)*&missparm+1)/nrow(timevec))*(nrow(timevec))-6;
muse=&ddfuse;
/*
```

```
*/
critval=finv(alval, L, muse); *So p(Wald<w) = p((1/lambda)*F < f);
power=1-probf(critval, L, muse, ncpuse);
ncp_dev=ncpnew+0.000001;
power_dev=((1-probf(critval, L, m, ncp_dev))-(1-probf(critval, L, m,
ncpnew)))/0.000001;
R2=1-inv(1+(1/L)*Lambda*ncp*(L/m));
create work.power_a from power;
append from power;
quit;

data work.power_b;
	set work.power_b work.power_a;
run;
%mend power;

%macro average;
	%do designs=1 %to &reps;
		%power(rep=&designs);
	%end;
%mend average;

%average; run;

proc sql;
	create table work.power_c as
		select COL1 as Power
		from work.power_b
		where COL1 ne 9999;
quit;

proc sql;
	create table work.AveragePower as
		select count(power) as Designs, mean(Power) as Power, std(power)
as DevPower, min(power) as MinPower, max(power) as maxPower,
				1*&n1 as n1final, 1*&n2 as n2final, 1*&n3 as n3final
		from work.power_c;
quit;
data work.MacroVals;
	set work.AveragePower;
	n1=round(&n1+&g1ratio);
	n2=round(&n2+&g2ratio);
	n3=round(&n3+&g3ratio);
	call symput('Power', Power);
	call symput('n1', n1);
	call symput('n2', n2);
	call symput('n3', n3);
run;
%end;
%mend m;
%m(n1=&n1,n2=&n2,n3=&n3); run;

proc sql;
	create table work.SampleSize1 as
		select n1final as n1, n2final as n2, n3final as n3, Power as
Power, Power as nominal_power
		from work.AveragePower;
quit;

data work.MacroVals;
	set work.SampleSize1;
/*
```

```
*/
       call symput('Power', Power);
       call symput('n1', n1);
       call symput('n2', n2);
       call symput('n3', n3);
run;

%macro m2(n1=&q1,n2=&q2,n3=&q3);
%do %while(&power>&despower);
data work.Treatment;
       seed=&misseed1+1;
       reps=&reps;
       do rep=1 to reps;
        do n=1 to &n1;
        ID=compress(put(n,best8.)||"A");
              do Time=min(&timevec) to max(&timevec);
                    z=ranuni(seed); /*Uses uniform(0,1) random variable "z" to
assign missing values if z>missparm, modified later to force baseline as
nonmissing*/
                    if  z>&missparm then u=0;
                    else if z<=&missparm then u=1;
                    output;
              end;
         end;
        end;
        drop seed z;
run;

proc sql;
       create table work.Treatobs as
             select rep, ID, Time, case Time when . then 1 else 1 end as
intercept,
                           case Time when . then 1 else 1 end as Treatment,
                     case Time when . then 0 else 0 end as Comparator, u
             from work.Treatment
             where Time in (&timevec);
quit;

data work.Placebo;
       seed=&misseed2+1;
       reps=&reps;
   do rep=1 to reps;
        do n=1 to &n2;
        ID=compress(put(n,best8.)||"B");
              do time=min(&timevec) to max(&timevec);
                    z=ranuni(seed); /*Uses uniform(0,1) random variable "z" to
assign missing values if z>missparm, modified later to force baseline as
nonmissing*/
                    if  z>&missparm then u=0;
                    else if z<=&missparm then u=1;
                    output;
              end;
         end;
        end;
        drop seed z;
run;

proc sql;
       create table work.Plbobs as
             select rep, ID, Time, case Time when . then 1 else 1 end as
intercept,
/*
```

```
*/
                              case Time when . then 0 else 0 end as Treatment,
                              case Time when . then 0 else 0 end as Comparator, u
              from work.Placebo
              where Time in (&timevec);
quit;

data work.Comparator;
       seed=&misseed3+1;
       reps=&reps;
   do rep=1 to reps;
        do n=1 to &n3;
        ID=compress(put(n,best8.)||"C");
               do time=min(&timevec) to max(&timevec);
                     z=ranuni(seed); /*Uses uniform(0,1) random variable "z" to
assign missing values if z>missparm, modified later to force baseline as
nonmissing*/
                     if  z>&missparm then u=0;
                     else if z<=&missparm then u=1;
                     output;
               end;
         end;
        end;
        drop seed z;
run;

proc sql;
       create table work.Cmptorobs as
              select rep, ID, Time, case Time when . then 1 else 1 end as
intercept,
                              case Time when . then 0 else 0 end as Treatment,
                              case Time when . then 1 else 1 end as Comparator, u
              from work.Comparator
              where Time in (&timevec);
quit;

data work.MixedData_a; *Combine placebo and treatment designs into one dataset;
       set work.Treatobs work.Plbobs work.Cmptorobs;
run;

proc sql; *Will create variable "M" to serve as missing data indicator and make
sure no baseline values are missing;
       create table work.MixedData_b as
              select rep, ID, intercept, Treatment, Comparator, Time,
Treatment*Time as TreatTime, Comparator*Time as CompTime,
                              case when u=0 and time ne min(&timevec) then . else
1 end as M
              from work.MixedData_a;
quit;

proc sql;
       create table work.MixedData as
              select rep, ID, intercept, Treatment, Comparator, Time, TreatTime,
CompTime
              from work.MixedData_b
              where M ne .;
quit;

proc sort data=work.MixedData;
       by rep id time;
run;
/*
```

```
*/
quit;

data work.power_b;
       COL1=9999;
run;

%macro power(rep=&sim);
proc sql;
       create table work.mixeddata_&rep as
              select *
              from work.MixedData
              where rep=&rep;
quit;
proc sort data=work.mixeddata_&rep;
       by rep id time;
run;
quit;
data work.seqids_&rep;
       set work.MixedData_&rep;
       by id;
       retain order 0;
              if first.id then order=order+1;
run;
ods select all;
proc iml;
Contrasts=({0 0 0 0 1 0, 0 0 0 0 0 1})`;
C=Contrasts;
L=trace(ginv(C)*C);
q=L+1; *+1 if to account for intercept;
ivar=&ranistd;
svar=&ransstd;
corr=&correlation;
errvar=&errdev*&errdev;
cov=(ivar*({1 0, 0 0})+svar*({0 0, 0 1}))*(corr*({0 1, 1 0})+I(2))*(ivar*({1 0,
0 0})+svar*({0 0, 0 1}));
bint=&xbar2;
btreat=&xbar1-&xbar2; *treatment effect;
bcompare=&xbar3-&xbar2;
btime=&btime2; *time effect in reference group 2;
btreattime=&btime1-&btime2; *group*time interaction;
bcomptime=&btime3-&btime2;
Beta=(bint||btreat||bcompare||btime||btreattime||bcomptime)`;
B=Beta;
n=&n1+&n2+&n3;
use work.seqids_&rep;
       do i=1 to n;
         read all var {intercept Treatment Comparator Time TreatTime CompTime}
where(order=i) into X;
         read all var {intercept Time} where(order=i) into Z;
         S=z*cov*z`+I((nrow(X)))@errvar;
         if i=1 then V=(X`*inv(S)*X);
     else V=V+(X`*inv(S)*X);
end;
PHI=inv(V);
use work.seqids_&rep;;
       do i=1 to n;
         read all var {intercept Treatment Comparator Time TreatTime CompTime}
where(order=i) into X;
         read all var {intercept Time} where(order=i) into Z;
         S=z*cov*z`+I((nrow(X)))@errvar;
/*
```

```
*/
         d1=z*({1 0, 0 0})*z`;
         d2=z*({0 0, 0 1})*z`;
         d3=I(nrow(X));
         d4=z*({0 1, 1 0})*z`;
         if i=1 then P1=-X`*inv(S)*d1*inv(S)*X; else P1=P1-
X`*inv(S)*d1*inv(S)*X;
         if i=1 then P2=-X`*inv(S)*d2*inv(S)*X; else P2=P2-
X`*inv(S)*d2*inv(S)*X;
         if i=1 then P3=-X`*inv(S)*d3*inv(S)*X; else P3=P3-
X`*inv(S)*d3*inv(S)*X;
         if i=1 then P4=-X`*inv(S)*d4*inv(S)*X; else P4=P4-
X`*inv(S)*d4*inv(S)*X;
         if i=1 then Q11=X`*inv(S)*d1*inv(S)*d1*inv(S)*X; else
Q11=Q11+X`*inv(S)*d1*inv(S)*d1*inv(S)*X;
         if i=1 then Q12=X`*inv(S)*d1*inv(S)*d2*inv(S)*X; else
Q12=Q12+X`*inv(S)*d1*inv(S)*d2*inv(S)*X;
         if i=1 then Q13=X`*inv(S)*d1*inv(S)*d3*inv(S)*X; else
Q13=Q13+X`*inv(S)*d1*inv(S)*d3*inv(S)*X;
         if i=1 then Q14=X`*inv(S)*d1*inv(S)*d4*inv(S)*X; else
Q14=Q14+X`*inv(S)*d1*inv(S)*d4*inv(S)*X;
         if i=1 then Q21=X`*inv(S)*d2*inv(S)*d1*inv(S)*X; else
Q21=Q21+X`*inv(S)*d2*inv(S)*d1*inv(S)*X;
         if i=1 then Q22=X`*inv(S)*d2*inv(S)*d2*inv(S)*X; else
Q22=Q22+X`*inv(S)*d2*inv(S)*d2*inv(S)*X;
         if i=1 then Q23=X`*inv(S)*d2*inv(S)*d3*inv(S)*X; else
Q23=Q23+X`*inv(S)*d2*inv(S)*d3*inv(S)*X;
         if i=1 then Q24=X`*inv(S)*d2*inv(S)*d4*inv(S)*X; else
Q24=Q24+X`*inv(S)*d2*inv(S)*d4*inv(S)*X;
         if i=1 then Q31=X`*inv(S)*d3*inv(S)*d1*inv(S)*X; else
Q31=Q31+X`*inv(S)*d3*inv(S)*d1*inv(S)*X;
         if i=1 then Q32=X`*inv(S)*d3*inv(S)*d2*inv(S)*X; else
Q32=Q32+X`*inv(S)*d3*inv(S)*d2*inv(S)*X;
         if i=1 then Q33=X`*inv(S)*d3*inv(S)*d3*inv(S)*X; else
Q33=Q33+X`*inv(S)*d3*inv(S)*d3*inv(S)*X;
         if i=1 then Q34=X`*inv(S)*d3*inv(S)*d4*inv(S)*X; else
Q34=Q34+X`*inv(S)*d3*inv(S)*d4*inv(S)*X;
         if i=1 then Q41=X`*inv(S)*d4*inv(S)*d1*inv(S)*X; else
Q41=Q41+X`*inv(S)*d4*inv(S)*d1*inv(S)*X;
         if i=1 then Q42=X`*inv(S)*d4*inv(S)*d2*inv(S)*X; else
Q42=Q42+X`*inv(S)*d4*inv(S)*d2*inv(S)*X;
         if i=1 then Q43=X`*inv(S)*d4*inv(S)*d3*inv(S)*X; else
Q43=Q43+X`*inv(S)*d4*inv(S)*d3*inv(S)*X;
         if i=1 then Q44=X`*inv(S)*d4*inv(S)*d4*inv(S)*X; else
Q44=Q44+X`*inv(S)*d4*inv(S)*d4*inv(S)*X;
*Information Matrices;
         if i=1 then I11a=trace(inv(S)*d1*inv(S)*d1); else
I11a=trace(I11a)+trace(inv(S)*d1*inv(S)*d1);
         if i=1 then I12a=trace(inv(S)*d1*inv(S)*d2); else
I12a=trace(I12a)+trace(inv(S)*d1*inv(S)*d2);
         if i=1 then I13a=trace(inv(S)*d1*inv(S)*d3); else
I13a=trace(I13a)+trace(inv(S)*d1*inv(S)*d3);
         if i=1 then I14a=trace(inv(S)*d1*inv(S)*d4); else
I14a=trace(I14a)+trace(inv(S)*d1*inv(S)*d4);
         if i=1 then I21a=trace(inv(S)*d2*inv(S)*d1); else
I21a=trace(I21a)+trace(inv(S)*d2*inv(S)*d1);
         if i=1 then I22a=trace(inv(S)*d2*inv(S)*d2); else
I22a=trace(I22a)+trace(inv(S)*d2*inv(S)*d2);
         if i=1 then I23a=trace(inv(S)*d2*inv(S)*d3); else
I23a=trace(I23a)+trace(inv(S)*d2*inv(S)*d3);
/*
```

```
*/
        if i=1 then I24a=trace(inv(S)*d2*inv(S)*d4); else
I24a=trace(I24a)+trace(inv(S)*d2*inv(S)*d4);
        if i=1 then I31a=trace(inv(S)*d3*inv(S)*d1); else
I31a=trace(I31a)+trace(inv(S)*d3*inv(S)*d1);
        if i=1 then I32a=trace(inv(S)*d3*inv(S)*d2); else
I32a=trace(I32a)+trace(inv(S)*d3*inv(S)*d2);
        if i=1 then I33a=trace(inv(S)*d3*inv(S)*d3); else
I33a=trace(I33a)+trace(inv(S)*d3*inv(S)*d3);
        if i=1 then I34a=trace(inv(S)*d3*inv(S)*d4); else
I34a=trace(I34a)+trace(inv(S)*d3*inv(S)*d4);
        if i=1 then I41a=trace(inv(S)*d4*inv(S)*d1); else
I41a=trace(I41a)+trace(inv(S)*d4*inv(S)*d1);
        if i=1 then I42a=trace(inv(S)*d4*inv(S)*d2); else
I42a=trace(I42a)+trace(inv(S)*d4*inv(S)*d2);
        if i=1 then I43a=trace(inv(S)*d4*inv(S)*d3); else
I43a=trace(I43a)+trace(inv(S)*d4*inv(S)*d3);
        if i=1 then I44a=trace(inv(S)*d4*inv(S)*d4); else
I44a=trace(I44a)+trace(inv(S)*d4*inv(S)*d4);
end;
I11=((0.5*trace(I11a)-0.5*trace(2*PHI*Q11-PHI*P1*PHI*P1)))*({1 0 0 0, 0 0 0 0,
0 0 0 0, 0 0 0 0});
I12=((0.5*trace(I12a)-0.5*trace(2*PHI*Q12-PHI*P1*PHI*P2)))*({0 1 0 0, 0 0 0 0,
0 0 0 0, 0 0 0 0});
I13=((0.5*trace(I13a)-0.5*trace(2*PHI*Q13-PHI*P1*PHI*P3)))*({0 0 1 0, 0 0 0 0,
0 0 0 0, 0 0 0 0});
I14=((0.5*trace(I14a)-0.5*trace(2*PHI*Q14-PHI*P1*PHI*P4)))*({0 0 0 1, 0 0 0 0,
0 0 0 0, 0 0 0 0});
I21=((0.5*trace(I21a)-0.5*trace(2*PHI*Q21-PHI*P2*PHI*P1)))*({0 0 0 0, 1 0 0 0,
0 0 0 0, 0 0 0 0});
I22=((0.5*trace(I22a)-0.5*trace(2*PHI*Q22-PHI*P2*PHI*P2)))*({0 0 0 0, 0 1 0 0,
0 0 0 0, 0 0 0 0});
I23=((0.5*trace(I23a)-0.5*trace(2*PHI*Q23-PHI*P2*PHI*P3)))*({0 0 0 0, 0 0 1 0,
0 0 0 0, 0 0 0 0});
I24=((0.5*trace(I24a)-0.5*trace(2*PHI*Q24-PHI*P2*PHI*P4)))*({0 0 0 0, 0 0 0 1,
0 0 0 0, 0 0 0 0});
I31=((0.5*trace(I31a)-0.5*trace(2*PHI*Q31-PHI*P3*PHI*P1)))*({0 0 0 0, 0 0 0 0,
1 0 0 0, 0 0 0 0});
I32=((0.5*trace(I32a)-0.5*trace(2*PHI*Q32-PHI*P3*PHI*P2)))*({0 0 0 0, 0 0 0 0,
0 1 0 0, 0 0 0 0});
I33=((0.5*trace(I33a)-0.5*trace(2*PHI*Q33-PHI*P3*PHI*P3)))*({0 0 0 0, 0 0 0 0,
0 0 1 0, 0 0 0 0});
I34=((0.5*trace(I34a)-0.5*trace(2*PHI*Q34-PHI*P3*PHI*P4)))*({0 0 0 0, 0 0 0 0,
0 0 0 1, 0 0 0 0});
I41=((0.5*trace(I41a)-0.5*trace(2*PHI*Q41-PHI*P4*PHI*P1)))*({0 0 0 0, 0 0 0 0,
0 0 0 0, 1 0 0 0});
I42=((0.5*trace(I42a)-0.5*trace(2*PHI*Q42-PHI*P4*PHI*P2)))*({0 0 0 0, 0 0 0 0,
0 0 0 0, 0 1 0 0});
I43=((0.5*trace(I43a)-0.5*trace(2*PHI*Q43-PHI*P4*PHI*P3)))*({0 0 0 0, 0 0 0 0,
0 0 0 0, 0 0 1 0});
I44=((0.5*trace(I44a)-0.5*trace(2*PHI*Q44-PHI*P4*PHI*P4)))*({0 0 0 0, 0 0 0 0,
0 0 0 0, 0 0 0 1});
I=ginv(I11+I12+I13+I14+I21+I22+I23+I24+I31+I32+I33+I34+I41+I42+I43+I44);
*Inverse of Expected Information Matrix;
*W Terms;
w11=({1 0 0 0})*I*({1 0 0 0})`;
w12=({1 0 0 0})*I*({0 1 0 0})`;
w13=({1 0 0 0})*I*({0 0 1 0})`;
w14=({1 0 0 0})*I*({0 0 0 1})`;
w21=({0 1 0 0})*I*({1 0 0 0})`;
w22=({0 1 0 0})*I*({0 1 0 0})`;
w23=({0 1 0 0})*I*({0 0 1 0})`;
/*
```

```
*/
w24=({0 1 0 0})*I*({0 0 0 1})`;
w31=({0 0 1 0})*I*({1 0 0 0})`;
w32=({0 0 1 0})*I*({0 1 0 0})`;
w33=({0 0 1 0})*I*({0 0 1 0})`;
w34=({0 0 1 0})*I*({0 0 0 1})`;
w41=({0 0 0 1})*I*({1 0 0 0})`;
w42=({0 0 0 1})*I*({0 1 0 0})`;
w43=({0 0 0 1})*I*({0 0 1 0})`;
w44=({0 0 0 1})*I*({0 0 0 1})`;
*Sum Terms;
c11=w11*(Q11-(P1*PHI*P1));
c12=w12*(Q12-(P1*PHI*P2));
c13=w13*(Q13-(P1*PHI*P3));
c14=w14*(Q14-(P1*PHI*P4));
c21=w21*(Q21-(P2*PHI*P1));
c22=w22*(Q22-(P2*PHI*P2));
c23=w23*(Q23-(P2*PHI*P3));
c24=w24*(Q24-(P2*PHI*P4));
c31=w31*(Q31-(P3*PHI*P1));
c32=w32*(Q32-(P3*PHI*P2));
c33=w33*(Q33-(P3*PHI*P3));
c34=w34*(Q34-(P3*PHI*P4));
c41=w41*(Q41-(P4*PHI*P1));
c42=w42*(Q42-(P4*PHI*P2));
c43=w43*(Q43-(P4*PHI*P3));
c44=w44*(Q44-(P4*PHI*P4));
sum=c11+c12+c13+c14+c21+c22+c23+c24+c31+c32+c33+c34+c41+c42+c43+c44;
PHIA=PHI+2*PHI*(sum)*PHI;
Astar=2*PHI*(sum)*PHI;
*Theta denoted as O in this code;
      O=C*ginv(C`*PHI*C)*C`;
*Traces;
*A1;
ta11=w11*trace(O*PHI*P1*PHI)*trace(O*PHI*P1*PHI);
ta12=w12*trace(O*PHI*P1*PHI)*trace(O*PHI*P2*PHI);
ta13=w13*trace(O*PHI*P1*PHI)*trace(O*PHI*P3*PHI);
ta14=w14*trace(O*PHI*P1*PHI)*trace(O*PHI*P4*PHI);
ta21=w21*trace(O*PHI*P2*PHI)*trace(O*PHI*P1*PHI);
ta22=w22*trace(O*PHI*P2*PHI)*trace(O*PHI*P2*PHI);
ta23=w23*trace(O*PHI*P2*PHI)*trace(O*PHI*P3*PHI);
ta24=w24*trace(O*PHI*P2*PHI)*trace(O*PHI*P4*PHI);
ta31=w31*trace(O*PHI*P3*PHI)*trace(O*PHI*P1*PHI);
ta32=w32*trace(O*PHI*P3*PHI)*trace(O*PHI*P2*PHI);
ta33=w33*trace(O*PHI*P3*PHI)*trace(O*PHI*P3*PHI);
ta34=w34*trace(O*PHI*P3*PHI)*trace(O*PHI*P4*PHI);
ta41=w41*trace(O*PHI*P4*PHI)*trace(O*PHI*P1*PHI);
ta42=w42*trace(O*PHI*P4*PHI)*trace(O*PHI*P2*PHI);
ta43=w43*trace(O*PHI*P4*PHI)*trace(O*PHI*P3*PHI);
ta44=w44*trace(O*PHI*P4*PHI)*trace(O*PHI*P4*PHI);
A1=ta11+ta12+ta13+ta14+ta21+ta22+ta23+ta24+ta31+ta32+ta33+ta34+ta41+ta42+ta43+t
a44;
*A2;
tb11=w11*trace(O*PHI*P1*PHI*O*PHI*P1*PHI);
tb12=w12*trace(O*PHI*P1*PHI*O*PHI*P2*PHI);
tb13=w13*trace(O*PHI*P1*PHI*O*PHI*P3*PHI);
tb14=w14*trace(O*PHI*P1*PHI*O*PHI*P4*PHI);
tb21=w21*trace(O*PHI*P2*PHI*O*PHI*P1*PHI);
tb22=w22*trace(O*PHI*P2*PHI*O*PHI*P2*PHI);
tb23=w23*trace(O*PHI*P2*PHI*O*PHI*P3*PHI);
tb24=w24*trace(O*PHI*P2*PHI*O*PHI*P4*PHI);
/*
```

```
*/
tb31=w31*trace(O*PHI*P3*PHI*O*PHI*P1*PHI);
tb32=w32*trace(O*PHI*P3*PHI*O*PHI*P2*PHI);
tb33=w33*trace(O*PHI*P3*PHI*O*PHI*P3*PHI);
tb34=w34*trace(O*PHI*P3*PHI*O*PHI*P4*PHI);
tb41=w41*trace(O*PHI*P4*PHI*O*PHI*P1*PHI);
tb42=w42*trace(O*PHI*P4*PHI*O*PHI*P2*PHI);
tb43=w43*trace(O*PHI*P4*PHI*O*PHI*P3*PHI);
tb44=w44*trace(O*PHI*P4*PHI*O*PHI*P4*PHI);
A2=tb11+tb12+tb13+tb14+tb21+tb22+tb23+tb24+tb31+tb32+tb33+tb34+tb41+tb42+tb43+t
b44;
*Mij component of E(F_NonCentral) terms;
M11=%Mij(i=1,j=1);
M12=%Mij(i=1,j=2);
M13=%Mij(i=1,j=3);
M14=%Mij(i=1,j=4);
M21=%Mij(i=2,j=1);
M22=%Mij(i=2,j=2);
M23=%Mij(i=2,j=3);
M24=%Mij(i=2,j=4);
M31=%Mij(i=3,j=1);
M32=%Mij(i=3,j=2);
M33=%Mij(i=3,j=3);
M34=%Mij(i=3,j=4);
M41=%Mij(i=4,j=1);
M42=%Mij(i=4,j=2);
M43=%Mij(i=4,j=3);
M44=%Mij(i=4,j=4);
*A3 component of E(F) and Var(F);
tc11=w11*B`*M11*B;
tc12=w12*B`*M12*B;
tc13=w13*B`*M13*B;
tc14=w14*B`*M14*B;
tc21=w21*B`*M21*B;
tc22=w22*B`*M22*B;
tc23=w23*B`*M23*B;
tc24=w24*B`*M24*B;
tc31=w31*B`*M31*B;
tc32=w32*B`*M32*B;
tc33=w33*B`*M33*B;
tc34=w34*B`*M34*B;
tc41=w41*B`*M41*B;
tc42=w42*B`*M42*B;
tc43=w43*B`*M43*B;
tc44=w44*B`*M44*B;
A3=0.5*(tc11+tc12+tc13+tc14+tc21+tc22+tc23+tc24+tc31+tc32+tc33+tc34+tc41+tc42+t
c43+tc44);
L=trace(ginv(C)*C);
*DF components;
E=inv(1-(A2/L));
B=(1/(2*L))*(A1+(6*A2));
gg=((L+1)*A1-(L+4)*A2)/((L+2)*A2);
dd1=gg/(3*L+2*(1-gg));
dd2=(L-gg)/(3*L+2*(1-gg));
dd3=(L-gg+2)/(3*L+2*(1-gg));
V=(2/L)*((1+(dd1*B))/((1-(dd2*B))*(1-(dd2*B))*(1-(dd3*B))));
rr=V/(2*E*E);
B=Beta;
*Degrees of Freedom M and Scale Factor Lambda;
m=4+((L+2)/((L*rr)-1));
Lambda=m/(E*(m-2));
/*
```

```
*/
*Expected Value;
E_null=(1/L)*(L+A2);
E_alt=E_null + (1/L)*(B`*C*ginv(C`*PHI*C)*C`*B+A3)-
(1/L)*(B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B);
ncp=lambda*(B`*C*ginv(C`*PHIA*C)*C`*B);
rat1=(lambda*E_null)/(m/(m-2));
ncpform=B`*C*ginv(C`*PHI*C)*C`*B;
E0=ginv(1-(A2/L)); *KR adjustment to E_null;
EA=(1/L)*ginv(ncpform-
A3+B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B)*((ncpform-
B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B)*(ncpform-
B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B));
ncpnew=L*Lambda*(EA/E0);
ncp1=ncpnew;
ncp2=ncp;
ncp3=ncpform;
ncp4=ncpform;
ncp5=ncpform;
ncpuse=&ncpuse;
scalera=B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B;
L=trace(ginv(C)*C);
*DF components;
E=inv(1-(A2/L));
B=(1/(2*L))*(A1+(6*A2));
gg=((L+1)*A1-(L+4)*A2)/((L+2)*A2);
dd1=gg/(3*L+2*(1-gg));
dd2=(L-gg)/(3*L+2*(1-gg));
dd3=(L-gg+2)/(3*L+2*(1-gg));
V=(2/L)*((1+(dd1*B))/((1-(dd2*B))*(1-(dd2*B))*(1-(dd3*B))));
rr=V/(2*E*E);
*Degrees of Freedom M and Scale Factor Lambda;
m=4+((L+2)/((L*rr)-1));
Lambda=m/(E*(m-2));
*Power;
alpha=&alpha;
alval=1-alpha;
timevec=({&timevec});
m1=m;
m2=m;
m3=m;
m4=(&n1+&n2+&n3)*(((nrow(timevec)-1)*&missparm+1)/nrow(timevec))-L-1;
m5=(&n1+&n2+&n3)*(((nrow(timevec)-
1)*&missparm+1)/nrow(timevec))*(nrow(timevec))-6;
muse=&ddfuse;
critval=finv(alval, L, muse); *So p(Wald<w) = p((1/lambda)*F < f);
power=1-probf(critval, L, muse, ncpuse);
ncp_dev=ncpnew+0.000001;
power_dev=((1-probf(critval, L, m, ncp_dev))-(1-probf(critval, L, m,
ncpnew)))/0.000001;
R2=1-inv(1+(1/L)*Lambda*ncp*(L/m));
create work.power_a from power;
append from power;
quit;

data work.power_b;
       set work.power_b work.power_a;
run;
%mend power;

/*
```

```
*/
%macro average;
	%do designs=1 %to &reps;
		%power(rep=&designs);
	%end;
%mend average;

%average; run;

proc sql;
	create table work.power_c as
		select COL1 as Power
		from work.power_b
		where COL1 ne 9999;
quit;

proc sql;
	create table work.AveragePower as
		select count(power) as Designs, mean(Power) as Power, std(power)
as DevPower, min(power) as MinPower, max(power) as maxPower,
				1*&n1 as n1final, 1*&n2 as n2final, 1*&n3 as n3final
		from work.power_c;
quit;
data work.MacroVals;
	set work.AveragePower;
	n1=round(&n1-&g1ratio);
	n2=round(&n2-&g2ratio);
	n3=round(&n3-&g3ratio);
	call symput('Power', Power);
	call symput('n1', n1);
	call symput('n2', n2);
	call symput('n3', n3);
run;
%end;
%mend m2;

%m2(n1=&n1, n2=&n2, n3=&n3); run;


proc sql;
	create table work.SampleSize2 as
		select round(n1final+&g1ratio) as n1, round(n2final+&g2ratio) as
n2, round(n3final+&g3ratio) as n3, n1final as n1small, n2final as n2small,
n3final as n3small, Power
		from work.AveragePower;
quit;

data work.MacroVals;
	set work.SampleSize2;
	call symput('Power', Power);
	call symput('n1', n1);
	call symput('n2', n2);
	call symput('n3', n3);
run;


%macro m3(n1=&q1,n2=&q2,n3=&q3);
data work.Treatment;
	seed=&misseed1+1;
	reps=&reps;
	do rep=1 to reps;
	 do n=1 to &n1;
/*
```

```
*/
        ID=compress(put(n,best8.)||"A");
               do Time=min(&timevec) to max(&timevec);
                     z=ranuni(seed); /*Uses uniform(0,1) random variable "z" to
assign missing values if z>missparm, modified later to force baseline as
nonmissing*/
                     if  z>&missparm then u=0;
                     else if z<=&missparm then u=1;
                     output;
               end;
         end;
        end;
        drop seed z;
run;

proc sql;
       create table work.Treatobs as
              select rep, ID, Time, case Time when . then 1 else 1 end as
intercept,
                                 case Time when . then 1 else 1 end as Treatment,
                           case Time when . then 0 else 0 end as Comparator, u
              from work.Treatment
              where Time in (&timevec);
quit;

data work.Placebo;
       seed=&misseed2+1;
       reps=&reps;
   do rep=1 to reps;
        do n=1 to &n2;
        ID=compress(put(n,best8.)||"B");
               do time=min(&timevec) to max(&timevec);
                     z=ranuni(seed); /*Uses uniform(0,1) random variable "z" to
assign missing values if z>missparm, modified later to force baseline as
nonmissing*/
                     if  z>&missparm then u=0;
                     else if z<=&missparm then u=1;
                     output;
               end;
         end;
        end;
        drop seed z;
run;

proc sql;
       create table work.Plbobs as
              select rep, ID, Time, case Time when . then 1 else 1 end as
intercept,
                                 case Time when . then 0 else 0 end as Treatment,
                                 case Time when . then 0 else 0 end as Comparator, u
              from work.Placebo
              where Time in (&timevec);
quit;

data work.Comparator;
       seed=&misseed3+1;
       reps=&reps;
   do rep=1 to reps;
        do n=1 to &n3;
        ID=compress(put(n,best8.)||"C");
               do time=min(&timevec) to max(&timevec);
/*
```

```
*/
                    z=ranuni(seed); /*Uses uniform(0,1) random variable "z" to
assign missing values if z>missparm, modified later to force baseline as
nonmissing*/
                    if  z>&missparm then u=0;
                    else if z<=&missparm then u=1;
                    output;
               end;
         end;
        end;
        drop seed z;
run;

proc sql;
       create table work.Cmptorobs as
              select rep, ID, Time, case Time when . then 1 else 1 end as
intercept,
                           case Time when . then 0 else 0 end as Treatment,
                           case Time when . then 1 else 1 end as Comparator, u
              from work.Comparator
              where Time in (&timevec);
quit;

data work.MixedData_a; *Combine placebo and treatment designs into one dataset;
       set work.Treatobs work.Plbobs work.Cmptorobs;
run;

proc sql; *Will create variable "M" to serve as missing data indicator and make
sure no baseline values are missing;
       create table work.MixedData_b as
              select rep, ID, intercept, Treatment, Comparator, Time,
Treatment*Time as TreatTime, Comparator*Time as CompTime,
                           case when u=0 and time ne min(&timevec) then . else
1 end as M
              from work.MixedData_a;
quit;

proc sql;
       create table work.MixedData as
              select rep, ID, intercept, Treatment, Comparator, Time, TreatTime,
CompTime
              from work.MixedData_b
              where M ne .;
quit;

proc sort data=work.MixedData;
       by rep id time;
run;
quit;

data work.power_b;
       COL1=9999;
run;

%macro power(rep=&sim);
proc sql;
       create table work.mixeddata_&rep as
              select *
              from work.MixedData
              where rep=&rep;
quit;
/*
```

```
*/
proc sort data=work.mixeddata_&rep;
       by rep id time;
run;
quit;
data work.seqids_&rep;
       set work.MixedData_&rep;
       by id;
       retain order 0;
              if first.id then order=order+1;
run;
ods select all;
proc iml;
Contrasts={0 0 0 0 1 0, 0 0 0 0 0 1}`;
C=Contrasts;
L=trace(ginv(C)*C);
q=L+1; *+1 if to account for intercept;
ivar=&ranistd;
svar=&ransstd;
corr=&correlation;
errvar=&errdev*&errdev;
cov=(ivar*({1 0, 0 0})+svar*({0 0, 0 1}))*(corr*({0 1, 1 0})+I(2))*(ivar*({1 0,
0 0})+svar*({0 0, 0 1}));
bint=&xbar2;
btreat=&xbar1-&xbar2; *treatment effect;
bcompare=&xbar3-&xbar2;
btime=&btime2; *time effect in reference group 2;
btreattime=&btime1-&btime2; *group*time interaction;
bcomptime=&btime3-&btime2;
Beta=(bint||btreat||bcompare||btime||btreattime||bcomptime)`;
B=Beta;
n=&n1+&n2+&n3;
use work.seqids_&rep;
       do i=1 to n;
         read all var {intercept Treatment Comparator Time TreatTime CompTime}
where(order=i) into X;
         read all var {intercept Time} where(order=i) into Z;
         S=z*cov*z`+I((nrow(X)))@errvar;
         if i=1 then V=(X`*inv(S)*X);
     else V=V+(X`*inv(S)*X);
end;
PHI=inv(V);
use work.seqids_&rep;;
       do i=1 to n;
         read all var {intercept Treatment Comparator Time TreatTime CompTime}
where(order=i) into X;
         read all var {intercept Time} where(order=i) into Z;
         S=z*cov*z`+I((nrow(X)))@errvar;
         d1=z*({1 0, 0 0})*z`;
         d2=z*({0 0, 0 1})*z`;
         d3=I(nrow(X));
         d4=z*({0 1, 1 0})*z`;
         if i=1 then P1=-X`*inv(S)*d1*inv(S)*X; else P1=P1-
X`*inv(S)*d1*inv(S)*X;
         if i=1 then P2=-X`*inv(S)*d2*inv(S)*X; else P2=P2-
X`*inv(S)*d2*inv(S)*X;
         if i=1 then P3=-X`*inv(S)*d3*inv(S)*X; else P3=P3-
X`*inv(S)*d3*inv(S)*X;
         if i=1 then P4=-X`*inv(S)*d4*inv(S)*X; else P4=P4-
X`*inv(S)*d4*inv(S)*X;
/*
```

```
*/
         if i=1 then Q11=X`*inv(S)*d1*inv(S)*d1*inv(S)*X; else
Q11=Q11+X`*inv(S)*d1*inv(S)*d1*inv(S)*X;
         if i=1 then Q12=X`*inv(S)*d1*inv(S)*d2*inv(S)*X; else
Q12=Q12+X`*inv(S)*d1*inv(S)*d2*inv(S)*X;
         if i=1 then Q13=X`*inv(S)*d1*inv(S)*d3*inv(S)*X; else
Q13=Q13+X`*inv(S)*d1*inv(S)*d3*inv(S)*X;
         if i=1 then Q14=X`*inv(S)*d1*inv(S)*d4*inv(S)*X; else
Q14=Q14+X`*inv(S)*d1*inv(S)*d4*inv(S)*X;
         if i=1 then Q21=X`*inv(S)*d2*inv(S)*d1*inv(S)*X; else
Q21=Q21+X`*inv(S)*d2*inv(S)*d1*inv(S)*X;
         if i=1 then Q22=X`*inv(S)*d2*inv(S)*d2*inv(S)*X; else
Q22=Q22+X`*inv(S)*d2*inv(S)*d2*inv(S)*X;
         if i=1 then Q23=X`*inv(S)*d2*inv(S)*d3*inv(S)*X; else
Q23=Q23+X`*inv(S)*d2*inv(S)*d3*inv(S)*X;
         if i=1 then Q24=X`*inv(S)*d2*inv(S)*d4*inv(S)*X; else
Q24=Q24+X`*inv(S)*d2*inv(S)*d4*inv(S)*X;
         if i=1 then Q31=X`*inv(S)*d3*inv(S)*d1*inv(S)*X; else
Q31=Q31+X`*inv(S)*d3*inv(S)*d1*inv(S)*X;
         if i=1 then Q32=X`*inv(S)*d3*inv(S)*d2*inv(S)*X; else
Q32=Q32+X`*inv(S)*d3*inv(S)*d2*inv(S)*X;
         if i=1 then Q33=X`*inv(S)*d3*inv(S)*d3*inv(S)*X; else
Q33=Q33+X`*inv(S)*d3*inv(S)*d3*inv(S)*X;
         if i=1 then Q34=X`*inv(S)*d3*inv(S)*d4*inv(S)*X; else
Q34=Q34+X`*inv(S)*d3*inv(S)*d4*inv(S)*X;
         if i=1 then Q41=X`*inv(S)*d4*inv(S)*d1*inv(S)*X; else
Q41=Q41+X`*inv(S)*d4*inv(S)*d1*inv(S)*X;
         if i=1 then Q42=X`*inv(S)*d4*inv(S)*d2*inv(S)*X; else
Q42=Q42+X`*inv(S)*d4*inv(S)*d2*inv(S)*X;
         if i=1 then Q43=X`*inv(S)*d4*inv(S)*d3*inv(S)*X; else
Q43=Q43+X`*inv(S)*d4*inv(S)*d3*inv(S)*X;
         if i=1 then Q44=X`*inv(S)*d4*inv(S)*d4*inv(S)*X; else
Q44=Q44+X`*inv(S)*d4*inv(S)*d4*inv(S)*X;
*Information Matrices;
         if i=1 then I11a=trace(inv(S)*d1*inv(S)*d1); else
I11a=trace(I11a)+trace(inv(S)*d1*inv(S)*d1);
         if i=1 then I12a=trace(inv(S)*d1*inv(S)*d2); else
I12a=trace(I12a)+trace(inv(S)*d1*inv(S)*d2);
         if i=1 then I13a=trace(inv(S)*d1*inv(S)*d3); else
I13a=trace(I13a)+trace(inv(S)*d1*inv(S)*d3);
         if i=1 then I14a=trace(inv(S)*d1*inv(S)*d4); else
I14a=trace(I14a)+trace(inv(S)*d1*inv(S)*d4);
         if i=1 then I21a=trace(inv(S)*d2*inv(S)*d1); else
I21a=trace(I21a)+trace(inv(S)*d2*inv(S)*d1);
         if i=1 then I22a=trace(inv(S)*d2*inv(S)*d2); else
I22a=trace(I22a)+trace(inv(S)*d2*inv(S)*d2);
         if i=1 then I23a=trace(inv(S)*d2*inv(S)*d3); else
I23a=trace(I23a)+trace(inv(S)*d2*inv(S)*d3);
         if i=1 then I24a=trace(inv(S)*d2*inv(S)*d4); else
I24a=trace(I24a)+trace(inv(S)*d2*inv(S)*d4);
         if i=1 then I31a=trace(inv(S)*d3*inv(S)*d1); else
I31a=trace(I31a)+trace(inv(S)*d3*inv(S)*d1);
         if i=1 then I32a=trace(inv(S)*d3*inv(S)*d2); else
I32a=trace(I32a)+trace(inv(S)*d3*inv(S)*d2);
         if i=1 then I33a=trace(inv(S)*d3*inv(S)*d3); else
I33a=trace(I33a)+trace(inv(S)*d3*inv(S)*d3);
         if i=1 then I34a=trace(inv(S)*d3*inv(S)*d4); else
I34a=trace(I34a)+trace(inv(S)*d3*inv(S)*d4);
         if i=1 then I41a=trace(inv(S)*d4*inv(S)*d1); else
I41a=trace(I41a)+trace(inv(S)*d4*inv(S)*d1);
/*
```

```
*/
        if i=1 then I42a=trace(inv(S)*d4*inv(S)*d2); else
I42a=trace(I42a)+trace(inv(S)*d4*inv(S)*d2);
        if i=1 then I43a=trace(inv(S)*d4*inv(S)*d3); else
I43a=trace(I43a)+trace(inv(S)*d4*inv(S)*d3);
        if i=1 then I44a=trace(inv(S)*d4*inv(S)*d4); else
I44a=trace(I44a)+trace(inv(S)*d4*inv(S)*d4);
end;
I11=((0.5*trace(I11a)-0.5*trace(2*PHI*Q11-PHI*P1*PHI*P1)))*({1 0 0 0, 0 0 0 0,
0 0 0 0, 0 0 0 0});
I12=((0.5*trace(I12a)-0.5*trace(2*PHI*Q12-PHI*P1*PHI*P2)))*({0 1 0 0, 0 0 0 0,
0 0 0 0, 0 0 0 0});
I13=((0.5*trace(I13a)-0.5*trace(2*PHI*Q13-PHI*P1*PHI*P3)))*({0 0 1 0, 0 0 0 0,
0 0 0 0, 0 0 0 0});
I14=((0.5*trace(I14a)-0.5*trace(2*PHI*Q14-PHI*P1*PHI*P4)))*({0 0 0 1, 0 0 0 0,
0 0 0 0, 0 0 0 0});
I21=((0.5*trace(I21a)-0.5*trace(2*PHI*Q21-PHI*P2*PHI*P1)))*({0 0 0 0, 1 0 0 0,
0 0 0 0, 0 0 0 0});
I22=((0.5*trace(I22a)-0.5*trace(2*PHI*Q22-PHI*P2*PHI*P2)))*({0 0 0 0, 0 1 0 0,
0 0 0 0, 0 0 0 0});
I23=((0.5*trace(I23a)-0.5*trace(2*PHI*Q23-PHI*P2*PHI*P3)))*({0 0 0 0, 0 0 1 0,
0 0 0 0, 0 0 0 0});
I24=((0.5*trace(I24a)-0.5*trace(2*PHI*Q24-PHI*P2*PHI*P4)))*({0 0 0 0, 0 0 0 1,
0 0 0 0, 0 0 0 0});
I31=((0.5*trace(I31a)-0.5*trace(2*PHI*Q31-PHI*P3*PHI*P1)))*({0 0 0 0, 0 0 0 0,
1 0 0 0, 0 0 0 0});
I32=((0.5*trace(I32a)-0.5*trace(2*PHI*Q32-PHI*P3*PHI*P2)))*({0 0 0 0, 0 0 0 0,
0 1 0 0, 0 0 0 0});
I33=((0.5*trace(I33a)-0.5*trace(2*PHI*Q33-PHI*P3*PHI*P3)))*({0 0 0 0, 0 0 0 0,
0 0 1 0, 0 0 0 0});
I34=((0.5*trace(I34a)-0.5*trace(2*PHI*Q34-PHI*P3*PHI*P4)))*({0 0 0 0, 0 0 0 0,
0 0 0 1, 0 0 0 0});
I41=((0.5*trace(I41a)-0.5*trace(2*PHI*Q41-PHI*P4*PHI*P1)))*({0 0 0 0, 0 0 0 0,
0 0 0 0, 1 0 0 0});
I42=((0.5*trace(I42a)-0.5*trace(2*PHI*Q42-PHI*P4*PHI*P2)))*({0 0 0 0, 0 0 0 0,
0 0 0 0, 0 1 0 0});
I43=((0.5*trace(I43a)-0.5*trace(2*PHI*Q43-PHI*P4*PHI*P3)))*({0 0 0 0, 0 0 0 0,
0 0 0 0, 0 0 1 0});
I44=((0.5*trace(I44a)-0.5*trace(2*PHI*Q44-PHI*P4*PHI*P4)))*({0 0 0 0, 0 0 0 0,
0 0 0 0, 0 0 0 1});
I=ginv(I11+I12+I13+I14+I21+I22+I23+I24+I31+I32+I33+I34+I41+I42+I43+I44);
*Inverse of Expected Information Matrix;
*W Terms;
w11=({1 0 0 0})*I*({1 0 0 0})`;
w12=({1 0 0 0})*I*({0 1 0 0})`;
w13=({1 0 0 0})*I*({0 0 1 0})`;
w14=({1 0 0 0})*I*({0 0 0 1})`;
w21=({0 1 0 0})*I*({1 0 0 0})`;
w22=({0 1 0 0})*I*({0 1 0 0})`;
w23=({0 1 0 0})*I*({0 0 1 0})`;
w24=({0 1 0 0})*I*({0 0 0 1})`;
w31=({0 0 1 0})*I*({1 0 0 0})`;
w32=({0 0 1 0})*I*({0 1 0 0})`;
w33=({0 0 1 0})*I*({0 0 1 0})`;
w34=({0 0 1 0})*I*({0 0 0 1})`;
w41=({0 0 0 1})*I*({1 0 0 0})`;
w42=({0 0 0 1})*I*({0 1 0 0})`;
w43=({0 0 0 1})*I*({0 0 1 0})`;
w44=({0 0 0 1})*I*({0 0 0 1})`;
*Sum Terms;
c11=w11*(Q11-(P1*PHI*P1));
c12=w12*(Q12-(P1*PHI*P2));
/*
```

```
*/
c13=w13*(Q13-(P1*PHI*P3));
c14=w14*(Q14-(P1*PHI*P4));
c21=w21*(Q21-(P2*PHI*P1));
c22=w22*(Q22-(P2*PHI*P2));
c23=w23*(Q23-(P2*PHI*P3));
c24=w24*(Q24-(P2*PHI*P4));
c31=w31*(Q31-(P3*PHI*P1));
c32=w32*(Q32-(P3*PHI*P2));
c33=w33*(Q33-(P3*PHI*P3));
c34=w34*(Q34-(P3*PHI*P4));
c41=w41*(Q41-(P4*PHI*P1));
c42=w42*(Q42-(P4*PHI*P2));
c43=w43*(Q43-(P4*PHI*P3));
c44=w44*(Q44-(P4*PHI*P4));
sum=c11+c12+c13+c14+c21+c22+c23+c24+c31+c32+c33+c34+c41+c42+c43+c44;
PHIA=PHI+2*PHI*(sum)*PHI;
Astar=2*PHI*(sum)*PHI;
*Theta denoted as O in this code;
       O=C*ginv(C`*PHI*C)*C`;
*Traces;
*A1;
ta11=w11*trace(O*PHI*P1*PHI)*trace(O*PHI*P1*PHI);
ta12=w12*trace(O*PHI*P1*PHI)*trace(O*PHI*P2*PHI);
ta13=w13*trace(O*PHI*P1*PHI)*trace(O*PHI*P3*PHI);
ta14=w14*trace(O*PHI*P1*PHI)*trace(O*PHI*P4*PHI);
ta21=w21*trace(O*PHI*P2*PHI)*trace(O*PHI*P1*PHI);
ta22=w22*trace(O*PHI*P2*PHI)*trace(O*PHI*P2*PHI);
ta23=w23*trace(O*PHI*P2*PHI)*trace(O*PHI*P3*PHI);
ta24=w24*trace(O*PHI*P2*PHI)*trace(O*PHI*P4*PHI);
ta31=w31*trace(O*PHI*P3*PHI)*trace(O*PHI*P1*PHI);
ta32=w32*trace(O*PHI*P3*PHI)*trace(O*PHI*P2*PHI);
ta33=w33*trace(O*PHI*P3*PHI)*trace(O*PHI*P3*PHI);
ta34=w34*trace(O*PHI*P3*PHI)*trace(O*PHI*P4*PHI);
ta41=w41*trace(O*PHI*P4*PHI)*trace(O*PHI*P1*PHI);
ta42=w42*trace(O*PHI*P4*PHI)*trace(O*PHI*P2*PHI);
ta43=w43*trace(O*PHI*P4*PHI)*trace(O*PHI*P3*PHI);
ta44=w44*trace(O*PHI*P4*PHI)*trace(O*PHI*P4*PHI);
A1=ta11+ta12+ta13+ta14+ta21+ta22+ta23+ta24+ta31+ta32+ta33+ta34+ta41+ta42+ta43+t
a44;
*A2;
tb11=w11*trace(O*PHI*P1*PHI*O*PHI*P1*PHI);
tb12=w12*trace(O*PHI*P1*PHI*O*PHI*P2*PHI);
tb13=w13*trace(O*PHI*P1*PHI*O*PHI*P3*PHI);
tb14=w14*trace(O*PHI*P1*PHI*O*PHI*P4*PHI);
tb21=w21*trace(O*PHI*P2*PHI*O*PHI*P1*PHI);
tb22=w22*trace(O*PHI*P2*PHI*O*PHI*P2*PHI);
tb23=w23*trace(O*PHI*P2*PHI*O*PHI*P3*PHI);
tb24=w24*trace(O*PHI*P2*PHI*O*PHI*P4*PHI);
tb31=w31*trace(O*PHI*P3*PHI*O*PHI*P1*PHI);
tb32=w32*trace(O*PHI*P3*PHI*O*PHI*P2*PHI);
tb33=w33*trace(O*PHI*P3*PHI*O*PHI*P3*PHI);
tb34=w34*trace(O*PHI*P3*PHI*O*PHI*P4*PHI);
tb41=w41*trace(O*PHI*P4*PHI*O*PHI*P1*PHI);
tb42=w42*trace(O*PHI*P4*PHI*O*PHI*P2*PHI);
tb43=w43*trace(O*PHI*P4*PHI*O*PHI*P3*PHI);
tb44=w44*trace(O*PHI*P4*PHI*O*PHI*P4*PHI);
A2=tb11+tb12+tb13+tb14+tb21+tb22+tb23+tb24+tb31+tb32+tb33+tb34+tb41+tb42+tb43+t
b44;
*Mij component of E(F_NonCentral) terms;
M11=%Mij(i=1,j=1);
/*
```

```
*/
M12=%Mij(i=1,j=2);
M13=%Mij(i=1,j=3);
M14=%Mij(i=1,j=4);
M21=%Mij(i=2,j=1);
M22=%Mij(i=2,j=2);
M23=%Mij(i=2,j=3);
M24=%Mij(i=2,j=4);
M31=%Mij(i=3,j=1);
M32=%Mij(i=3,j=2);
M33=%Mij(i=3,j=3);
M34=%Mij(i=3,j=4);
M41=%Mij(i=4,j=1);
M42=%Mij(i=4,j=2);
M43=%Mij(i=4,j=3);
M44=%Mij(i=4,j=4);
*A3 component of E(F) and Var(F);
tc11=w11*B`*M11*B;
tc12=w12*B`*M12*B;
tc13=w13*B`*M13*B;
tc14=w14*B`*M14*B;
tc21=w21*B`*M21*B;
tc22=w22*B`*M22*B;
tc23=w23*B`*M23*B;
tc24=w24*B`*M24*B;
tc31=w31*B`*M31*B;
tc32=w32*B`*M32*B;
tc33=w33*B`*M33*B;
tc34=w34*B`*M34*B;
tc41=w41*B`*M41*B;
tc42=w42*B`*M42*B;
tc43=w43*B`*M43*B;
tc44=w44*B`*M44*B;
A3=0.5*(tc11+tc12+tc13+tc14+tc21+tc22+tc23+tc24+tc31+tc32+tc33+tc34+tc41+tc42+t
c43+tc44);
L=trace(ginv(C)*C);
*DF components;
E=inv(1-(A2/L));
B=(1/(2*L))*(A1+(6*A2));
gg=((L+1)*A1-(L+4)*A2)/((L+2)*A2);
dd1=gg/(3*L+2*(1-gg));
dd2=(L-gg)/(3*L+2*(1-gg));
dd3=(L-gg+2)/(3*L+2*(1-gg));
V=(2/L)*((1+(dd1*B))/((1-(dd2*B))*(1-(dd2*B))*(1-(dd3*B))));
rr=V/(2*E*E);
B=Beta;
*Degrees of Freedom M and Scale Factor Lambda;
m=4+((L+2)/((L*rr)-1));
Lambda=m/(E*(m-2));
*Expected Value;
E_null=(1/L)*(L+A2);
E_alt=E_null + (1/L)*(B`*C*ginv(C`*PHI*C)*C`*B+A3)-
(1/L)*(B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B);
ncp=lambda*(B`*C*ginv(C`*PHIA*C)*C`*B);
rat1=(lambda*E_null)/(m/(m-2));
ncpform=B`*C*ginv(C`*PHI*C)*C`*B;
E0=ginv(1-(A2/L)); *KR adjustment to E_null;
EA=(1/L)*ginv(ncpform-
A3+B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B)*((ncpform-
B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B)*(ncpform-
B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B));
/*
```

```
*/
ncpnew=L*Lambda*(EA/E0);
ncp1=ncpnew;
ncp2=ncp;
ncp3=ncpform;
ncp4=ncpform;
ncp5=ncpform;
ncpuse=&ncpuse;
scalera=B`*C*ginv(C`*PHI*C)*C`*Astar*C*ginv(C`*PHI*C)*C`*B;
L=trace(ginv(C)*C);
*DF components;
E=inv(1-(A2/L));
B=(1/(2*L))*(A1+(6*A2));
gg=((L+1)*A1-(L+4)*A2)/((L+2)*A2);
dd1=gg/(3*L+2*(1-gg));
dd2=(L-gg)/(3*L+2*(1-gg));
dd3=(L-gg+2)/(3*L+2*(1-gg));
V=(2/L)*((1+(dd1*B))/((1-(dd2*B))*(1-(dd2*B))*(1-(dd3*B))));
rr=V/(2*E*E);
*Degrees of Freedom M and Scale Factor Lambda;
m=4+((L+2)/((L*rr)-1));
Lambda=m/(E*(m-2));
*Power;
alpha=&alpha;
alval=1-alpha;
timevec=({&timevec});
m1=m;
m2=m;
m3=m;
m4=(&n1+&n2+&n3)*(((nrow(timevec)-1)*&missparm+1)/nrow(timevec))-L-1;
m5=(&n1+&n2+&n3)*(((nrow(timevec)-
1)*&missparm+1)/nrow(timevec))*(nrow(timevec))-6;
muse=&ddfuse;
critval=finv(alval, L, muse); *So p(Wald<w) = p((1/lambda)*F < f);
power=1-probf(critval, L, muse, ncpuse);
ncp_dev=ncpnew+0.000001;
power_dev=((1-probf(critval, L, m, ncp_dev))-(1-probf(critval, L, m,
ncpnew)))/0.000001;
R2=1-inv(1+(1/L)*Lambda*ncp*(L/m));
create work.power_a from power;
append from power;
quit;
data work.power_b;
      set work.power_b work.power_a;
run;
%mend power;

%macro average;
      %do designs=1 %to &reps;
            %power(rep=&designs);
      %end;
%mend average;

%average; run;

proc sql;
      create table work.power_c as
            select COL1 as Power
            from work.power_b
            where COL1 ne 9999;
quit;
/*
```

```
*/
proc sql;
	create table work.AveragePower as
		select count(power) as Designs, mean(Power) as Power, std(power)
as DevPower, min(power) as MinPower, max(power) as maxPower,
				1*&n1 as n1final, 1*&n2 as n2final, 1*&n3 as n3final
		from work.power_c;
quit;
%mend m3;

%m3(n1=&n1, n2=&n2, n3=&n3); run;


proc sql;
	create table work.SampleSize as
		select n1final as n1, n2final as n2, n3final as n3, Power as
Nominal_Power "Nominal Power", sum(0,&reps) as reps
		from work.AveragePower;
quit;

proc print data=work.SampleSize noobs;
run;
quit;
```