



VCU

Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations

Graduate School

2009

Joint Mixed-Effects Models for Longitudinal Data Analysis: An Application for the Metabolic Syndrome

John Thorp III
Virginia Commonwealth University

Follow this and additional works at: <https://scholarscompass.vcu.edu/etd>



Part of the [Biostatistics Commons](#)

© The Author

Downloaded from

<https://scholarscompass.vcu.edu/etd/1943>

This Thesis is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

JOINT MIXED-EFFECTS MODELS FOR LONGITUDINAL DATA ANALYSIS:
AN APPLICATION FOR THE METABOLIC SYNDROME

A Thesis submitted in partial fulfillment of the requirements for the degree of
Masters of Biostatistics at Virginia Commonwealth University.

by

JOHN MERCER THORP, III

B.S. Statistics, North Carolina State University, 2007

Director: JESSICA M. KETCHUM, PH.D.

ASSISTANT PROFESSOR, DEPARTMENT OF BIostatISTICS

Virginia Commonwealth University

Richmond, Virginia

November 2009

Acknowledgements

With most lists, there is an inherent emphasis placed on the first and last slots, as these are the most unique. Regarding my acknowledgements, I have specifically chosen those to thank first and last, in between, I do not have any specific order or agenda. That being noted, I would like to begin with someone very special to me; Rosemary Steinbach. She has been a very important part of my life since beginning my graduate school studies, and her love, support, advice and occasional pushing have made all the difference when things might have been harder than expected. I look forward to joining in with her on a new adventure, as she has been there for me on mine. I want her to know she means the world to me and that I love her very much.

As far as I know, there hasn't been a thesis or dissertation produced without the dedicated assistance of an advisor, and this thesis is certainly no exception, it would not have been completed without the supervision and guidance of Dr. Jessica Ketchum. All the way through to the last day she was willing to spend all the time needed (regardless of what time of day, or night, it may have been) to complete the task at hand. I have a tremendous amount of gratitude towards her for the myriad ways in which she helped this work come together.

Family and friends have also played a large role in my well being, both mentally and academically. I have been fortunate to have found a city in which to do my graduate school work where I had a support system of family. Specifically, my grandparents have

been there with me along the way by offering everything from a blessed distraction from the day-to-day with a nice dinner out to monetary support to their thoughts and prayers. They each have an extremely special place in my heart. Furthermore, I have made more friends than I might ever have had the audacity to hope for, who have been there for laughter and counsel, fun and study alike. To list all these names with adequate attention to what they've meant to me would take much more space than available here, I trust they know who they are. The support of my siblings and extended family has also been an added boost to what I have been able to accomplish.

I now wish to bestow the honor of my final acknowledgement on my parents. Through the years our relationship has seen many phases and many faces, though the past years of my life, spent on post graduate work, has found us closer than ever. In ways too innumerable to touch upon they have gotten me through the rigors of graduate school, from enrollment to employment, and I doubt I would have found the success and accomplishment I have without their love. Here is to the next chapter; thank you mom and dad, I love you both dearly.

JMT

November 15, 2009

Table of Contents

	Page
Acknowledgements	ii
List of Tables	vii
List of Figures	ix
Abstract	x
1 Introduction	1
1.1. Introduction	1
1.2. Modeling Longitudinal Data	2
1.2.1 Some Notation	2
1.2.2 General Approaches	4
1.2.3 Modeling Different Types of Longitudinal Data	8
1.3. Modeling Joint Longitudinal Data	10
1.3.1 General Approaches	11
1.3.2 Choosing Between the Univariate Models and a Multivariate Model	15
1.4. The Metabolic Syndrome	16
1.4.1 Diagnosis of the Metabolic Syndrome	16
1.5. Jointly Modeling Risk Factors of the Metabolic Syndrome over Time	21
2 Methodology	22
2.1. Introduction	22

2.2.	Univariate Normal Mixed-Effects Model Definition.....	23
2.3.	Estimation in the Normal Random-Effects Model.....	24
2.3.1	Methods of Estimation.....	24
2.3.2	Estimation of Fixed-Effects Parameters.....	27
2.3.3	Estimation of Random-Effects Parameters.....	28
2.3.4	Estimation of Variance Parameters.....	31
2.4.	Multivariate Methods.....	31
2.5.	Multivariate Normal Random-Effects Model Definition.....	32
2.5.1	Special Cases for the G Matrix.....	34
2.5.2	Association of the Evolutions (AOE).....	36
2.5.3	Evolution of the Association (EOA).....	36
2.6.	Summary.....	38
3	Data Analysis.....	39
3.1.	Introduction.....	39
3.2.	The Fels Longitudinal Study.....	39
3.3.	The Metabolic Syndrome Measurements.....	40
3.3.1	Blood Pressure.....	40
3.3.2	Waist Circumference.....	40
3.3.3	Blood Measures.....	41
3.3.4	Additional Measures.....	41

3.4.	Data Management and Preparation	42
3.4.1	Childhood Subset	43
3.4.2	Adulthood Subset.....	44
3.4.3	Further Considerations	47
3.5.	Results	48
3.5.1	Description of the Sample.....	48
3.5.2	Univariate and Multivariate Models for Male Subjects.....	49
3.5.3	Univariate and Multivariate Models for Female Subjects	62
3.5.4	Comparisons.....	75
4	Conclusions and Future Work.....	83
4.1.	Overview of Work.....	83
4.2.	Limitations	85
5	List of References	87
5.1.....		88
6	Appendices.....	96
6.1.	Details for Data Processing.....	96
6.2.	Details for Mixed-Effects Modeling	104

List of Tables

	Page
Table 1: ATP III Criteria Categorical Cut Points for the Metabolic Syndrome	20
Table 2: Distribution of Adulthood Risks for the Metabolic Syndrome.....	47
Table 3: Fixed-Effects Tests for Males (Univariate)	50
Table 4: Male's LS Means Differences (Univariate).....	52
Table 5: Estimated G Variance-Covariance Matrix for Males (Univariate).....	54
Table 6: Estimated G Correlation Matrix for Males (Univariate)	55
Table 7: Fixed-Effects Tests for Males (Multivariate)	57
Table 8: Male's LS Means Differences (Multivariate)	59
Table 9: Estimated G Variance-Covariance Matrix for Males (Multivariate).....	60
Table 10: Estimated G Correlation Matrix for Males (Multivariate)	60
Table 11: Fixed-Effects Tests for Females (Univariate).....	63
Table 12: Female's LS Means Differences (Univariate)	65
Table 13: Estimated G Variance-Covariance Matrix for Females (Univariate)	67
Table 14: Estimated G Correlation Matrix for Females (Univariate).....	68
Table 15: Fixed-Effects Tests for Females (Multivariate).....	69
Table 16: Female's LS Means Differences (Multivariate)	71
Table 17: Estimated G Variance-Covariance Matrix for Females (Multivariate)	72
Table 18: Estimated G Correlation Matrix for Females (Multivariate).....	72

Table 19: REML Estimates for the Covariance and Correlation Parameters in the Univariate and Multivariate Models for the Male Participants.....	76
Table 20: Covariate Comparison for Male Subjects	78
Table 21: REML Estimates for the Covariance and Correlation Parameters in the Univariate and Multivariate Models for the Female Participants	80
Table 22: Covariate Comparison for Female Subjects	82

List of Figures

	Page
Figure 1: Distribution of the Number of Childhood Visits	44
Figure 2: Distribution of the Number of Adulthood Visits.....	45
Figure 3: DBP and SBP Plots for Male Subjects (Univariate and Multivariate Models)	53
Figure 4: Evolution of the Associations (Males)	62
Figure 5: DBP and SBP Plots for Female Subjects (Univariate and Multivariate Models)	66
Figure 6: Associations of the Evolution (Females).....	74

Abstract

JOINT MIXED-EFFECTS MODELS FOR LONGITUDINAL DATA ANALYSIS:
AN APPLICATION FOR THE METABOLIC SYNDROME

By John Mercer Thorp, III, M.S.

A Thesis submitted in partial fulfillment of the requirements for the degree Masters of
Biostatistics at Virginia Commonwealth University.

Virginia Commonwealth University, 2009

Major Directors

Jessica M. Ketchum

Assistant Professor, Department of Biostatistics

Christine M. Schubert

Associate Professor, Department of Biostatistics

Edmond P. Wickham

Assistant Professor, Departments of Internal Medicine and Pediatrics

x

Mixed-effects models are commonly used to model longitudinal data as they can appropriately account for within and between subject sources of variability. Univariate mixed effect modeling strategies are well developed for a single outcome (response) variable that may be continuous (e.g. Gaussian) or categorical (e.g. binary, Poisson) in nature. Only recently have extensions been discussed for jointly modeling multiple outcome variables measures longitudinally. Many diseases processes are a function of several factors that are correlated. For example, the metabolic syndrome, a constellation of cardiovascular risk factors associated with an increased risk of cardiovascular disease and type 2 diabetes, is often defined as having three of the following: elevated blood pressure, high waist circumference, elevated glucose, elevated triglycerides, and decreased HDL. Clearly these multiple measures within a subject are not independent. A model that could jointly model two or more of these risk factors and appropriately account for between subjects sources of variability as well as within subject sources of variability due to the longitudinal and multivariate nature of the data would be more useful than several univariate models. In fact, the univariate mixed-effects model can be extended in a relatively straightforward fashion to define a multivariate mixed-effects model for longitudinal data by appropriately defining the variance-covariance structure for the random-effects. Existing software such as the PROC MIXED in SAS can be used to fit the multivariate mixed-effects model.

The Fels Longitudinal Study data were used to illustrate both univariate and multivariate mixed-effects modeling strategies. Specifically, jointly modeled longitudinal measures of systolic (SBP) and diastolic (DBP) blood pressure during childhood (ages two to eighteen) were compared between participants who were diagnosed with at least three of the metabolic syndrome risk factors in adulthood (ages thirty to fifty-five) and those who were never diagnosed with any risk factors. By identifying differences in risk factors, such as blood pressure, early in childhood between those who go on to develop the metabolic syndrome in adulthood and those who do not, earlier interventions could be used to prevent the development cardiovascular disease and type 2 diabetes.

As demonstrated by these analyses, the multivariate model is able to not only answer the same questions addressed as the univariate model, it is also able to answer additional important questions about the association in the evolutions of the responses as well as the evolution of the associations. Furthermore, the additional information gained by incorporating information about the correlations between the responses was able to reduce the variability (standard errors) in both the fixed-effects estimates (e.g. differences in groups, effects of covariates) as well as the random-effects estimates (e.g. variability).

1 Introduction

1.1. Introduction

Multivariate longitudinal data occur when multiple measures are collected on the same individual repeatedly over time. For example, systolic blood pressure (SBP) and diastolic blood pressure (DBP) measures are collected simultaneously from a patient every time they visit the doctor's office. Together these measurements give the physician an indication of the health and functioning of an individual's circulatory system at a given time point, and longitudinal measures of SBP and DBP can alert the physician to changes in the health of an individual. SBP is the pressure while the heart contracts to pump blood to the body, while DBP is the pressure when the heart relaxes between beats. Measures of SBP and DBP are highly related and changes in either often affect changes in the other. A great deal of interest then lies in how the evolution of SBP is related to the evolution of DBP, as well as how the association changes, or evolves, over time. A modeling approach that jointly models longitudinal measures of SBP and DBP over time can lend insight into these questions.

1.2. Modeling Longitudinal Data

Longitudinal data are a series of measurements of the same event taken from the same individual repeatedly over time. The most unique characteristic of longitudinal data is the ability to directly study change. The primary goal of most longitudinal studies is to characterize the change in response over time and the factors that influence this change. Great strides have been made over the past three decades involving development of statistical methodology for longitudinal data analysis. Longitudinal data require special methodology because the series of data from one subject are likely intercorrelated, and this correlation must be taken into account to draw valid statistical inferences. In fact, longitudinal data usually exhibit a positive correlation, with the strength of the association decreasing as a function of time separation (i.e. observations further apart as less correlated than those closer together). The next section will cover some basic notation.

1.2.1 Some Notation

Focusing briefly on some specific points of notation, let Y_{ij} represent a response variable and let \mathbf{x}_{ij} be a $p \times 1$ vector of explanatory variables observed at time t_{ij} , for observations $i = 1, \dots, n_j$ on subject $j = 1, \dots, S$. The mean and variance of Y_{ij} are represented by $E(Y_{ij}) = \mu_{ij}$ and $\text{Var}(Y_{ij}) = v_{ij}$, respectively. The set of repeated

outcomes for the j^{th} subject are collected into an $n_j \times 1$ vector, $\mathbf{Y}_j = (Y_{1j} \ Y_{2j} \ \dots \ Y_{n_j j})$,

with mean $E(\mathbf{Y}_j) = \boldsymbol{\mu}_j$ and $n_j \times n_j$ variance-covariance matrix $\text{Var}(\mathbf{Y}_j) = \mathbf{V}_j$, where

the ii' element of \mathbf{V}_j is the covariance between Y_{ij} and $Y_{i'j}$, denoted by

$\text{Cov}(Y_{ij}, Y_{i'j}) = v_{ii'j}$. The $n_j \times n_j$ correlation matrix of \mathbf{Y}_j is denoted by \mathbf{R}_j . The

responses for all subjects are then collected into the $N \times N$ vector

$$\mathbf{Y} = (\mathbf{Y}_1 \ \mathbf{Y}_2 \ \dots \ \mathbf{Y}_S), \text{ with } N = \sum_{j=1}^S n_j.$$

Most longitudinal analyses are based on a regression model such as the usual linear model,

$$\begin{aligned} Y_{ij} &= \beta_1 x_{ij1} + \beta_2 x_{ij2} + \dots + \beta_p x_{ijp} + \varepsilon_{ij} \\ &= \mathbf{x}_{ij}' \boldsymbol{\beta} + \varepsilon_{ij}, \end{aligned}$$

Where, $\boldsymbol{\beta} = (\beta_1 \ \beta_2 \ \dots \ \beta_p)$ is a $p \times 1$ vector of unknown regression coefficients and ε_{ij} is a zero-mean random variable representing the deviation of the response from the model prediction $\mathbf{x}_{ij}' \boldsymbol{\beta}$. Typically, $x_{ij1} = 1$ for all j and i , while β_1 is the intercept term in the linear model.

In matrix notation, the regression equation for the j^{th} subject takes the form

$$\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta} + \boldsymbol{\varepsilon}_j,$$

where \mathbf{X}_j is an $n_j \times p$ matrix with \mathbf{x}_{ij} in the i^{th} row and $\boldsymbol{\varepsilon}_j = (\varepsilon_{1j} \ \varepsilon_{2j} \ \dots \ \varepsilon_{n_j j})$.

The following section will cover some general approaches to handling longitudinal data.

1.2.2 General Approaches

Researchers have been collecting longitudinal data for centuries, and have likely been spending an equal amount of time deriving methods for the analysis of such data. An in depth historical perspective of the methods for analyzing longitudinal data can be found in Fitzmaurice (2008) and Ware and Liang (1996). Researchers can use several general approaches to model longitudinal data, as described by Diggle *et al.* (2002), Fahrmeir and Tutz (1994, 2001), and Molenberghs and Verbeke (2006); these are summarized below.

A simple and sometimes effective strategy is to reduce the series of repeated measures to one (or two) summary measure, and then analyze each summary measure as a function of the covariates, \mathbf{x}_j . This “two-stage”, or “derived variable”, analysis dates back to the early contributions of growth curve analysis by Wishart (1938), Box (1950), and Rao (1958) and agricultural experiments by Roswell and Walters (1976). The method is applicable only when the covariates are time-invariant (i.e. $\mathbf{x}_{ij} = \mathbf{x}_j$ for all i and j), since the summary value can only be regressed on \mathbf{x}_j . However, this approach is less useful if important explanatory variables change over time. While this approach has certain appeal due to its simplicity, it has a number of additional drawbacks. Clearly, there is a considerable loss of useful information. Also, subjects with different profiles can produce the same summary measures. Finally, some summary measures that have

been proposed are not well defined where there are missing data or irregularly spaced repeated measures. Rather than reduce the series of data to summary measures, there are three other modeling strategies that can be employed to model the individual Y_{ij} as a function of \mathbf{x}_{ij} .

The first strategy considered models the marginal mean. The term marginal is used here to indicate that the model for the mean response at each occasion depends only on the covariates of interest and not on random effects or previous responses. A marginal model can be specified using the framework of generalized linear models (GLM) (Nedler and Wedderburn, 1972). The model is typically specified in three parts. First, the mean of each response, $E(Y_{ij} | \mathbf{X}_{ij}) = \mu_{ij}$, is assumed to depend on the covariates through a known link function, $h^{-1}(\mu_{ij}) = \mathbf{X}_{ij}'\boldsymbol{\beta}$ (for example, the logit link for binary responses or the log link for Poisson responses). Second, the variance of each Y_{ij} , given the covariates, is assumed to depend on the mean according to $\text{Var}(Y_{ij} | \mathbf{X}_{ij}) = \phi v(\mu_{ij})$, where $v(\mu_{ij})$ is a known variance function and ϕ is a scale parameter that may be known or may need to be estimated. Third, and lastly, the conditional within-subject association among the vector of repeated responses, given the covariates, is assumed to be a function of an additional set of association parameters. This approach has the advantage of separately modeling the mean and covariance. The separation of the model for the mean response from the model for the within-subject associations ensures that the

marginal model regression coefficients have interpretations that do not depend on the assumptions made about the within-subject associations. Furthermore, valid inference about β can often still be made even when an incorrect value for V is assumed. There are limitations of the marginal model that are worth mentioning. While the marginal model can handle missing data, there are only a limited number of variance-covariance structures that can handle unbalanced data (i.e. when the data collection time points vary between subjects.) Also, the marginal model is unable to distinguish between within-subject and between-subject variability. Liang and Zeger (1986) developed generalized estimating equations (GEE) to estimate consistent estimators of the regression parameters in a marginal model, and their variances, under relatively mild conditions. While the semi-parametric GEE methods have been the dominant methods for estimation of marginal models, they are certainly not the only approach.

A second strategy which may be implemented for modeling longitudinal data is a mixed-effects, or random-effects, modeling approach (Laird and Ware, 1982). Here, the mean response depends not only on the covariates of interest, but also on a vector of random effects. The idea with this model is that some subset of the regression parameters vary randomly from one individual to the next, thereby accounting for sources of natural heterogeneity in the population. Each subject is assumed to have their own subject-specific mean response trajectory over time and a subset of the regression parameters are now regarded as being random. The linear mixed-effects model fits the mean response as a combination of population characteristics (fixed-effects) assumed to be shared by all

individuals and subject-specific effects (random-effects) that are unique to a particular individual. By including random-effects in the model, linear mixed-effects models are able to explicitly distinguish between within-subject and between-subject sources of variation. With a linear mixed-effects model it is not only possible to estimate parameters that describe how the mean responses change over time, but it is also possible to predict how an individual's response trajectories change over time. Mixed-effects models are highly attractive due to their ability to handle missing and unbalanced data reasonably well.

A third approach for modeling longitudinal data, is often referred to as the conditional models method. This method works by modeling the mean and the time dependence simultaneously by conditioning an outcome on other outcomes or on a subset of other outcomes. The most common example of a conditional model is the classical log-linear model (Agresti, 2002). A particular case of conditional models is known as the transition model. Transition models are considered conditional in the sense that they model the conditional distribution of the response at any point in time, given the previous responses and the covariates. The dependence in the repeated measures is thought to be due to the past values influence of the response the present observation. Well known transition models are Markov type models. There are many examples of the use of Markov chains to model equally spaced discrete longitudinal data with a finite number of categories (Anderson and Goodman, 1957; Cox, 1958; Billingsley, 1961). Other discussions of transition models applied to longitudinal data can be found in Cox (1972),

Korn and Whittemore (1979), Zeger *et al.* (1985), and Ware *et al.* (1988). While transition models have a long history of use with longitudinal data, their application has been limited. In general, transition models have, for the most part, only been developed for equally spaced repeated measures, and are more difficult to apply with missing data, mistimed measurements, and non-equidistant intervals between time points. Another limitation is that estimation of the regression parameters can be highly affected by the assumptions regarding time dependence. Furthermore, the effects of the covariates may be diminished when conditioning on past responses.

With all three approaches, both the dependence of the responses on the explanatory variables and the autocorrelations among the responses are modeled. There are at least three serious consequences to ignoring inherent correlations in longitudinal data (Diggle *et al.* 2002). First, incorrect inferences about the regression coefficients, β , can be made. Second, estimates of β may be inefficient (i.e. less precise than possible). Third, there is suboptimal protection against biases caused by missing data.

1.2.3 Modeling Different Types of Longitudinal Data

An important characteristic of longitudinal data is the type of outcome being measured. Methods for continuous, normally distributed data form the most developed and advanced body of research spanning the literature from the early works of British astronomer Airy (1861) through the landmark paper on linear mixed-effects models by Laird and Ware (1982). Unfortunately, researchers are often confronted with different

types of data such as binary, multinomial, or count. One view researchers take, supported by large sample results, is that normal theory should be used as much as possible, even with regards to non-normal data such as ordinal scores and counts (Fitzmaurice, *et al.*, 2009). A more predominant view is that each type of outcome should be analyzed using methods that exploit the nature of the data.

The GLM described earlier for the marginal models is able to handle any type of response in the exponential family (i.e. Gaussian, binary, Poisson) and has been extended to allow for random-effects. In the same sense that the linear model defined in section 1.2.1 is a special case of the GLM, the linear mixed-effects model is a special case of the generalized linear mixed-effects model (GLMM). In this case, given a vector of random-effects, the responses are assumed to be conditionally independent and to have distributions from the exponential family. The conditional mean depends on both fixed-effects and random-effects for some known link function. The conditional variance is assumed to depend on the conditional mean. Finally, the random-effects are assumed to be independent of the covariates and to have a zero-mean multivariate normal distribution. Work by Ashford and Sowden (1970), Pierce and Sands (1975), and Korn and Whittemore (1979) laid the framework for GLMM's with much of work that followed focusing on the issues of estimation. It is worth noting, however, that the extensions to GLMM's can have important implications for the interpretation of the regression coefficients. In GLMM's, the regression coefficients have subject-specific interpretations and represent the effects of the covariates on changes in an individual's

(possibly transformed) mean response per unit change in the covariate (Molenberghs and Verbeke, 2006; Fitzmaurice *et al.*, 2009). With marginal models, on the other hand, the regression coefficients have population based interpretations and represent the effects of covariates on changes in the population's (possibly transformed) mean response per unit change in the covariate. In the special case of the linear model, where an identity link function is assumed, the fixed-effects in the model for the conditional mean also happen to have interpretations in terms of the population means.

The GLM has also been extended to handle most types of conditional transition models and many of the previously cited literature deals with non-Gaussian data. An extensive discussion on the extensions can be found in Molenberghs and Verbeke part III (2006).

1.3. Modeling Joint Longitudinal Data

When there are multiple outcomes measured repeatedly over time, the most basic approach would be to model each longitudinal outcome independently. However, this does not account for the multivariate nature of the data. As compared to traditional univariate approaches, methods for analyzing multivariate longitudinal data can be challenging for several reasons. For one, the variability for each response is likely to be different. Also, in addition to the correlation due to repeated measures over time, there is likely to be correlations between the outcomes. The primary objective of joint modeling is to provide a framework where questions of scientific interest pertaining to relationships

among and between the multiple outcomes and other factors (i.e. treatment, dose, or covariates) may be formalized. In order for there to be valid inference, the joint models must be able to appropriately account for correlations within and between each outcome. General approaches to what has been discussed thus far, will be cover next.

1.3.1 General Approaches

Consider modeling two outcomes Y_1 and Y_2 , where one, or both, may have been collected longitudinally. Here, attention will be restricted to two outcomes, noting that extensions for more than two outcomes are relatively straightforward. There are several methods available in the literature for jointly modeling longitudinal data. Four of these are briefly discussed below, while more thorough descriptions can be found in recent texts by Fitzmaurice *et al.* (2009) and Molenberghs and Verbeke (2006).

One strategy is referred to as the multivariate marginal models (Galecki, 1994; Molenberghs and Verbeke, 2006). The idea here is to directly specify the joint density $f(y_1, y_2)$ of (Y_1, Y_2) . The model includes the necessary assumptions about the marginal association among the longitudinal measurements within each vector Y_1 and Y_2 , as well as assumptions of the nature of the associations between elements of Y_1 and Y_2 . This may become difficult if Y_1 and Y_2 are of different types (e.g. continuous and binary, or continuous and time-to-event) or in the case of highly unbalanced data. Furthermore, extensions beyond two outcomes are especially

challenging since assumptions about larger, more complex association structures must be made. The primary attraction to this model is the ability to directly make inferences about the marginal characteristics of the outcomes (e.g. average evolutions).

A second strategy is the conditional model described in section 1.2.2. Here, direct specification of a joint distribution for (Y_1, Y_2) is avoided by factorizing the density into a product of a marginal and a conditional density. That is,

$$f(y_1, y_2) = f(y_1 | y_2) f(y_2) = f(y_2 | y_1) f(y_1).$$

This reduces the problem to specifications of the models for each of the outcomes separately, with a marginally specified model for one outcome, and a conditionally specified model for the other. Careful attention must be given to the choice of which outcome is modeled marginally and which is modeled conditionally. For example, it may be plausible that one plays the role of a time-dependent covariate. Different choices can lead to very different, and perhaps opposite, results and conclusions. Another limitation with this methodology is that it does not directly lead to marginal inferences. Finally, when both outcomes are highly correlated and thought to be manifestations of a common underlying treatment effect, conditioning on one will attenuate the treatment effect of the other. Therefore, conditional models are not the preferred approach when there are more than two outcomes since there may be many more possible factorizations. Examples of conditional models used to jointly model a continuous and binary outcome can be found in Tate (1954), Olkin and Tate (1961), Little and Schluchter (1985), Krzanowski (1988), and Cox and Wermuth (1992, 1994).

A third strategy is the shared-parameters modeling approach. As described earlier, random-effects can be introduced into a model to account for associations in the longitudinal measures. This same idea can be extended to account for additional associations in the multivariate longitudinal data. For the shared-parameters model, define $\boldsymbol{\beta}$ as a vector of random-effects, common to the model for \mathbf{Y}_1 and the model for \mathbf{Y}_2 , and assume independences of both outcomes, conditionally on $\boldsymbol{\beta}$. The joint density of $(\mathbf{Y}_1, \mathbf{Y}_2)$ is then obtained from

$$f(y_1, y_2) = \int f(y_1, y_2 | \boldsymbol{\beta}) f(\boldsymbol{\beta}) d\boldsymbol{\beta} = \int f(y_1 | \boldsymbol{\beta}) f(y_2 | \boldsymbol{\beta}) f(\boldsymbol{\beta}) d\boldsymbol{\beta},$$

where $f(\boldsymbol{\beta})$ denotes the marginal density of the random effects. In this formulation, the random-effects, $\boldsymbol{\beta}$, is a “shared-parameter” that induces a correlation between \mathbf{Y}_1 and \mathbf{Y}_2 through their joint dependence of $\boldsymbol{\beta}$. That \mathbf{Y}_1 and \mathbf{Y}_2 are conditionally independent given the random-effects, $\boldsymbol{\beta}$, is interpreted as a belief that a common set of underlying characteristics of the individual governs both outcome processes. An advantage of this type of model is that \mathbf{Y}_1 and \mathbf{Y}_2 do not need to be of the same type (e.g. \mathbf{Y}_1 could be continuous responses and \mathbf{Y}_2 could be binary responses). Another advantage is that the parameters in the joint shared-parameters model have the same interpretations as they do in each of the corresponding “univariate” models. In addition, extensions for more than two outcomes are straightforward and because dimensionality in the above integration does not increase, computational intensity does not increase. This type of model has been

used repeatedly in the literature to jointly model a longitudinal outcome and a time-to-event outcome (Degruittola and Tu, 1994; Tsiatis *et al.*, 1995; Faucett and Thomas, 1996; Tsiatis and Davidian, 2004). The main disadvantage of the shared-parameters approach is that it implies very strong assumptions about the association between the outcomes being modeled. Specifically, the correlation between pairs of measurements from different outcomes must be equal to the product of the correlation between measurements from the first outcome and measurements from the second outcome (Fitzmaurice *et al.*, 2009).

That is,

$$\text{Corr}\{Y_1(s), Y_2(t)\} = \sqrt{\text{Corr}\{Y_1(s), Y_1(t)\}} \sqrt{\text{Corr}\{Y_2(s), Y_2(t)\}}.$$

As such, the shared-parameter model may not accurately represent the association structure of the data. This motivated the need for more flexible correlation patterns, at the expense of model complexity.

A fourth, and final, approach is the random-effects model method. The assumptions in the correlation structure in the shared parameters model could be relaxed by allowing the models for \mathbf{Y}_1 and \mathbf{Y}_2 to depend on separate vectors of the random-effects parameters, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, which are themselves correlated. Under these assumptions,

$$\text{Corr}\{Y_1(s), Y_2(t)\} \leq \sqrt{\text{Corr}\{Y_1(s), Y_1(t)\}} \sqrt{\text{Corr}\{Y_2(s), Y_2(t)\}}.$$

Random-effects models can easily be extended to model more than two outcomes, however, the dimensionality of the vector of random-effects increases as the number of

outcomes increases, leading to computational difficulties. Pair-wise modeling approaches may be a more appropriate means to side-step the issue of dimensionality (Verbeke *et al.*, 2006).

1.3.2 Choosing Between the Univariate Models and a Multivariate Model

The availability of the joint model does not necessarily mean that a researcher must use one. It is possible that several univariate models will suffice to answer the research question, or questions, at hand. In other cases, it is necessary to combine all of the responses into a single model in order to draw specific inferences. Sometimes, the association structure may be of direct importance. For example, a researcher may be interested in how the changes in one response are associated with changes in another response (the association in the evolution). There may also be interest in how the association between the responses changes or evolves over time (the evolution in the associations). Analyses pertaining to these questions have been described in Fieuws and Verbeke (2004, 2006) in bivariate and multivariate settings. A joint model would also be necessary if a researcher were interested in testing fixed-effects referring to a set of outcomes simultaneously. For example, the definition of success of an intervention may depend on more than one response variable. Joint models may also have additional benefits for prediction purposes. Furthermore, the additional information from multiple outcomes may better discriminate between groups (Fieuws *et al.*, 2007). When outcomes share parameters, a joint model can lead to a gain in efficiency. Even when all parameters

are outcome-specific, there is a potential gain in efficiency when there are missing data. For example, one may be able to assume the data are missing at random in a multivariate model, but not be able to do so when looking at each outcome separately.

1.4. The Metabolic Syndrome

The primary focus of this thesis is to jointly model two continuous outcomes as a function of fixed-effects and random-effects. More specifically, there is interest in modeling longitudinal measures of childhood SBP and DBP, as a function of birth weight and physical maturation, to determine if there exist differences in the childhood profiles between those who go on to develop the metabolic syndrome later in life and those who never develop any of the risk factors associated with the metabolic syndrome. The following sections give a definition of the metabolic syndrome and describe the measurement procedures for the risk factors associated with this syndrome, along with a more formal definition of the purpose of this thesis.

1.4.1 Diagnosis of the Metabolic Syndrome

The metabolic syndrome, which is a constellation of cardiovascular risk factors, has also been referred to as syndrome X (Raven, 1993), the “deadly quartet” (Kaplan, 1989), insulin resistance syndrome (Stern., 1994), and hypertriglyceride waist (Lemieux, *et al.*, 2000), and has been defined by several major organizations over the years. The term, metabolic syndrome, is most commonly used in the cardiovascular field. The metabolic

syndrome has been receiving increased attention over the past years in part due to its promotion of atherosclerotic cardiovascular disease (ASCVD). The metabolic syndrome is generally classified as a “constellation of interrelated risk factors of metabolic origin--*metabolic risk factors*--that appear to directly promote the development of atherosclerotic cardiovascular disease” (Grundy, *et al.*, 2005). Furthermore, it has been proposed that not only is the metabolic syndrome related to ASCVD, but the risk for ASCVD associated with the metabolic syndrome risk factors is more than the sum of its parts (Grundy, 2006). There is also a relationship between the metabolic syndrome and an increased risk of developing type 2 diabetes mellitus (Grundy, *et al.*, 2005). Both ASCVD and type 2 diabetes mellitus are growing health concerns that plague numerous Americans as the prevalence of obesity grows in the United States. The Center for Disease Control and Prevention defines obesity for an individual as that person having a body mass index (BMI) of thirty or greater (note, BMI is calculated from a person’s weight and height and provides a reasonable indicator of body fatness and weight). In 2008, the CDC determined that forty-nine states in America had a prevalence of obesity of more than 20%. These numbers also appear to be on the rise. The metabolic syndrome takes into account risk factors other than just BMI, making it a potential barometer of future medical comorbidities related to ASCVD and type 2 diabetes mellitus.

Given the hefty sums of money spent each year on health problems associated with the metabolic syndrome, ASCVD, and type 2 diabetes mellitus, it is of great interest to investigate methods for the early prediction for the metabolic syndrome. Not only is

obesity on the rise amongst the general U.S. population, it is also on the rise amongst children. According to a July report by the National Center for Health Statistics, 15% of children ages six to eighteen were overweight in 2000, this percentage is up from just 6% in 1980. Given these alarming numbers regarding the prevalence of obesity in children, it is desirable to investigate a predictive model that would be capable of identifying children who are at increased risk of developing the metabolic syndrome as adults. Early detection is can be important in combating most health related problems an individual may develop. Therefore, identifying those between the ages of two and eighteen who may be at increased risk to develop the metabolic syndrome as adults is of significant interest in the early detection and prevention of cardiovascular disease and type 2 diabetes.

Though, many proposals have been made in the past decade to best define what set of criteria determines if an individual has the metabolic syndrome, only one will be considered for the purposes of this analysis; the National Cholesterol Education Program (NCEP) Adult Treatment Panel III (ATP III) criteria. Historically, the two underlying risk factors that appear to be most serious are abdominal obesity (Lemieux *et al.*, 2000) and insulin resistance (Park *et al.*, 2003). Many other possible associated factors have been considered over the years including physical inactivity (Gustat *et al.*, 2002), aging (Ford *et al.*, 2002), and hormonal imbalance (Apridonidze, *et al.*, 2004).

Despite the numerous definitions proposed over the years since the metabolic syndrome was first proposed as a medical diagnosis, the definition which will be

considered for the purposes of this thesis is that proposed in 2001 by the NCEP ATP III. This was the first proposed definition that did not specifically required documentation of an elevated insulin resistance, noting that direct measures of insulin resistance are laborious and not well standardized (Grundy *et al.* 2005). Since the ATP III criteria for the metabolic syndrome removed the requirement of insulin resistance, there was now no single factor necessary for diagnosis. Instead, five risk factors are considered under the ATP III criteria, they are as follows: elevated waist circumference, reduced HDL-c, elevated serum glucose, elevated triglycerides, and elevated blood pressure. The presence of at least three out of the aforementioned five risk factors is the basis for diagnosing the metabolic syndrome under the ATP III proposal. The ATP III criteria for the metabolic syndrome will be used given that it is the most simple to use in a clinical setting and also has the benefit of not emphasizing a single risk factor. So, in summary, the metabolic syndrome shall be defined here as having any three of the five risk factors in accordance with the ATP III criteria.

For each risk factor outlined by the ATP III criteria, there is a categorical cut point, for certain risk factors these are the same for males and females, for other risk factors the cut points are different. If an individual falls above (or below in the case of HDL-c) this cut point, they are considered to have that particular risk factor. These cut points are laid out in Table 1, which is followed by a more detailed definition of each risk factor.

Table 1: ATP III Criteria Categorical Cut Points for the Metabolic Syndrome

	Male	Female
Waist Circumference	≥ 102 cm (≥ 40 inches)	≥ 88 cm (≥ 35 inches)
HDL-c	< 40 mg/dL (1.03 mmol/L)	< 50 mg/dL (1.3 mmol/L)
Serum Glucose	≥ 100 mg/dL	
Triglycerides	≥ 150 mg/dL (1.7 mmol/L)	
Systolic Blood Pressure	≥ 130 mm Hg	
Diastolic Blood Pressure	≥ 85 mm Hg	
Having either SBP or DBP ≥ the cut point, means having the blood pressure risk factor		

First considering the risk factor of waist circumference, literature pertaining to the metabolic syndrome gives the following recommendation for waist measurement:

“...locate top of right iliac crest. Place a measuring tape in a horizontal plane around the abdomen at level of iliac crest.” (Grundy *et al.*, 2005). The second metabolic syndrome

risk factor, fasting serum glucose, is measured by collecting a serum sample for

measuring glucose after an individual has abstained from food for a set period of time.

Fasting triglycerides are measured using a simple standardized blood test know as a

lipoprotein panel; which is used to determine an individual’s triglyceride levels as well as their cholesterol levels. From the lipoprotein panel HDL-c, the “good cholesterol”, which

carries cholesterol from the body’s tissues to the liver, and can protect against heart

attack and stroke, is measured (recall, HDL-c is the only metabolic syndrome risk factor

with a minimum categorical cut point). Finally, blood pressure, determined by measures

of SBP and DBP, make up the fifth metabolic syndrome risk factor. The systolic blood

pressure measurement is a measure of the peak pressure in the arteries, while the diastolic

blood pressure is a measurement of the minimum pressure in the arteries. One specific instrument used to measure blood pressure is a standard mercury sphygmomanometer.

1.5. Jointly Modeling Risk Factors of the Metabolic Syndrome over Time

Now that both the metabolic syndrome and the methodology for analyzing joint longitudinal data have been introduced, the statistical and biological questions of interest can be better described. It is of interest to see if those individuals who develop the metabolic syndrome in adulthood and those who develop no risk factors in adulthood, have significantly different SBP and DBP profiles during childhood. The question will be addressed separately for males and females. These models will adjust for both time-invariant (e.g. sex and birth weight) and time-variant (physical maturation) covariates. Furthermore, it is of interest to consider both univariate mixed-effects models designed to handle each outcome measure individually and a single multivariate mixed-effects model with a multivariate outcome representation.

This analysis will probe the differences between multiple univariate models and a single multivariate model, allowing discussion between the pros and cons of both methods, as well what may be gained, or lost, using one versus the other. The metabolic syndrome is ideal for this type of analysis given it is inherently a syndrome comprised of multiple outcomes, lending itself to a multivariate mixed-effects model.

2 Methodology

2.1. Introduction

In this chapter, the methods for modeling longitudinal data with a univariate normal mixed-effects (random-effects) model are discussed, along with the methods for jointly modeling multivariate longitudinal data. For the purposes of this thesis, the terms mixed-effects and random-effects will be used interchangeably, and are considered to be the same. In section 2.2, the univariate normal random-effects model is introduced and defined, along with the notation which will be used henceforth. In section 2.3, a discussion of estimation in normal mixed-effects models is presented. The joint mixed-effects model will be presented in section 2.4, with subsequent sections focusing on special cases of the \mathbf{G} matrix, the association of the evolutions, and the evolution of the association.

2.2. Univariate Normal Mixed-Effects Model Definition

The normal random-effects model introduced by Laird and Ware (1982) can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.1)$$

where

\mathbf{y} is the $N \times 1$ vector of observed response values,

$\boldsymbol{\alpha}$ is the $P \times 1$ vector of fixed-effects parameters,

\mathbf{X} is the $N \times P$ observed design matrix corresponding to the fixed-effects,

$\boldsymbol{\beta}$ is the $Q \times 1$ vector of random-effects parameters,

\mathbf{Z} is the $N \times Q$ observed design matrix corresponding to the random-effects, and

$\boldsymbol{\varepsilon}$ is the $N \times 1$ vector of residuals.

A corresponding assumption of model (2.1) is, $\boldsymbol{\beta} \sim N_Q(\mathbf{0}, \mathbf{G})$; that is $\boldsymbol{\beta}$ has a Q -variate normal density with mean vector $\mathbf{0}$ and a variance-covariance matrix \mathbf{G} . Furthermore, we assume $\boldsymbol{\beta}$ is independent of the vector of residuals, $\boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim N_N(\mathbf{0}, \mathbf{R})$; that is $\boldsymbol{\varepsilon}$ has an N -variate normal density with mean vector $\mathbf{0}$ and variance covariance matrix \mathbf{R} . Given the assumptions listed above, $\mathbf{y} \sim N_N(\mathbf{X}\boldsymbol{\alpha}, \mathbf{V})$; that is \mathbf{y} has a N -variate normal density with mean vector $\mathbf{X}\boldsymbol{\alpha}$ and variance-covariance matrix \mathbf{V} . For the above, note that

$$\mathbf{V} = \text{Var}[\mathbf{y}] = \text{Var}[\mathbf{Z}\boldsymbol{\beta}] + \text{Var}[\boldsymbol{\varepsilon}] = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}.$$

2.3. Estimation in the Normal Random-Effects Model

In this section, the methods used for estimation in the normal random-effects model, described above in section 2.2, are discussed. Section 2.3.1, will cover the likelihood function and approaches to its maximization, more specifically the maximum likelihood method and the residual maximum likelihood method. The subsequent three sections will cover estimation of the fixed-effects parameters, random-effects parameters, and variance parameters.

2.3.1 Methods of Estimation

Suppose a random sample of N observations is obtained from a univariate normal random-effects model as defined in equation (2.1), then the likelihood of the model parameters, given the vector of N observations, is defined as

$$L = L(\boldsymbol{\alpha}, \boldsymbol{\gamma}; \mathbf{y}) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha})\right\}}{(2\pi)^{(1/2)N} |\mathbf{V}|^{(1/2)}}, \quad (2.2)$$

where $\boldsymbol{\alpha}$ is a vector of fixed-effects parameters and $\boldsymbol{\gamma}$ is a vector containing the variance parameters. Given its simplicity in comparison to the likelihood function, the log of the likelihood function is generally used in practice. Its maximum value coincides with that of the likelihood function. The log-likelihood of the model parameters, is defined as

$$\begin{aligned} l = l(\boldsymbol{\alpha}, \boldsymbol{\gamma}; \mathbf{y}) &= -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha}) \\ &= K - \frac{1}{2} \left[\log |\mathbf{V}| + (\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha}) \right]. \end{aligned} \quad (2.3)$$

where, $K = -\frac{1}{2} N \log(2\pi)$, which is a constant that may be ignored in the maximization process. Now the values in the model parameters which maximize the log-likelihood may be determined. Estimates of the parameters for the model given in equation (2.1) are found by maximizing the log-likelihood given in equation (2.3) with respect to α and γ . One such method that may be used to maximize the log-likelihood function is the maximum likelihood (ML) method. The ML method, first maximizes the log-likelihood with respect to the variance parameters, while treating the fixed-effects parameters, α , as constant. Upon determining the variance parameter estimates, the fixed-effects parameters are then determined by finding the values of α which maximize the log-likelihood, while treating the variance parameters as constant. It is important to note, the maximum likelihood approach may produce variance parameters that are biased downwards since they are based on the assumption that the fixed-effects parameters are known (Brown and Prescott, 2006).

Another method that may be used to maximize the log-likelihood function is the residual maximum likelihood (REML) method. Sometimes this method is referred to as the restricted maximum likelihood method. For this approach, the fixed-effects parameters, α , are eliminated from the log-likelihood equation, such that it will only be defined in terms of the variance parameters. Then, a likelihood function based on the full residuals, $(\mathbf{y} - \mathbf{X}\hat{\alpha})$, instead of the ordinary residuals, $(\mathbf{y} - \mathbf{X}\hat{\alpha} - \mathbf{Z}\hat{\beta})$ needs to be determined. It may be noted that the full residuals are a linear combination of \mathbf{y} and

furthermore $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})$ and $\hat{\boldsymbol{\alpha}}$ are independent (Diggle *et al.*, 1994). From these facts, the joint-likelihood for $\boldsymbol{\alpha}$ and the variance parameters, $\boldsymbol{\gamma}$, may be express as a product of the likelihoods based on $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})$ and $\hat{\boldsymbol{\alpha}}$

$$L(\boldsymbol{\gamma}, \boldsymbol{\alpha}; \mathbf{y}) = L(\boldsymbol{\gamma}; \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})L(\boldsymbol{\alpha}; \hat{\boldsymbol{\alpha}}, \boldsymbol{\gamma}). \quad (2.4)$$

Thus,

$$L(\boldsymbol{\gamma}; \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}) = \frac{L(\boldsymbol{\gamma}, \boldsymbol{\alpha}; \mathbf{y})}{L(\boldsymbol{\alpha}; \hat{\boldsymbol{\alpha}}, \boldsymbol{\gamma})}. \quad (2.5)$$

From the above, and equation (2.2),

$$L(\boldsymbol{\gamma}, \boldsymbol{\alpha}; \mathbf{y}) \propto |\mathbf{V}|^{-1} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})\right\}. \quad (2.6)$$

Furthermore, $\hat{\boldsymbol{\alpha}}$, has a multivariate normal distribution with mean and variance given by the maximum likelihood estimates discussed later in equations (2.11), and (2.12), respectively. Hence,

$$L(\boldsymbol{\alpha}; \hat{\boldsymbol{\alpha}}, \boldsymbol{\gamma}) \propto |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|^{1/2} \exp\left\{\frac{1}{2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})\right\}. \quad (2.7)$$

Taking the ratio of equations (2.6) and (2.7), yields the REML, defined as

$$L(\boldsymbol{\gamma}; \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}) \propto |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|^{-1/2} |\mathbf{V}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})\right\}. \quad (2.8)$$

Therefore, the REML log-likelihood is defined as

$$\log L(\boldsymbol{\gamma}; \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}) = K - \frac{1}{2} \left\{ \log |\mathbf{V}| - \log |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|^{-1} + (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}) \right\}. \quad (2.9)$$

Despite $\hat{\alpha}$ appearing in the REML log-likelihood in equation (2.9), it is present only as a function of the variance parameters. As with the maximum likelihood method, the variance parameters are now estimated by maximizing the REML log-likelihood given in equation (2.9), with regards to the variance parameters. Given the nature of the REML likelihood, and its treatment of the fixed-effects as parameters, rather than as constants, the resulting variance parameter estimates are unbiased. In the same fashion of the maximum likelihood method, values of α are found by maximizing the REML log-likelihood with regards to the fixed-effects parameters, while treating the variance parameters as fixed.

2.3.2 Estimation of Fixed-Effects Parameters

For both the maximum log-likelihood method and the REML log-likelihood methods the fixed effects solutions may be calculated by maximizing either likelihood by differentiating the log-likelihood with respect to α , and subsequently setting the resulting expression to zero. That is,

$$\mathbf{X}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\alpha) = \mathbf{0}. \quad (2.10)$$

The solutions to this equation are the maximum likelihood estimates for the fixed-effects parameters:

$$\hat{\alpha} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \quad (2.11)$$

The variance of $\hat{\alpha}$ is given by,

$$\begin{aligned}
\text{var}[\hat{\boldsymbol{\alpha}}] &= \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{V}^{-1} \text{var}[\mathbf{y}] \mathbf{V}^{-1}\mathbf{X} \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1} \\
&= \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{V}^{-1} \mathbf{V} \mathbf{V}^{-1} \mathbf{X} \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1} \\
&= \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}.
\end{aligned} \tag{2.12}$$

This expression assumes that \mathbf{V} is known when in fact \mathbf{V} is estimated. Thus there may be some downward bias in the variance of $\hat{\boldsymbol{\alpha}}$, although this is usually small (Brown and Prescott, 1999).

2.3.3 Estimation of Random-Effects Parameters

Recall, $\boldsymbol{\beta}$, the $Q \times 1$ vector of random-effects, is assumed to follow a Q -variate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix \mathbf{G} , $\boldsymbol{\beta} \sim N_Q(\mathbf{0}, \mathbf{G})$. The specific values of the random-effects must be thought of as realizations of a sample from a distribution. Therefore their expected values are zero, by definition. However, it is possible to obtain predictions (“estimates”) of them. To obtain these estimates, a joint likelihood function in terms of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ is defined. This is expressed by taking the product of the likelihoods for $\mathbf{y} | \boldsymbol{\beta}$ and $\boldsymbol{\beta}$ as

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}; \mathbf{y}) = L(\boldsymbol{\alpha}, \boldsymbol{\gamma}; \mathbf{R}; \mathbf{y} | \boldsymbol{\beta}) L(\boldsymbol{\gamma}; \mathbf{G}; \boldsymbol{\beta}). \tag{2.13}$$

The likelihood of $\boldsymbol{\beta}$ is written as

$$L(\boldsymbol{\gamma}; \mathbf{G}; \boldsymbol{\beta}) = (2\pi)^{-Q/2} |\mathbf{G}|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta})' \mathbf{G} (\boldsymbol{\beta})\right\}. \tag{2.14}$$

The likelihood for $\mathbf{y} | \boldsymbol{\beta}$ is written as

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}; \mathbf{y} | \boldsymbol{\beta}) \propto |\mathbf{R}|^{-1/2} \exp \left\{ -\frac{1}{2} \left((\mathbf{y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{Z}\boldsymbol{\beta})' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{Z}\boldsymbol{\beta}) \right) \right\}. \quad (2.15)$$

Therefore, the joint likelihood for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ is defined as

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}; \mathbf{y}) \propto |\mathbf{R}|^{-1/2} |\mathbf{G}|^{-1/2} \exp \left\{ -\frac{1}{2} \left[(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{Z}\boldsymbol{\beta})' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{Z}\boldsymbol{\beta}) + \boldsymbol{\beta}' \mathbf{G}^{-1} \boldsymbol{\beta} \right] \right\}. \quad (2.16)$$

From equation (2.16) the corresponding log-likelihood is expressed as

$$l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}; \mathbf{y}) \propto -\frac{1}{2} \{ \log |\mathbf{R}| + \log |\mathbf{G}| \} - \frac{1}{2} \left\{ (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{Z}\boldsymbol{\beta})' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{Z}\boldsymbol{\beta}) + \boldsymbol{\beta}' \mathbf{G}^{-1} \boldsymbol{\beta} \right\}. \quad (2.17)$$

To obtain the maximum likelihood solution for the random-effects parameter, $\boldsymbol{\beta}$, the derivative of the log-likelihood, or REML log-likelihood, is taken with respect to $\boldsymbol{\beta}$, and the subsequent expression is set to zero. That is,

$$\begin{aligned} \frac{\partial l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}; \mathbf{y})}{\partial \boldsymbol{\beta}} &= \mathbf{Z}' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha} - \mathbf{Z}\boldsymbol{\beta}) - \mathbf{G}^{-1} \boldsymbol{\beta} \\ &= \mathbf{Z}' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}) - \left(\mathbf{Z}' \mathbf{R} \mathbf{Z} + \mathbf{G}^{-1} \right) \boldsymbol{\beta}. \end{aligned} \quad (2.18)$$

Setting the above expression to zero and using the fact that $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$,

yields the maximum likelihood solution for $\hat{\boldsymbol{\beta}}$.

$$\begin{aligned}
\hat{\beta} &= \left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \right)^{-1} \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\alpha) \\
&= \left[\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \right)^{-1} \left(\mathbf{R}\mathbf{Z}'^{-1} \right)^{-1} \right] (\mathbf{y} - \mathbf{X}\alpha) \\
&= \left[\left(\mathbf{R}\mathbf{Z}'^{-1} \right) \left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \right) \right]^{-1} (\mathbf{y} - \mathbf{X}\alpha) \\
&= \left[\mathbf{Z} + \mathbf{R}\mathbf{Z}'^{-1}\mathbf{G}^{-1} \right]^{-1} (\mathbf{y} - \mathbf{X}\alpha) \tag{2.19} \\
&= \left[(\mathbf{V} - \mathbf{R})\mathbf{Z}'^{-1}\mathbf{G}^{-1} + \mathbf{R}\mathbf{Z}'^{-1}\mathbf{G}^{-1} \right]^{-1} (\mathbf{y} - \mathbf{X}\alpha) \\
&= \left[\mathbf{V}\mathbf{Z}'^{-1}\mathbf{G}^{-1} \right]^{-1} (\mathbf{y} - \mathbf{X}\alpha) \\
&= \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\alpha).
\end{aligned}$$

The variance of $\hat{\beta}$ is given by

$$\begin{aligned}
\text{var}[\hat{\beta}] &= \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1} \text{var}[\mathbf{y} - \mathbf{X}\alpha] \mathbf{V}^{-1}\mathbf{Z}\mathbf{G} \\
&= \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1} \text{var} \left[\mathbf{y} - \mathbf{X} \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \right] \mathbf{V}^{-1}\mathbf{Z}\mathbf{G} \\
&= \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1} \text{var} \left[\left(\mathbf{I} - \mathbf{X} \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{V}^{-1} \right) \mathbf{y} \right] \mathbf{V}^{-1}\mathbf{Z}\mathbf{G} \\
&= \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1} \left[\mathbf{I} - \mathbf{X} \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{V}^{-1} \right] \mathbf{V} \left[\mathbf{I} - \mathbf{V}^{-1}\mathbf{X} \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}' \right] \mathbf{V}^{-1}\mathbf{Z}\mathbf{G} \tag{2.20} \\
&= \left[\mathbf{G}\mathbf{Z}' - \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{X} \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}' \right] \mathbf{V}^{-1} \left[\mathbf{Z}\mathbf{G} - \mathbf{X} \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{G} \right] \\
&= \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1} - \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{X} \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{G}.
\end{aligned}$$

Once again, the variance of $\hat{\beta}$ will tend to have a small downward bias since it is assumed that V is known (Brown and Prescott, 1999).

2.3.4 Estimation of Variance Parameters

The variance parameters are also obtained by maximizing the log-likelihood function. The derivatives of the log-likelihood functions with respect to the variance parameters, however, are nonlinear. Thus, an iterative approach, such as the Newton-Raphson algorithm, is often used to find the maximum likelihood solutions. Variance-covariance estimates of the variance parameters for any given structure can also be obtained by using large sample theory. These estimates are based on asymptotic theory and should be interpreted with caution.

2.4. Multivariate Methods

The mixed-model in equation (2.1) can be easily extended to include multiple response variables by further stacking the data and defining a specific variance-covariance structure for the random effects. For simplicity consider modeling two response variables (Y_1 and Y_2) over time and incorporating random intercepts and slopes in order to model the correlations over time between responses.

2.5. Multivariate Normal Random-Effects Model Definition

Let y_{ijk} represent the i^{th} observation, from the j^{th} subject, for the k^{th} response variable,

where $i = 1, \dots, n_{jk}, j = 1, \dots, S$, and $k = 1, \dots, K$. Also, define $N_k = \sum_{j=1}^S n_{jk}$ and

$N = \sum_{k=1}^K N_k$. The vector $\mathbf{Y}_{jk} = [y_{1jk} \quad y_{2jk} \quad \dots \quad y_{n_{jk}jk}]'$ then represents the n_{jk}

observations of the k^{th} response variable from the j^{th} subject and the vector

$\mathbf{Y}_k = [\mathbf{Y}_{1k} \quad \mathbf{Y}_{2k} \quad \dots \quad \mathbf{Y}_{Sk}]'$ represents the N_k observations for the k^{th} response

variable across all subjects. Finally, the vector $\mathbf{Y} = [\mathbf{Y}_1 \quad \mathbf{Y}_2 \quad \dots \quad \mathbf{Y}_K]'$ represents the N

observations across all response variables and subjects.

In the context of modeling two response variables, the linear mixed-effects models for each response variable for subject j taken at time t can be specified as

$$\begin{aligned} Y_{j1}(t) &= \mu_1(t) + a_{j1} + b_{j1}(t) + \varepsilon_{j1}(t) \\ Y_{j2}(t) &= \mu_2(t) + a_{j2} + b_{j2}(t) + \varepsilon_{j2}(t) \end{aligned} \quad (2.21)$$

where $\mu_k(t)$ refers to the average evolution (of the k^{th} response over time) and is a function of the fixed effects. The subject specific random intercepts a_{jk} and slopes $b_{jk}(t)$ describe how the subject specific profiles deviate from the average profile for the k^{th} response. The two response trajectories are joined together by assuming a joint distribution for the vector of random-effects, $\boldsymbol{\beta}$, such as

$$\boldsymbol{\beta} = \begin{bmatrix} a_{j1} \\ b_{j1} \\ a_{j2} \\ b_{j2} \end{bmatrix} \sim N(\mathbf{0}, \mathbf{G}),$$

where the variance-covariance matrix for the random effects, \mathbf{G} , has the following structure:

$$\mathbf{G} = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_1 b_1} & \sigma_{a_1 a_2} & \sigma_{a_1 b_2} \\ \sigma_{b_1 a_1} & \sigma_{b_1}^2 & \sigma_{b_1 a_2} & \sigma_{b_1 b_2} \\ \sigma_{a_2 a_1} & \sigma_{a_2 b_1} & \sigma_{a_2}^2 & \sigma_{a_2 b_2} \\ \sigma_{b_2 a_1} & \sigma_{b_2 b_1} & \sigma_{b_2 a_2} & \sigma_{b_2}^2 \end{bmatrix}. \quad (2.22)$$

The error components for each response, which are independent of the random effects, can be taken to be correlated or uncorrelated ($\sigma_{12} = 0$), such that the error components are defined as;

$$\begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \right).$$

Assuming $\sigma_{12} = 0$ implies that, conditional on the random-effects, both response trajectories are independent. The assumption of conditional independence could alternatively be relaxed and the random errors could be taken to be dependent by allowing for a nonzero covariance between the error components ($\sigma_{12} \neq 0$).

2.5.1 Special Cases for the \mathbf{G} Matrix

Special case can now be obtained by making specific assumptions for the variance-covariance matrix \mathbf{G} . Two such specific variance-covariance structures are described in the following subsections, a complete independence structure and a shared-parameters structure.

2.5.1.1 Complete Independence

The two response variable could be taken to be completely independent at any point in time, thereby imposing the following structure for \mathbf{G} :

$$\mathbf{G} = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_1 b_1} & 0 & 0 \\ \sigma_{b_1 a_1} & \sigma_{b_1}^2 & 0 & 0 \\ 0 & 0 & \sigma_{a_2}^2 & \sigma_{a_2 b_2} \\ 0 & 0 & \sigma_{b_2 a_2} & \sigma_{b_2}^2 \end{bmatrix}.$$

Within a response variable, the random intercept and slope induce within-subject correlations in the repeated measures over time, while assuming independence between subjects. Moreover, this model assumes that the two responses are completely independent. The results for the model would be identical, in theory, to fitting two separate random-effect models.

2.5.1.2 Shared-Parameters

Now that a complete independence structure has been considered for the \mathbf{G} matrix one may consider the other end of the spectrum where the two response variables could be taken to be completely dependent. In this case, the two responses essentially “share” the same set of random effect parameters (intercept and slope). When two parameters are completely dependent, the correlation between them is equal to one. This occurs when the covariance between the parameters is equal to the square root of the product of their respective variances. Most notation, however, define the model with a 2×1 vector of random effects, such as

$$\beta = \begin{bmatrix} a_j \\ b_j \end{bmatrix} \sim N(\mathbf{0}, \mathbf{G}), \text{ with } \mathbf{G} = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ba} & \sigma_b^2 \end{bmatrix}.$$

Clearly, the aforementioned structure imposes strong assumptions on the relationship between the two response variables. It is very unlikely that the two responses would exhibit complete dependence in the association between the random slopes and between the random intercepts. One advantage of this model, when the assumption is tenable, is that it drastically reduces the number of random effects that must be estimated when the number of response variables is large. For models with a large number of response variables, estimation would likely be impossible if the shared-parameters (or alternative approach) were not used.

2.5.2 Association of the Evolutions (AOE)

One important question that may be addressed with a joint mixed-effects model is how the evolution of one response is associated with the evolution of another response (“association of the evolutions”). By definition, the correlation between the evolutions for the two random slopes is given by

$$r_E = \frac{\text{Cov}(b_1, b_2)}{\sqrt{\text{Var}(b_1)}\sqrt{\text{Var}(b_2)}} = \frac{\sigma_{b_1, b_2}}{\sqrt{\sigma_{b_1}^2} \sqrt{\sigma_{b_2}^2}}. \quad (2.23).$$

It may be noted that the above expression is produced using those values from the **G** matrix defined in equation (2.22).

2.5.3 Evolution of the Association (EOA)

A similar idea that may be investigated using a joint mixed effects model is how the association between the responses evolves over time (“evolution of the association”). Assuming uncorrelated errors, the marginal correlation between the two responses as a function of time is given by

$$\begin{aligned} r_M(t) &= \frac{\text{Cov}(Y_{j1}(t), Y_{j2}(t))}{\sqrt{\text{Var}(Y_{j1}(t))}\sqrt{\text{Var}(Y_{j2}(t))}} \\ &= \frac{\sigma_{a_1, a_2} + t\sigma_{a_1, b_2} + t\sigma_{a_2, b_1} + t^2\sigma_{b_1, b_2}}{\sqrt{\sigma_{a_1}^2 + 2t^2\sigma_{a_1, b_1} + 2t^2\sigma_{b_1}^2 + \sigma_1^2} \sqrt{\sigma_{a_2}^2 + 2t^2\sigma_{a_2, b_2} + 2t^2\sigma_{b_2}^2 + \sigma_2^2}}. \end{aligned} \quad (2.24)$$

Assuming correlated errors, the marginal correlation between the two responses as a function of time is given by

$$r_M(t) = \frac{\text{Cov}(Y_{j1}(t), Y_{j2}(t))}{\sqrt{\text{Var}(Y_{j1}(t))} \sqrt{\text{Var}(Y_{j2}(t))}} \quad (2.25)$$

$$= \frac{\sigma_{a_1, a_2} + t\sigma_{a_1, b_2} + t\sigma_{a_2, b_1} + t^2\sigma_{b_1, b_2} + \sigma_{12}}{\sqrt{\sigma_{a_1}^2 + 2t^2\sigma_{a_1, b_1} + 2t^2\sigma_{b_1}^2 + \sigma_1^2} \sqrt{\sigma_{a_2}^2 + 2t^2\sigma_{a_2, b_2} + 2t^2\sigma_{b_2}^2 + \sigma_2^2}}$$

The delta method could be used to obtain 95% confidence bounds for $r_M(t)$ at any particular point in time.

Two observations can be made from equation (2.24). First, notice that when $t = 0$ the marginal correlation reduces to

$$r_M(t) = \frac{\sigma_{a_1, a_2}}{\sqrt{\sigma_{a_1}^2 + \sigma_1^2} \sqrt{\sigma_{a_2}^2 + \sigma_2^2}},$$

which is essentially the correlation between the two random intercepts. In fact, when the error components are small, the closer the marginal correlation at $t = 0$ approximates the correlation between the random intercepts. Also, as t increases $r_M(t)$ converges to r_E for the case with uncorrelated errors, and to

$$r_M(t) = \frac{\sigma_{a_1, a_2} + \sigma_{12}}{\sqrt{\sigma_{a_1}^2 + \sigma_1^2} \sqrt{\sigma_{a_2}^2 + \sigma_2^2}},$$

for the case of correlated errors, which indicates that the absolute value of the marginal correlation at $t = 0$ cannot be higher than the correlation between the random intercepts. It may also be noted that as t increases the marginal correlation converges to the correlation between the random slopes, while the variance-covariance parameters of the random effects determine the shape of the marginal correlation function (Fieuw, *et al.* 2004).

2.6. Summary

The univariate normal random-effects model was first defined and methods for estimation were discussed. The univariate definition was extended to a more general definition for the multivariate normal random-effects model. Univariate models can essentially be joined together with a specific definition of the variance-covariance structure for the random-effects.

Various structures of the **G** matrix are imposed to allow for different assumptions of the random-effects. Both a complete independence approach and a shared-parameters approach were considered. Finally, the methodology for determining the association of the evolutions and the evolution of the association were discussed. Software packages such as PROC MIXED and PROC GLIMMIX in SAS could be used to implement multivariate longitudinal data analysis. In the following chapter, an application of the multivariate mixed-model will be demonstrated using data from the Fels Longitudinal Study.

3 Data Analysis

3.1. Introduction

The methods for univariate and multivariate longitudinal analysis using the Fels Longitudinal Study data are presented in this chapter. In sections 3.2 and 3.3 the Fels study and the relevant data collected by the study are described. The methods used for data management are detailed in section 3.4, while the results from the study are discussed and conclusions drawn in section 3.5.

3.2. The Fels Longitudinal Study

Data from the Fels Longitudinal Study was chosen because it contains a large sample of participants with serial measurements of the factors defining the metabolic syndrome from the time of a participant's birth until their death. The Fels Longitudinal Study is an ongoing multidisciplinary serial study which began in 1929 as a means to investigate child growth and development (Roche, 1992). To date, a myriad of data has been collected on not just the original participants but also their children, grandchildren and great grandchildren. The study does routine data collection multiple times per year for participants in the study from birth until the age of twenty-one, then continues collecting information every two years until death. Only a small percentage of participants (approximately 8%) are lost to follow-up monitoring. The Fels Longitudinal Study collects serial information on such areas as anthropometry, dual energy x-ray absorptiometry, hydrodensitometry, residual volume, bioelectric impedance, total body

water, grip strength and blood pressure, skeletal maturation, health history, menstruation, smoking, alcohol, function, and physical activity. All procedures were approved by the Wright State University institutional review board, and all participants gave written consent in order to join the Fels Longitudinal Study. Recall, the specific measurements of interest related to the metabolic syndrome, are waist circumference, blood pressure, triglycerides, high density lipoprotein cholesterol (HDL-c), and serum glucose.

3.3. The Metabolic Syndrome Measurements

3.3.1 Blood Pressure

The measurements of a participant's diastolic and systolic blood pressure (mm Hg) were taken with a standard mercury sphygmomanometer every six months from age two through age eighteen, and every two years from then on. Three measurements were taken at each examination by a trained technician while participants were seated; the mean of the last two measurements was recorded (Roche *et al.*, 1992).

3.3.2 Waist Circumference

The measurement of an individual's waist circumference (cm) was taken as part of the anthropometric data collected by the Fels Longitudinal Study. The circumference of one's abdomen was taken using techniques similar to corresponding measurements in the *Anthropometric Standardization Reference Manual* (Lohman *et al.*, 1988). For the Fels

data set, the measuring of waist circumference changed during the course of the study. Specifically the study shifted from collecting waist circumference measurements made at the iliac crest to collecting measurements made at the umbilicus, which is a more modern method of waist circumference measurement. For the purposes of this research and the proposed analysis, the difference in measurement location is not expected to have a dramatic effect. For this analysis, if the iliac measurement for a participant is available it will be used as the waist circumference measurement in the analysis, while the umbilicus measurement will be used if the iliac is not available. All waist circumference measurements recorded for the Fels Longitudinal Study were the average of two measurements taken by independent observers. These measures were repeated if the two observers differed by more than a preset error limit.

3.3.3 Blood Measures

The measurements of a participant's serum glucose (mg/dL), triglyceride (mg/dL) and high density lipoprotein cholesterol (HDL-c) (mg/dL) levels were observed using fasting blood samples collected on participants via a venous blood draw. This information is not collected until a participant is approximately eight years of age at the time of their visit.

3.3.4 Additional Measures

In addition to the measurements described above and the age at each visit, weight and skeletal age at each visit and the sex of each participant were available. Weight was

measured to 0.1 kg using a SECA brand scale. Skeletal age was assessed using a hand radiograph of each participant through the age of eighteen. These radiographs were scored using various established skeletal maturation assessment procedures, given that skeletal maturation involves the changes in the bones as an individual becomes an adult. Since hand-wrist x-rays were taken on most participants in the Fels study through the age of eighteen, the “Fels method” was developed to determine skeletal ages using hand-wrist x-rays through dual energy a-ray absorptiometry. The Fels method is based on ninety-eight indicators and thirteen measurements, of which only 25-30% may be used for a given chronological age since there is a short range in which certain indicators actually provide useful information (Roche *et al.*, 1988). For the purposes of this analysis the relative bone age will be used as a proxy for biological maturity. Relative bone age is calculated by subtracting the chronological age from the skeletal bone age.

3.4. Data Management and Preparation

Several data management steps were taken to prepare the original data set for analysis. This included preparing childhood data and determining which participants developed the metabolic syndrome in adulthood. All data management was performed with SAS 9.2, see Appendix section 6.1 and section 6.2.

3.4.1 Childhood Subset

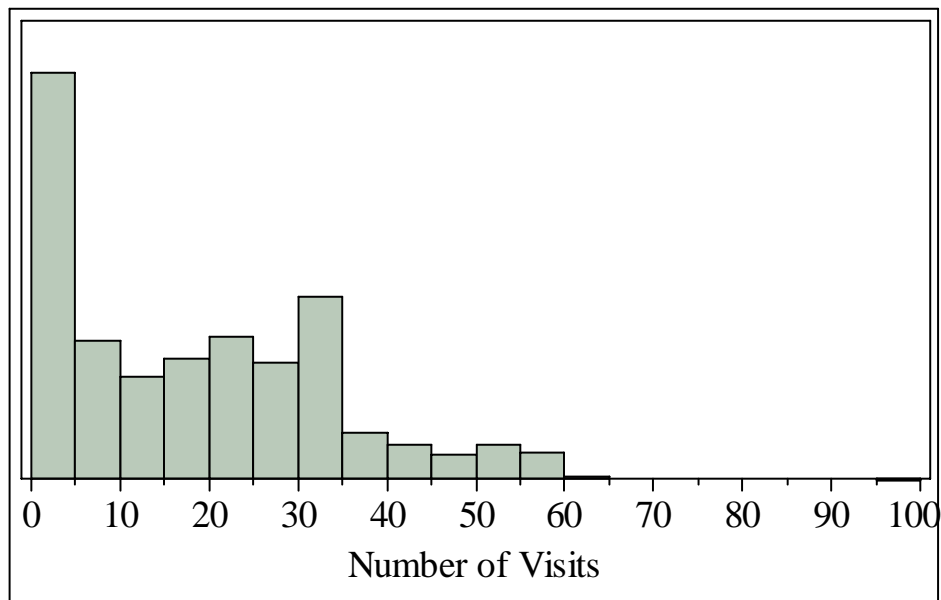
The outcome variables of interest for the analyses include the childhood (ages two through eighteen) measures of systolic and diastolic blood pressure. While the other factors of the metabolic syndrome (waist circumference, HDL-c, triglycerides, and glucose) were not analyzed as outcome variables, this information was retained from participant's childhood visits for descriptive purposes and future considerations. The independent childhood variables included sex, birth weight, and skeletal age. Skeletal age is a time dependent covariate, while sex and birth weight are independent of age at the time of visit. It should be noted that birth weight was not included in the original data set as a variable; rather, a participant's earliest weight measurement recorded (no more than five days after birth) was used.

In organizing the data a subset was created (CHILD) which included multiple rows per subject; one row per visit for ages two (inclusive) through nineteen (exclusive). The columns consisted of the outcome measures of interest (systolic and diastolic blood pressure, as well as waist circumference, triglycerides, HDL-c, and serum glucose), the chronological age at which each measurement was collected, the participant's skeletal age at each visit, birth weight, and sex, as well as a unique participant identification number.

The CHILD subset consisted of 1361 subjects with a total of 24745 visits. Each participant was seen an average of approximately 18.18 times (standard deviation = 15.24, median = 17). The fewest number of visits for a single participant during

childhood was one, while the highest number of childhood visits for a single participant was ninety-five. The distribution of the number of childhood visits for the 1361 participants is shown in Figure 1.

Figure 1: Distribution of the Number of Childhood Visits



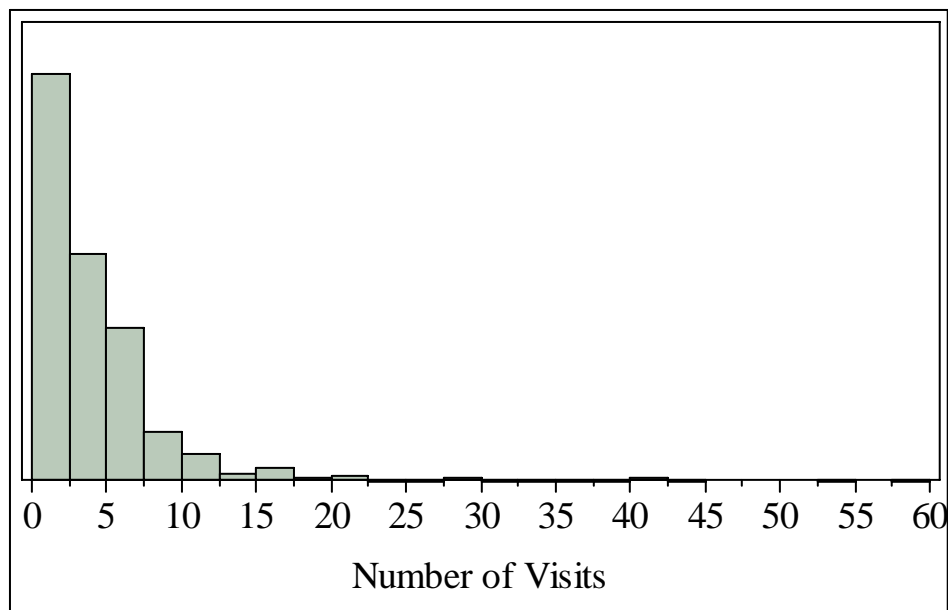
3.4.2 Adulthood Subset

A second subset (ADULT) was created which contained the adulthood measurements of participants, between the ages of thirty and fifty-five. This subset contained those risk factor measurements specific to the metabolic syndrome for participants in adulthood such that there were multiple rows per subject (one row per visit). Recall, the categorical cut-points for the risk factors are: a systolic blood pressure that is more than 130mm Hg

or a diastolic blood pressure that is more than 85mm Hg, a waist circumference that is greater than 102 cm for men and greater than 88 cm for women, a triglyceride count that is greater than 150 mg/dL, an HDL-c that is less than 40 mg/dL for men and less than 50 mg/dL for women, and a fasting serum glucose that is greater than 100 mm/dL.

There were a total of 5165 visits from 1194 subjects in the ADULT subset with 4.3 visits on average (standard deviation = 5.46, median = 3) between the ages of thirty and fifty-five. The minimum number of visits for a subject between the ages of thirty and fifty-five was one while the maximum number of visits was fifty-eight. The distribution of the number of adulthood visits for the 1194 participants is shown in Figure 2.

Figure 2: Distribution of the Number of Adulthood Visits



Given the measurements for each participant in adulthood, the number of risk factors each participant had at each adulthood visit was calculated. More specifically, it was determined whether a participant had no risk factors, exactly one, two, three, four or five risk factors at each visit in adulthood. In addition to the exact number of risk factors at each visit, it was determined whether a participant had at least one, two, three, four, or five risk factors at each visit in adulthood (for example, if a person had three risk factors, they would be flagged as having ‘at least one risk factor’, ‘at least two risk factors’, and ‘at least three risk factors’).

These risk factor variables for the 1194 subjects in the ADULT subset were merged with the 1361 subjects in the CHILD subset in order to create the dataset used for analysis (FINAL). A participant was included in the FINAL dataset if, and only if, they had measurements recorded in childhood (ages two through eighteen) and recorded in adulthood (ages thirty through fifty-five). The total number of subjects across both CHILD and ADULT subsets was 2035; however, 841 subjects were in the CHILD subset but not in the ADULT subset, while 674 subjects were in the ADULT subset but not in the CHILD subset. Thus, there were only 520 subjects in both the CHILD and ADULT subsets, meaning the FINAL subset to be used for analysis contained 520 subjects. All henceforth summaries are based on this sample of 520 individuals from the FINAL subset.

The distribution of the number of participants who were never diagnosed with any of the metabolic syndrome risk factors in adulthood and those who were diagnosed with

at least one, two, three, four and five risk factors (at some point during adulthood) are summarized in Table 2.

Table 2: Distribution of Adulthood Risks for the Metabolic Syndrome

Number of Risk Factors	Count	Total	Percent
None	67	466	14.38%
At Least One	399	466	85.62%
At Least Two	258	470	54.89%
At Least Three	153	464	32.97%
At Least Four	82	477	17.19%
At Least Five	21	510	4.12%

A participant was considered to have developed the metabolic syndrome in adulthood if they developed three or more risk factors, simultaneously, for at least one visit between the ages of thirty and fifty-five. The primary goal of this study is to determine if the 153 participants who developed the metabolic syndrome (at least three risk factors) differ from the sixty-seven participants who did not develop any of the metabolic syndrome risk factors in adulthood, with respect to systolic and diastolic blood pressure measures during childhood.

3.4.3 Further Considerations

Investigation of the response variables of interest during childhood yields a high degree of missing data of those factors related to blood measures (triglycerides, HDL-c, and glucose). Of the 14848 visits available from ages two through eighteen across the 520 participants, diastolic blood pressure was missing for 7802 visits (52.55%), systolic blood

pressure was missing for 7787 visits (52.44%), waist circumference was missing for 2870 visits (19.33%), triglycerides was missing for 14247 visits (95.95%), HDL-c was missing for 14250 visits (95.97%), and glucose was missing for 14756 visits (99.38%). It is unclear why the missing data rate is so high for all of the measures, especially the measures obtained from a venous blood sample. While the large degree of missing blood measures could be attributed to these measures not commencing until age eight, rates above 95% are still unusually high. Another reason for the missing data rates could be attributed to visits spanning over more than one day. For example, a participant could be seen over several days, with blood pressure recorded on day one, waist circumference on day two, and blood taken on day three. This would result in the participants having three visits, with data missing on each of these measures 67% of the time. Regardless of the reason for missing data, it is difficult to consider including the childhood blood measures (HDL-c, triglyceride, and glucose) for analysis in this study.

3.5. Results

3.5.1 Description of the Sample

There were 153 individuals (69.5%) in the Metabolic Syndrome (MetS) Group; that is, the group of adults who were diagnosed with three or more of the metabolic syndrome risk factors at least once between the ages of thirty and fifty-five. There were sixty-seven (30.5%) individuals in the No Risk Factor (No MetS) Group; that is, the group of adults

who were never diagnosed with any of the metabolic syndrome risk factors between the ages of thirty and fifty-five. For this sample of 220 individuals, 112 (50.9%) were male, while 108 (49.1%) were female. The mean birth weight was 3.29 kg (standard deviation = 0.58). There were nineteen participants missing birth weight information; six were in the No MetS Group and thirteen were in the MetS Group. The average age the participants in the MetS Group were first determined to have at least three of the metabolic syndrome risk factors in adulthood was 41.7 years (standard deviation = 7.58).

The following sections describe the results from the univariate and multivariate models separately for male and female participants. Male subjects are considered in section 3.5.2, with the univariate models being discussed first, followed by a multivariate model section. Female subjects are considered in section 3.5.3, with the univariate models being discussed first, followed by a multivariate model section. The chapter concludes with a comparison section (3.5.4) that discusses the different models for male and female participants.

3.5.2 Univariate and Multivariate Models for Male Subjects

3.5.2.1 Univariate Models

For the subset of 112 male participants a multivariate mixed-effects model was fit for the two response variables, diastolic blood pressure (DBP) and systolic blood pressure (SBP), assuming a complete independence variance-covariance structure as described in

section 2.5.1.1. This model produces the same results as two separate univariate mixed-effects models, but allows for a single likelihood for the model parameters enabling direct comparison with the correlated multivariate model fit in the subsequent section. This model was fit allowing for a separate quartic age effect for each metabolic syndrome group (MetS, No MetS), and adjusted for relative age and birth weight. Random intercepts and (linear first order) slopes were fit for each participant to account for within-subject correlations. The fixed-effects tests for the two response variables are summarized in Table 3.

Table 3: Fixed-Effects Tests for Males (Univariate)

Effect	DBP				SBP			
	NDF	DDF	F	<i>p</i> -value	NDF	DDF	F	<i>p</i> -value
Relative Age	1	803	0.35	0.5544	1	714	5.77	0.0165
Birth Weight	1	77.1	0.06	0.8127	1	87.4	1.11	0.2958
Age	1	1099	7.44	0.0065	1	1113	10.48	0.0012
METS Group	1	1044	0.66	0.4157	1	1055	5.09	0.0242
Age x MetS Group	1	1094	0.07	0.7908	1	1109	5.45	0.0197
Age²	1	1101	6.06	0.0140	1	1116	9.80	0.0018
Age² x MetS Group	1	1095	0.03	0.8627	1	1111	4.71	0.0303
Age³	1	1100	5.52	0.0190	1	1116	10.51	0.0012
Age³ x MetS Group	1	1094	0.01	0.9096	1	1111	4.10	0.0431
Age⁴	1	1099	4.83	0.0281	1	1116	10.12	0.0015
Age⁴ x MetS Group	1	1093	<0.01	0.9682	1	1109	3.61	0.0577

NDF, DDF = numerator, denominator degrees of freedom

Examining Table 3, there was evidence of a statistically significant quartic relationship between age and both DBP (p -value = 0.0281) and SBP (p -value = 0.0015) for males. However, this quartic relationship was not significantly different between the two MetS Groups (MetS and No MetS) for either DBP (p -value = 0.9682) or SBP (p -value = 0.0577). The differences in DBP and SBP between the metabolic syndrome groups at each age for males are summarized in Table 4.

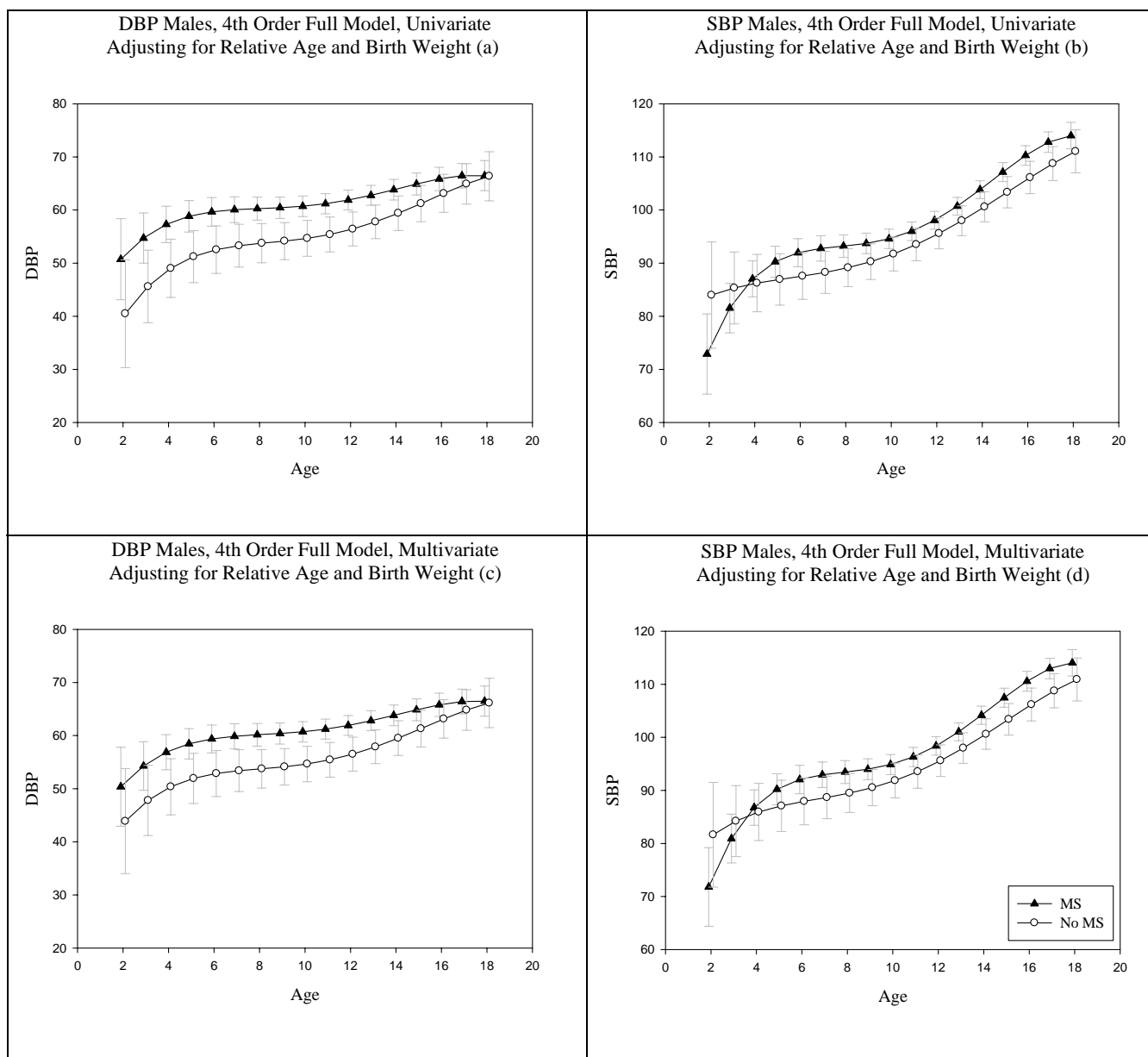
The univariate plots for DBP and SBP are summarized in Panels (a) and (b) of Figure 3. From these plots notice that both DBP and SBP increased with age, with greater increases associated with SBP than with DBP. In general, it also appears that those who developed at least three metabolic syndrome risk factors in adulthood had higher blood pressure measures during childhood than those who did not develop any metabolic syndrome risk factors in adulthood.

Table 4: Male's LS Means Differences (Univariate)

Age	DBP (MetS-No MetS)				SBP (MetS-No MetS)			
	Difference	SE	p-value	95% CI	Difference	SE	p-value	95% CI
2	10.250	6.466	0.1135	(-2.451, 22.952)	-11.010	6.384	0.0827	(-23.640, 1.441)
3	9.116	4.222	0.0320	(0.793, 17.438)	-3.791	4.171	0.3644	(-12.011, 4.430)
4	8.248	3.259	0.0127	(1.794, 14.702)	0.792	3.221	0.8062	(-5.583, 7.168)
5	7.598	2.887	0.0098	(1.875, 13.322)	3.322	2.852	0.2465	(-2.327, 8.971)
6	7.120	2.630	0.0079	(1.904, 12.335)	4.388	2.592	0.0933	(-0.748, 9.524)
7	6.767	2.373	0.0053	(2.058, 11.476)	4.495	2.328	0.0562	(-0.120, 9.110)
8	6.498	2.157	0.0033	(2.214, 10.782)	4.064	2.099	0.0556	(-0.100, 8.228)
9	6.270	2.024	0.0026	(2.250, 10.289)	3.433	1.947	0.0806	(-0.426, 7.293)
10	6.043	1.958	0.0026	(2.158, 9.929)	2.856	1.855	0.1265	(-0.820, 6.532)
11	5.780	1.914	0.0032	(1.984, 9.576)	2.502	1.780	0.1625	(-1.023, 6.027)
12	5.445	1.870	0.0045	(1.733, 9.157)	2.458	1.610	0.1510	(-0.909, 5.825)
13	5.002	1.855	0.0083	(1.319, 8.686)	2.725	1.642	0.1000	(-0.531, 5.981)
14	4.421	1.904	0.0225	(0.638, 8.203)	3.223	1.653	0.0539	(-0.055, 6.500)
15	3.669	2.010	0.0712	(-0.324, 7.661)	3.786	1.727	0.0304	(0.364, 7.208)
16	2.718	2.112	0.2017	(-1.481, 6.916)	4.165	1.796	0.0224	(0.603, 7.727)
17	1.540	2.228	0.4915	(-2.894, 5.974)	4.028	1.881	0.0349	(0.293, 7.761)
18	0.110	2.736	0.9680	(-5.302, 5.522)	2.955	2.406	0.2211	(-1.794, 7.704)

Using an $\alpha = .05$ decision rule, there are significant differences between the MetS Group and the No MetS Group at ages three through fourteen (p -values ≤ 0.0320) for DBP, and at ages fifteen through seventeen (p -values < 0.0350) for SBP. However, when using a more conservative Bonferroni adjustment for multiple comparisons of $\alpha = .05/17 = .0029$ the only statistically significant differences that remain were at ages nine and ten (both p -values = .0026) for DBP.

Figure 3: DBP and SBP Plots for Male Subjects (Univariate and Multivariate Models)



There was also evidence that relative age had a significant effect on SBP (p -value = 0.0165), but not on DBP (p -value = 0.5544). Amongst males, a one year increase in relative age (i.e. physical maturity) was significantly associated with a 0.931 increase (SE = 0.388; 95% CI = [0.170, 1.692]) in SBP. A one year increase in relative age was associated with a nominal decrease of 0.240 (SE = 0.406; 95% CI = [-1.036, 0.556]) in DBP. There was not evidence of a significant relationship between birth weight and either DBP (p -value = 0.8127) or SBP (p -value = 0.2958). For males, a one kg increase in birth weight was associated with a nominal decrease of 1.186 (SE = 1.128; 95% CI = [-3.428, 1.055]) in SBP, and a nominal decrease of 0.302 (SE = 1.268; 95% CI = [-2.827, 2.224]) in DBP.

Now that the fixed effects have been considered for the male univariate models, the random effects (intercepts and slopes) will be investigated. The estimated variance-covariance matrix, **G**, and the estimated **G** correlation matrix for both the DBP and the SBP response variables may be seen below in Table 5 and Table 6, respectively.

Table 5: Estimated **G** Variance-Covariance Matrix for Males (Univariate)

		DBP		SBP	
		Intercept	Slope	Intercept	Slope
DBP	Intercept	173.240	-10.441		
	Slope	-10.441	0.845		
SBP	Intercept			169.550	-9.962
	Slope			-9.962	0.722

Table 6: Estimated **G** Correlation Matrix for Males (Univariate)

		DBP		SBP	
		Intercept	Slope	Intercept	Slope
DBP	Intercept	1.000	-0.863		
	Slope	-0.863	1.000		
SBP	Intercept			1.000	-0.901
	Slope			-0.901	1.000

From Table 5, it may be seen that variability in the random intercepts and slopes is relatively similar for both DBP and SBP, though the variability for DBP is slightly higher. The same may be said of the covariance between the intercepts and slopes for SBP and DBP; while both are similar, DBP appears to be more extreme. Also, while not only being similar, the covariance's for both DBP and SBP are negative, which is indicative of a negative correlation, as seen in the **G** correlation matrix in Table 6. This negative correlation indicates that participants with lower SBP and DBP intercepts (blood pressure at “age = 0”) have higher increases in blood pressure over childhood (slopes), while participants with higher SBP and DBP intercepts have lower increases in blood pressure over time.

3.5.2.2 Multivariate Model

For the subset of 112 male participants, a multivariate mixed-effects model was fit with for the two response variables, DBP and SBP, assuming an unstructured variance-covariance structure as discussed in section 2.5. This model is the same as the

univariate model discussed in the previous section, except the sets of random intercepts and slopes for each response are now correlated rather than independent. This multivariate model was fit allowing for a separate quartic age effect for each metabolic syndrome group (MetS, No MetS), and adjusted for relative age and birth weight. Random intercepts and (linear first order) slopes were fit for each participant to account for within-subject correlations. The fixed-effects tests for the two response variables are summarized in Table 7.

The multivariate plots for DBP and SBP are summarized in Panels (c) and (d) of Figure 3. From these plots notice that both DBP and SBP increased with age, with greater increases associated with SBP than with DBP. In general, it also appears that those who developed at least three metabolic syndrome risk factors in adulthood had higher blood pressure measures during childhood than those who did not develop any metabolic syndrome risk factors in adulthood.

Table 7: Fixed-Effects Tests for Males (Multivariate)

Effect	Multivariate Males			
	NDF	DDF	F	p-value
Response	1	810	4.97	0.0261
Relative Age	1	1209	0.13	0.7144
Birth Weight	1	81.8	0.61	0.4381
Relative Age x Response	1	399	8.59	0.0036
Birth Weight x Response	1	76.5	0.82	0.3693
Age	1	2113	18.15	<.0001
METS Group	1	1749	1.21	0.2720
Age x MetS group	1	2107	2.06	0.1512
Age x Response	1	1319	0.96	0.3274
Response x MetS Group	1	989	3.29	0.0701
Age*Response x MetS Group	1	1303	2.38	0.1228
Age²	1	2128	16.03	<.0001
Age² x Response	1	1462	1.00	0.3178
Age² x MetS Group	1	2120	1.76	0.1851
Age² x Response x MetS Group	1	1441	2.07	0.1507
Age³	1	2131	16.06	<.0001
Age³ x Response	1	1579	1.19	0.2753
Age³ x MetS Group	1	2123	1.53	0.2161
Age³ x Response x MetS group	1	1557	1.82	0.1771
Age⁴	1	1721	14.29	0.0002
Age⁴ x Response	1	1665	2.63	0.1053
Age⁴ x MetS Group	1	1674	2.29	0.1302
Age⁴ x Response x MetS Group	1	1663	1.57	0.2111

NDF, DDF = numerator, denominator degrees of freedom

As with the two univariate models, the multivariate model showed evidence of a significant quartic relationship between age and both responses, DBP (p -value = 0.0425) and SBP (p -value = 0.0003). Furthermore, there was not evidence of a significant

difference between the two MetS Groups for either DBP (p -value = 0.9987) or SBP (p -value = 0.0835). The differences in DBP and SBP between the metabolic syndrome groups at each age for males are summarized in Table 8.

Using an $\alpha = .05$ decision rule, there were significant differences between the MetS Group and the No MetS Group at ages four through fourteen (p -values ≤ 0.0409) for DBP, and at ages fourteen through seventeen (p -values ≤ 0.0382) for SBP. However, when using a more conservative Bonferroni adjustment for multiple comparisons of $\alpha = .05/17 = 0.0029$, the only statistically significant differences that remain were at ages nine and ten (both p -values ≤ 0.0023) for DBP.

Table 8: Male's LS Means Differences (Multivariate)

AGE	DBP (MetS-No MetS)				SBP (MetS-No MetS)			
	Difference	SE	p-value	95% CI	Difference	SE	p-value	95% CI
2	6.477	6.299	0.3044	(-5.900, 18.854)	-9.830	6.281	0.1183	(-22.170, 2.513)
3	6.481	4.082	0.1140	(-1.570, 14.532)	-3.290	4.116	0.4257	(-11.400, 4.828)
4	6.497	3.140	0.0409	(0.275, 12.719)	0.833	3.199	0.7950	(-5.499, 7.164)
5	6.510	2.789	0.0216	(0.977, 12.043)	3.128	2.850	0.2748	(-2.519, 8.775)
6	6.508	2.553	0.0123	(1.443, 11.572)	4.126	2.607	0.1163	(-1.040, 9.292)
7	6.476	2.315	0.0062	(1.881, 11.070)	4.277	2.358	0.0725	(-0.398, 8.952)
8	6.401	2.115	0.0032	(2.200, 10.602)	3.955	2.141	0.0676	(-0.292, 8.201)
9	6.269	1.995	0.0022	(2.308, 10.231)	3.458	1.994	0.0859	(-0.497, 7.412)
10	6.068	1.938	0.0023	(2.222, 9.913)	3.011	1.903	0.1165	(-0.760, 6.782)
11	5.783	1.902	0.0030	(2.012, 9.554)	2.762	1.826	0.1331	(-0.855, 6.378)
12	5.401	1.866	0.0047	(1.698, 9.104)	2.782	1.743	0.1134	(-0.672, 6.236)
13	4.909	1.859	0.0097	(1.217, 8.602)	3.070	1.683	0.0711	(-0.269, 6.408)
14	4.294	1.917	0.0275	(0.487, 8.102)	3.546	1.689	0.0382	(0.197, 6.896)
15	3.543	2.027	0.0838	(-0.483, 7.569)	4.058	1.754	0.0226	(0.581, 7.534)
16	2.643	2.131	0.2182	(-1.592, 6.878)	4.375	1.813	0.0176	(0.779, 7.972)
17	1.580	2.246	0.4838	(-2.890, 6.051)	4.194	1.888	0.0287	(0.446, 7.943)
18	0.343	2.750	0.9010	(-5.098, 5.783)	3.134	2.403	0.1939	(-1.608, 7.876)

There was evidence that relative age had a significant effect on SBP (p -value = 0.0342), but not on DBP (p -value = 0.1401). Amongst males, a one year increase in relative age (i.e. physical maturity) was significantly associated with a 0.794 increase (SE = 0.374, 95% CI = [0.059, 1.528]) in SBP. A one year increase in relative age was associated with a nominal decrease of 0.573 (SE = 0.388; 95% CI = [-1.334, 0.189]) in DBP. There was not evidence of a significant relationship between birth weight and either DBP (p -value = 0.7677) or SBP (p -value = 0.2610). For males, a one kg increase in

birth weight was associated with a nominal decrease of 0.366 (SE = 1.235; 95% CI = [-2.825, 2.093]) in DBP, and a nominal decrease of 1.314 (SE = 1.161; 95% CI = [-3.623, 0.995]) in SBP.

Now that the fixed effects have been considered for the male multivariate model, the random effects (intercepts and slopes) will be investigated. For the male multivariate model the estimated variance-covariance matrix, **G**, and the estimated **G** correlation matrix for both the DBP and the SBP response variables may be seen in Table 9 and Table 10, respectively.

Table 9: Estimated **G** Variance-Covariance Matrix for Males (Multivariate)

		DBP		SBP	
		Intercept	Slope	Intercept	Slope
DBP	Intercept	162.010	-9.6903	136.180	-7.424
	Slope	-9.6903	0.800	-7.199	0.469
SBP	Intercept	136.18	-7.199	171.930	-9.718
	Slope	-7.4235	0.469	-9.718	0.687

Table 10: Estimated **G** Correlation Matrix for Males (Multivariate)

		DBP		SBP	
		Intercept	Slope	Intercept	Slope
DBP	Intercept	1.000	-0.851	0.816	-0.704
	Slope	-0.851	1.000	-0.614	0.633
SBP	Intercept	0.816	-0.614	1.000	-0.894
	Slope	-0.704	0.633	-0.894	1.000

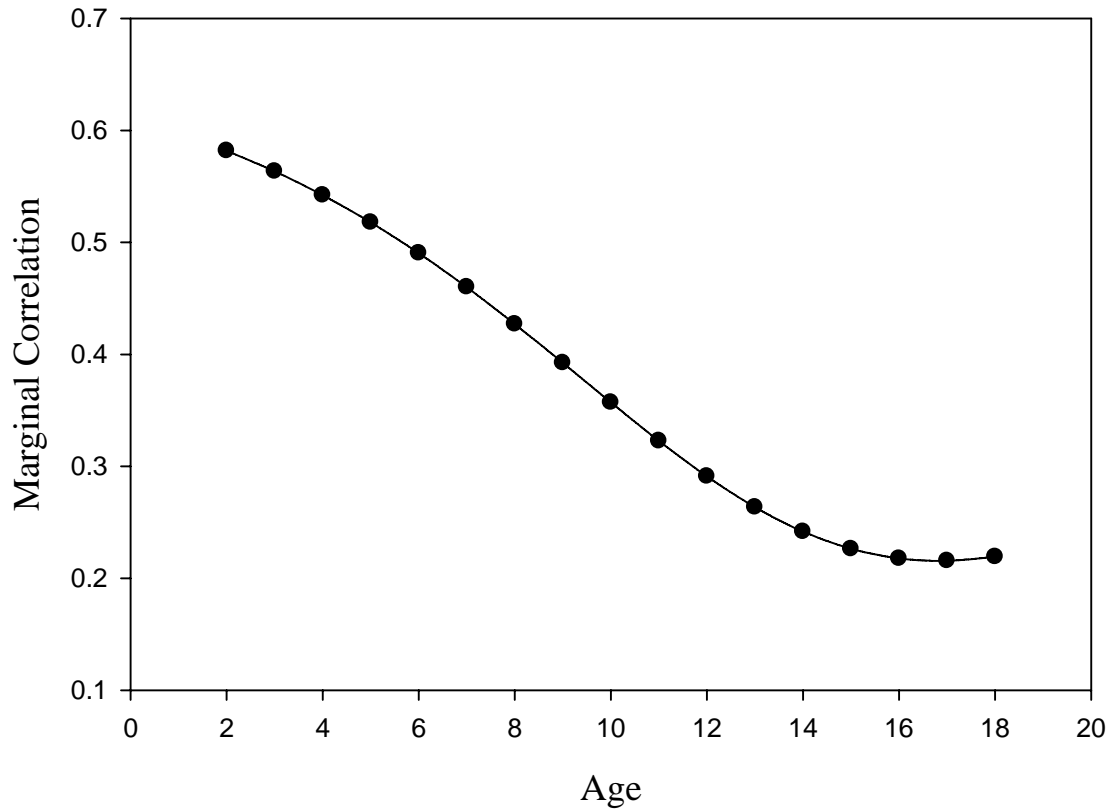
From the random effects, it may be seen that variability is relatively similar for both DBP and SBP, though the variability for DBP is slightly higher. The same may be said of the covariance for SBP and DBP, while both are similar, DBP appears to be more extreme. Also, while not only being similar, the covariance's for both DBP and SBP are negative, which is indicative of a negative correlation, which is seen in the **G** correlation matrix.

With the multivariate mixed-effects model is possible to investigate how the evolution of DBP is associated with the evolution of SBP, the association of the evolutions (AOE). It is also possible to determine how the association between DBP and SBP evolves over time, the evolution of the association (EOA).

The AOE can be determined by using equation (2.23) from section 2.5.2 or by reading the correlation between the two slopes directly from the estimated **G** correlation matrix (Table 10). Here the AOE between the random slope for DBP and the random slope for SBP is 0.633.

The EOA can be determined, and then visualized, using the marginal correlation between DBP and SBP, equation (2.25) from section 2.5.3. To visualize this, the implied correlation has been calculated and plotted over time using the marginal correlation between both frequencies in Figure 4. Notice that the association is strongest at age two at around 0.6, and this association decreases over time, leveling out at approximately 0.2 between fourteen and sixteen years of age.

Figure 4: Evolution of the Associations (Males)



3.5.3 Univariate and Multivariate Models for Female Subjects

3.5.3.1 Univariate Models

For the subset of 108 female participants a multivariate mixed-effects model was fit for the two response variables, diastolic blood pressure (DBP) and systolic blood pressure

(SBP), assuming a complete independence variance-covariance structure as described in section 2.5.1.1. This model produces the same results as two separate univariate mixed-effects models, but allows for a single likelihood for the model parameters enabling direct comparison with the correlated multivariate model fit in the subsequent section. This model was fit allowing for a separate quadratic age effect for each metabolic syndrome group (MetS, No MetS), and adjusted for relative age and birth weight. Random intercepts and (linear first order) slopes were fit for each participant to account for within-subject correlations. The fixed-effects tests for the two response variables are summarized in

Table 11.

Table 11: Fixed-Effects Tests for Females (Univariate)

Effect	DBP Females				SBP Females			
	NDF	DDF	F	<i>p</i> -value	NDF	DDF	F	<i>p</i> -value
Relative Age	1	809	1.60	0.2065	1	826	3.03	0.0821
Birth Weight	1	74	1.80	0.1843	1	77.7	0.30	0.5885
Age	1	713	13.89	0.0002	1	732	29.38	<.0001
MetS Group	1	322	0.26	0.6093	1	326	0.05	0.8151
Age x MetS Group	1	706	3.17	0.0757	1	725	0.19	0.6661
Age²	1	855	6.12	0.0136	1	881	11.02	0.0009
Age² x MetS Group	1	852	4.71	0.0303	1	877	0.19	0.6599

NDF, DDF = numerator, denominator degrees of freedom

The univariate plots for DBP and SBP are shown in Panels (a) and (b) of Figure 5. From these plots, notice that both DBP and SBP increase with age, with greater increases associated with SBP than with DBP. In general, it also appears that those who develop at least three metabolic syndrome risk factors in adulthood had higher blood pressure measures during childhood than those who did not develop any metabolic syndrome risk factors in adulthood.

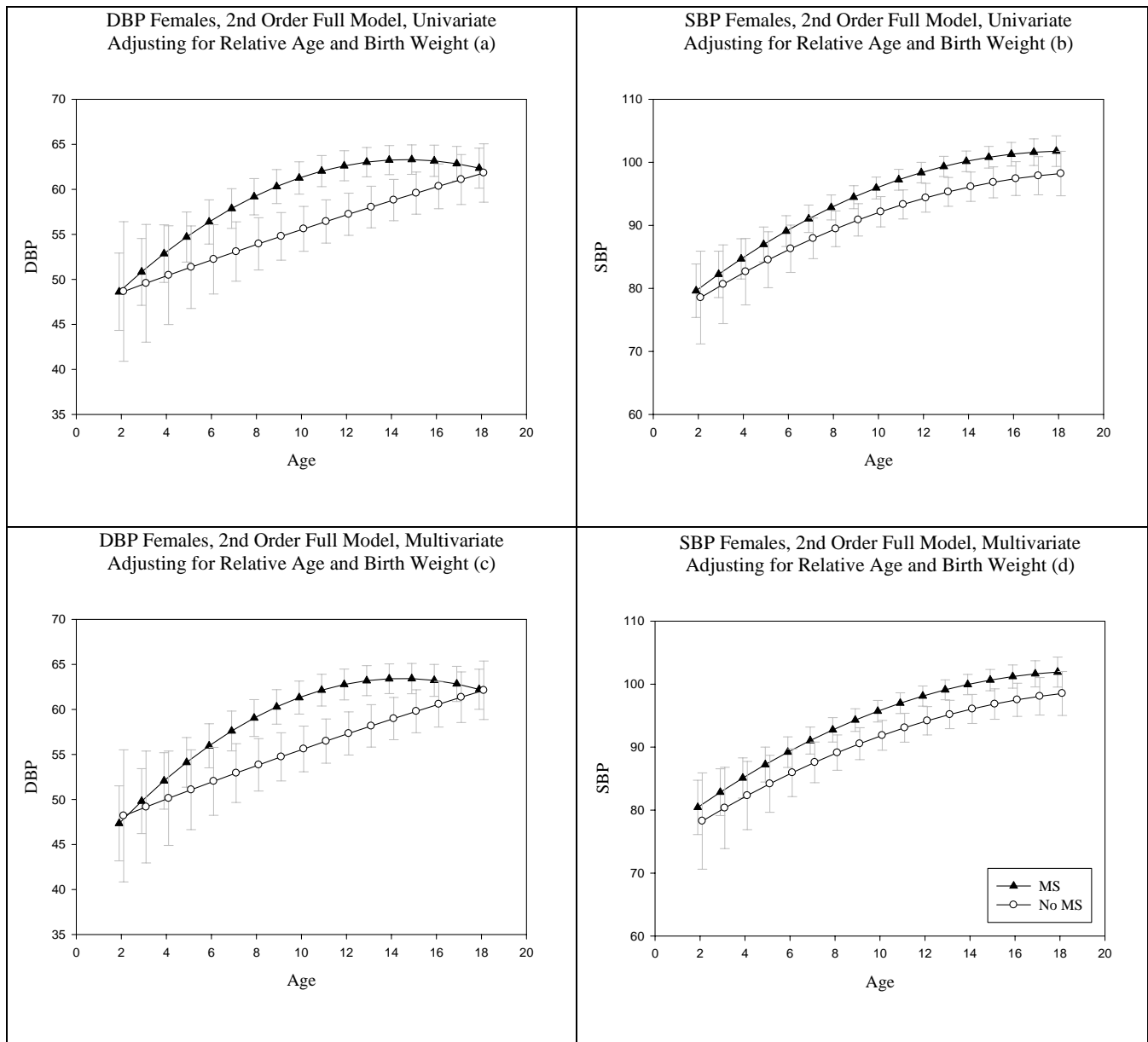
Statistically, there was evidence of a significant quadratic relationship between age and both DBP (p -value = 0.0136) and SBP (p -value = 0.0009) for females. This quadratic relationship was significantly different between the two MetS Groups (MetS and No MetS) for DBP (p -value = 0.0303) but not SBP (p -value = 0.6599). The differences in DBP and SBP between the metabolic syndrome groups at each age for females are summarized in Table 12.

Table 12: Female's LS Means Differences (Univariate)

Age	DBP (MetS-No MetS)				SBP (MetS-No MetS)			
	Difference	SE	<i>p</i> -value	95% CI	Difference	SE	<i>p</i> -value	95% CI
2	-0.040	4.501	0.9934	(-8.906, 8.831)	2.171	4.462	0.6271	(-6.620, 10.960)
3	1.263	3.812	0.7407	(-6.255, 8.782)	2.496	3.783	0.5102	(-4.960, 9.956)
4	2.395	3.215	0.4574	(-3.954, 8.744)	2.788	3.19	0.3835	(-3.510, 9.087)
5	3.358	2.712	0.2179	(-2.008, 8.723)	3.046	2.686	0.2589	(-2.270, 8.360)
6	4.152	2.306	0.0746	(-0.420, 8.724)	3.272	2.274	0.1531	(-1.240, 7.778)
7	4.778	1.998	0.0187	(0.812, 8.744)	3.464	1.954	0.0795	(-0.410, 7.342)
8	5.235	1.779	0.0041	(1.700, 8.770)	3.623	1.722	0.0382	(0.202, 7.044)
9	5.524	1.637	0.0011	(2.273, 8.775)	3.748	1.568	0.0189	(0.634, 6.862)
10	5.644	1.548	0.0004	(2.570, 8.717)	3.841	1.474	0.0106	(0.915, 6.766)
11	5.595	1.493	0.0003	(2.633, 8.557)	3.900	1.422	0.0072	(1.079, 6.721)
12	5.378	1.454	0.0004	(2.493, 8.264)	3.926	1.397	0.0060	(1.154, 6.698)
13	4.993	1.427	0.0007	(2.160, 7.825)	3.918	1.394	0.0060	(1.150, 6.687)
14	4.438	1.416	0.0024	(1.623, 7.253)	3.878	1.419	0.0076	(1.058, 6.698)
15	3.715	1.439	0.0118	(0.849, 6.581)	3.804	1.486	0.0124	(0.847, 6.761)
16	2.824	1.525	0.068	(-0.214, 5.862)	3.697	1.616	0.0248	(0.480, 6.914)
17	1.764	1.699	0.3023	(-1.616, 5.144)	3.557	1.828	0.0549	(-0.080, 7.191)
18	0.535	1.979	0.7875	(-3.391, 4.461)	3.384	2.132	0.1157	(-0.850, 7.613)

Using an $\alpha = .05$ decision rule, there were significant differences between the MetS Group and the No MetS Group at ages seven through fifteen (p -values ≤ 0.0187) for DBP, and at ages eight through sixteen (p -values ≤ 0.0382) for SBP. However, when using a more conservative Bonferroni adjustment for multiple comparisons of $\alpha = .05/17 = 0.0029$ the only statistically significant differences that remain were at ages nine through fourteen (p -values ≤ 0.0024) for DBP.

Figure 5: DBP and SBP Plots for Female Subjects (Univariate and Multivariate Models)



There was not evidence that relative age had a significant effect on DBP (p -value = 0.2065) or on SBP (p -value = 0.0821). For females, a one year increase in relative age was associated with a nominal increase of 0.462 (SE = 0.365; 95% CI = [-0.255, 1.179]) in DBP and a nominal increase of 0.623 (SE = 0.358; 95% CI = [-0.079, 1.324]) in SBP. There was not evidence that birth weight had a significant effect on DBP (p -value = 0.1843) or SBP (p -value = 0.5885). For females, a one kg increase in birth weight was associated with a nominal increase of 1.679 (SE = 1.253; 95% CI = [-0.818, 4.175]) in DBP and a nominal increase of 0.664 (SE = 1.222; 95% CI = [-1.77, 3.098]) in SBP.

Now that the fixed effects have been considered for the female univariate models, the random effects (intercepts and slopes) will be investigated. For the female univariate models, the estimated variance-covariance matrix, \mathbf{G} , and the estimated \mathbf{G} correlation matrix for both the DBP and the SBP response variables may be seen in Table 13 and Table 14, respectively.

Table 13: Estimated \mathbf{G} Variance-Covariance Matrix for Females (Univariate)

		DBP		SBP	
		Intercept	Slope	Intercept	Slope
DBP	Intercept	138.910	-7.605		
	Slope	-7.605	0.530		
SBP	Intercept			153.230	-9.626
	Slope			-9.626	0.747

Table 14: Estimated **G** Correlation Matrix for Females (Univariate)

		DBP		SBP	
		Intercept	Slope	Intercept	Slope
DBP	Intercept	1.000	-0.886		
	Slope	-0.886	1.000		
SBP	Intercept			1.000	-0.899
	Slope			-0.899	1.000

From Table 13, it may be seen that variability in the random intercepts and slopes is relatively similar for both DBP and SBP, though the variability for SBP is slightly higher (recall, this is the opposite of what was seen among the male participants). The same may be said of the covariance between the random intercepts and slopes for SBP and DBP, while both are similar, SBP appears to be slightly more extreme. Also, while not only being similar, the covariance's for both DBP and SBP are negative, which is indicative of a negative correlation, as seen in the **G** correlation matrix.

3.5.3.2 Multivariate Model

For the subset of 108 female participants, a multivariate mixed-effects model was fit with for the two response variables, DBP and SBP, assuming an unstructured variance-covariance structure as discussed in section 2.5. This model is the same as the univariate model discussed in the previous section, except the sets of random intercepts and slopes for each response are now correlated rather than independent. This multivariate model was fit allowing for a separate quadratic age effect for each metabolic syndrome group

(MetS, No MetS), and adjusted for relative age and birth weight. Random intercepts and (linear first order) slopes were fit for each participant to account for within-subject correlations. The fixed-effects tests for the two response variables are summarized in Table 15.

Table 15: Fixed-Effects Tests for Females (Multivariate)

Effect	NDF	DDF	F	p-value
Response	1	209	49.87	<.0001
Relative Age	1	1478	1.35	0.2448
Birth Weight	1	72	0.99	0.3235
Relative Age x Response	1	423	2.06	0.1515
Birth Weight x Response	1	92.4	0.49	0.4869
Age	1	1158	43.16	<.0001
MetS Group	1	336	0.20	0.6523
Age x MetS group	1	1147	3.19	0.0743
Age x Response	1	370	1.75	0.1863
Response x MetS Group	1	200	0.47	0.4915
Age*Response x MetS Group	1	362	1.51	0.2202
Age²	1	1638	19.73	<.0001
Age² x Response	1	506	0.48	0.4874
Age² x MetS Group	1	1630	4.36	0.0370
Age² x Response x MetS Group	1	496	2.48	0.1158

NDF, DDF = numerator, denominator degrees of freedom

The multivariate plots for DBP and SBP are shown in Panels (c) and (d) of Figure 5. From these plots, notice that both DBP and SBP increases with age, with greater increases associated with SBP than with DBP. In general, it also appears that those who

developed at least three metabolic syndrome risk factors in adulthood had higher blood pressure measures during childhood than those who did not develop any metabolic syndrome risk factors in adulthood.

As with the two univariate models, the multivariate model indicated evidence of a significant quadratic relationship between age and both responses, DBP (p -value = 0.0030) and SBP (p -value < 0.0001). Furthermore, there was evidence of a significant difference between the two MetS Groups for DBP (p -value = 0.0099), however, there was not evidence of a significant difference between the two MetS Groups for SBP (p -value = 0.5229). The differences in DBP and SBP between the metabolic syndrome groups at each age for females are summarized in Table 16.

Using an $\alpha = .05$ decision rule, there were significant differences between the MetS Group and the No MetS Group at ages seven through fifteen (p -values ≤ 0.0210) for DBP, and at ages nine through seventeen (p -values ≤ 0.0479) for SBP. However, when using a more conservative Bonferroni adjustment for multiple comparisons of $\alpha = .05/17 = 0.0029$ the only statistically significant differences that remain were at ages nine through thirteen (p -values ≤ 0.0012) for DBP.

Table 16: Female's LS Means Differences (Multivariate)

Age	DBP (MetS-No MetS)				SBP (MetS-No MetS)			
	Difference	SE	<i>p</i> -value	95% CI	Difference	SE	<i>p</i> -value	95% CI
2	-0.838	4.280	0.8450	(-9.274, 7.599)	1.080	4.313	0.8024	(-7.417, 9.577)
3	0.643	3.644	0.8602	(-6.548, 7.833)	1.578	3.675	0.6681	(-5.671, 8.827)
4	1.934	3.095	0.5331	(-4.182, 8.050)	2.030	3.120	0.5161	(-4.131, 8.192)
5	3.035	2.636	0.2518	(-2.182, 8.253)	2.437	2.648	0.3592	(-2.802, 7.675)
6	3.947	2.268	0.0847	(-0.550, 8.444)	2.797	2.262	0.2188	(-1.684, 7.278)
7	4.67	1.989	0.0210	(0.721, 8.619)	3.112	1.960	0.1155	(-0.777, 7.000)
8	5.203	1.792	0.0046	(1.644, 8.761)	3.381	1.738	0.0548	(-0.071, 6.832)
9	5.546	1.661	0.0012	(2.248, 8.844)	3.604	1.587	0.0254	(0.453, 6.754)
10	5.700	1.577	0.0005	(2.570, 8.830)	3.781	1.491	0.0128	(0.822, 6.740)
11	5.665	1.522	0.0003	(2.644, 8.685)	3.912	1.435	0.0076	(1.064, 6.760)
12	5.440	1.483	0.0004	(2.495, 8.384)	3.997	1.409	0.0055	(1.202, 6.793)
13	5.025	1.456	0.0009	(2.132, 7.918)	4.037	1.407	0.0051	(1.242, 6.832)
14	4.421	1.446	0.0030	(1.543, 7.299)	4.031	1.436	0.0063	(1.174, 6.888)
15	3.628	1.471	0.0160	(0.696, 6.559)	3.979	1.510	0.0102	(0.971, 6.987)
16	2.645	1.554	0.0932	(-0.454, 5.744)	3.881	1.646	0.0210	(0.601, 7.161)
17	1.472	1.720	0.3947	(-1.952, 4.897)	3.737	1.860	0.0479	(0.036, 7.439)
18	0.111	1.986	0.9557	(-3.830, 4.051)	3.548	2.162	0.1039	(-0.742, 7.838)

There was not evidence that relative age had a significant effect on either DBP (*p*-value = 0.944), or on SBP (*p*-value = 0.0708). Amongst females, a one year increase in relative age was associated with a nominal increase of 0.024 (SE = 0.344, 95% CI = [-0.651, 0.699]) in DBP and a nominal increase of 0.611 (SE = 0.338, 95% CI = [-0.052, 1.274]) in SBP. There was also not evidence of a significant relationship between birth weight and either DBP (*p*-value = 0.2442) or SBP (*p*-value = 0.5189). For females, a one kg increase in birth weight was associated with a nominal increase of 1.511 (SE = 1.287;

95% CI = [-1.054, 4.075]) in DBP, and a nominal increase of 0.8137 (SE = 1.256; 95% CI = [-1.687, 3.315]) in SBP.

Now that the fixed effects have been considered for the female Multivariate model, the random-effects (intercept and slopes) will be investigated. For the female multivariate model the estimated variance-covariance matrix, **G**, and the estimated **G** correlation matrix for both the DBP and the SBP response variables may be seen in Table 17 and Table 18.

Table 17: Estimated **G** Variance-Covariance Matrix for Females (Multivariate)

		DBP		SBP	
		Intercept	Slope	Intercept	Slope
DBP	Intercept	139.800	-7.453	142.500	-9.157
	Slope	-7.453	0.518	-7.777	0.603
SBP	Intercept	142.500	-7.777	165.830	-10.510
	Slope	-9.157	0.603	-10.510	0.817

Table 18: Estimated **G** Correlation Matrix for Females (Multivariate)

		DBP		SBP	
		Intercept	Slope	Intercept	Slope
DBP	Intercept	1.000	-0.876	0.936	-0.857
	Slope	-0.876	1.000	-0.839	0.928
SBP	Intercept	0.936	-0.839	1.000	-0.903
	Slope	-0.857	0.928	-0.903	1.000

From Table 16, it may be seen that variability is relatively similar for both DBP and SBP, though the variability for SBP is slightly higher. The same may be said of the

covariance for DBP and SBP, while both are similar, SBP appears to be slightly more extreme. Also, while not only being similar, the covariance's for both DBP and SBP are negative, which is indicative of a negative correlation, as seen in the **G** correlation matrix.

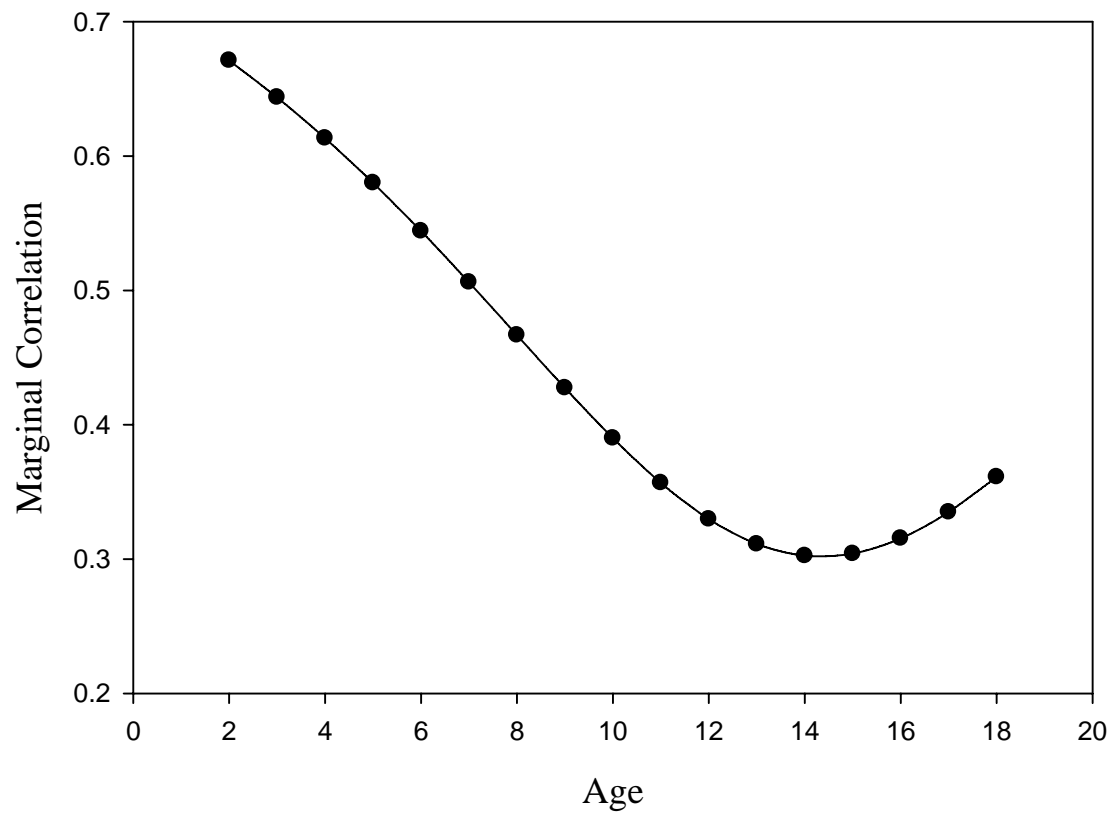
With the multivariate mixed-effects model is possible to investigate how the evolution of DBP is associated with the evolution of SBP, the association of the evolutions (AOE). It is also possible to determine how the association between DBP and SBP evolves over time, the evolution of the association (EOA).

The AOE can be determined by using equation (2.23) from section 2.5.2 or by reading the correlation between the two slopes directly from the estimated **G** correlation matrix (Table 18). Here the AOE between the random slope for DBP and the random slope for SBP is 0.928. The association in the evolutions of DBP and SBP is much higher than that seen with males (recall, for the male participants $AOE = 0.663$).

The EOA can be determined, and then visualized, using the marginal correlation between DBP and SBP, equation (2.25) from section 2.5.3. To visualize this, the implied correlation has been calculated and plotted over time using the marginal correlation between both frequencies in Figure 6. Notice, that at its strongest the correlation is just less than 0.7, at age two, and this association decreases over time through age fourteen where it bottoms out at approximately 0.3. At age fourteen the correlation begins to increase through age eighteen, nearing an association of approximately 0.4. For females,

this increase in the AOE in the later ages of childhood is far more extreme than that seen with males.

Figure 6: Associations of the Evolution (Females)



3.5.4 Comparisons

Now that both univariate and multivariate mixed-effects models have been considered for male and female participants, some time will be taken to compare both types of mixed-models for each sex. For the male participants, the estimated values for both the variance-covariance **G** matrix (Table 5) and the correlation **G** matrix (Table 6) have been combined for comparison purposes in Table 19, for both the univariate models and the multivariate model. When comparing the results from the independent setting to the results from the multivariate setting there are several points of interest. The -2 log-likelihood value corresponding to the two univariate models for males (fit as a joint model with appropriate covariance terMetS equal to zero) was equal to 17000.1. The -2 log-likelihood value for the multivariate model was 16962.4. A likelihood ratio test indicated that the multivariate model provided a significantly better fit than the two univariate models ($\chi^2 = 37.7$, $df=4$, p -value <0.0001). With regards to Akaike's information criterion (AIC), the multivariate model (AIC = 16962.4) is also indicated as a better fit than the univariate model (AIC = 17016.1) (since a smaller AIC value indicates a better fit). Notice how the multivariate model seems to decrease the variability in the random effects, this may be seen in Table 19, or by recalling Table 5 and Table 9. Taking into account the SE's for the variance and covariance estimates, the multivariate model in general allowed for more accurate prediction (smaller errors) of the variability in the random effects, though just slightly.

Table 19: REML Estimates for the Covariance and Correlation Parameters in the Univariate and Multivariate Models for the Male Participants

	Covariance Estimates		Correlation Estimates	
	Univariate	Multivariate	Univariate	Multivariate
-2 Res II	17000.1	16938.4	17000.1	16938.4
AIC	17016.1	16962.4	17016.1	16962.4
$\sigma_{a_1}^2$	173.240 (37.131)	162.010 (35.439)	1.000	1.000
$\sigma_{a_1a_2}$	-	136.180 (29.048)	-	0.816
$\sigma_{a_2}^2$	169.550 (35.441)	171.930 (34.975)	1.000	1.000
$\sigma_{a_1b_1}$	-10.441 (2.062)	-9.690 (2.496)	-0.086	-0.851
$\sigma_{a_2b_1}$	-	-7.199 (2.029)	-	-0.614
$\sigma_{b_1}^2$	0.845 (0.203)	0.800 (0.196)	1.000	1.000
$\sigma_{a_1b_2}$	-	-7.424 (1.932)	-	-0.704
$\sigma_{a_2b_2}$	-9.962 (2.404)	-9.718 (2.305)	-0.901	-0.894
$\sigma_{b_1b_2}$	-	0.468 (0.137)	-	0.633
$\sigma_{b_2}^2$	0.722 (0.176)	0.687 (0.165)	1.000	1.000
σ_1^2	56.322 (2.529)	56.509 (2.546)	56.322 (2.529)	56.509 (2.546)
σ_2^2	55.289 (2.460)	55.330 (2.457)	55.289 (2.460)	55.330 (2.457)

(-2 Res II = log-likelihood, AIC = Akaike's information criterion.)

Comparing the fixed effects for the univariate and multivariate mixed-models for the male subjects, some interesting things may be considered. First, and foremost, there is the question of whether the different models reached the same bottom-line conclusion.

Recall, that for the univariate model (Table 4), significant differences were found between the two MetS groups at ages three through fourteen, for DBP, and fifteen through seventeen, for SBP. Compare this to the multivariate model (Table 8), where significant differences were found between the two MetS groups at ages four through fourteen, for DBP, and fourteen through seventeen, for SBP. Essentially the same conclusion was drawn for both models. Moreover, both the univariate and multivariate models found that only ages nine and ten, for DBP, were significant when using a more conservative Bonferroni correction.

What is even more telling, when comparing the two models' LS means differences (i.e. comparing Table 4 and Table 8), was the precision of these differences. For each age, not only were the estimates for the differences slightly smaller for the multivariate model, the SE's were also slightly smaller, when compared to the univariate model. The SE's provided further evidence that, when applicable, fitting a joint multivariate mixed-effects model is superior to fitting multiple univariate models.

Comparing the covariates between the two types of models will yield further information of interest. Recall, from section 3.5.2, that both the univariate and multivariate models found a significant relationship between relative age and SBP. Both were positively associated with SBP (slopes of 0.794 compared to 0.931), however, the SE (0.374 compared to 0.388) was smaller for the multivariate model, hence the 95% CI was also tighter for the multivariate model. Both models also concluded a nominal

decrease with regards to relative age for DBP, and with regards to birth weight for both DBP and SBP, see Table 20.

Table 20: Covariate Comparison for Male Subjects

	DBP			SBP		
	Slope (SE)	95% CI	<i>p</i> -value	Slope (SE)	95% CI	<i>p</i> -value
Relative Age						
Univariate	-0.240 (0.406)	(-1.036, 0.556)	0.5544	0.931 (0.388)	(0.170, 1.692)	0.0165
Multivariate	-0.573 (0.388)	(-1.334, 0.189)	0.1401	0.794 (0.374)	(0.059, 1.528)	0.0342
Birth Weight						
Univariate	-0.302 (1.268)	(-2.827, 2.224)	0.8127	-1.186 (1.128)	(-3.428, 1.055)	0.2958
Multivariate	-0.366 (1.235)	(-2.825, 2.093)	0.7677	-1.314 (1.161)	(-3.623, 0.995)	0.2610

For the female participants, the estimated values for both the variance-covariance **G** matrix (Table 13) and the correlation **G** matrix (Table 10Table 14) have been combined for comparison purposes in Table 21, for both the univariate models and the multivariate models. When comparing the results from the independent setting to the results from the multivariate setting there are several points of interest. The -2 log-likelihood value corresponding to the two univariate models (fit as a joint model with appropriate covariance terMetS equal to zero) was equal to 14396.4. The -2 log-likelihood value for the multivariate model was 14297.4. A likelihood ratio test indicated that the multivariate model provided a significantly better fit than the two univariate models ($\chi^2 = 99$, $df=4$, p -value <0.0001). With regards to Akaike's information criterion

(AIC) the multivariate model (AIC = 16962.4) is also a better fit than the univariate model (AIC = 17016.1) (recall, a smaller value indicates a better fit). Notice how the multivariate model seems to decrease the variability in the random effects. Taking into account the SE's for the estimates of variance and covariance the multivariate model, in general, allows for more accurate prediction, though just slightly.

Table 21: REML Estimates for the Covariance and Correlation Parameters in the Univariate and Multivariate Models for the Female Participants

	Covariance Estimates		Correlation Estimates	
	Univariate	Multivariate	Univariate	Multivariate
-2 Res ll	14396.4	14297.4	14396.4	14297.4
AIC	14412.4	14321.4	14412.4	14321.4
$\sigma_{a_1}^2$	138.910 (31.920)	139.820 (31.367)	1.000	1.000
$\sigma_{a_1 a_2}$	-	142.500 (28.260)	-	0.936
$\sigma_{a_2}^2$	153.230 (33.032)	165.830 (34.286)	1.000	1.000
$\sigma_{a_1 b_1}$	-7.605 (2.124)	-7.453 (2.058)	-0.886	-0.876
$\sigma_{a_2 b_1}$	-	-7.776 (1.832)	-	-0.839
$\sigma_{b_1}^2$	0.530 (0.154)	0.518 (0.147)	1.000	1.000
$\sigma_{a_1 b_2}$	-	-9.157 (2.063)	-	-0.857
$\sigma_{a_2 b_2}$	-9.626 (2.348)	-10.513 (2.443)	-0.899	-0.903
$\sigma_{b_1 b_2}$	-	0.603 (0.139)	-	0.9279
$\sigma_{b_2}^2$	0.747 (0.180)	0.817 (0.188)	1.000	1.000
σ_1^2	47.956 (2.325)	47.645 (2.296)	47.956 (2.325)	47.645 (2.296)
σ_2^2	44.803 (2.161)	44.407 (2.124)	44.803 (2.161)	44.407 (2.124)

(-2 Res ll = log-likelihood, AIC = Akaike's information criterion.)

Comparing the fixed effects for the univariate and multivariate mixed-models for the female participants some interesting things may be considered. First, and foremost,

there is the question of whether the different models reached the same final conclusion. Recall, that for the univariate model (Table 12), significant differences were found between the two METS groups at ages seven through fifteen, for DBP, and eight through sixteen, for SBP. Compare this to the multivariate model (Table 16), where significant differences were found between the two METS groups at ages seven through fifteen, for DBP, and nine through seventeen, for SBP. Essentially the same conclusion was drawn for both models. Moreover, both the univariate and multivariate models concluded similarly that no significant differences were found for SBP when using a more conservative Bonferroni correction, while for DBP there were significant differences at ages nine through fourteen, and ages nine through thirteen, respectively.

When comparing the two models' LS means differences (i.e. comparing Table 12 and Table 16), what was even more telling was the precision of these differences. For each age, not only were the estimates for the differences slightly smaller for the multivariate model, the SE's were also slightly smaller, when compared to the univariate model. The SE's were further evidence that, when applicable, fitting a joint mixed-effects model was a better method.

Another item to touch upon was how the covariates compare between the two types of models. Recall, from section 3.5.3, that both the univariate and multivariate models did not find significant relationships between either covariate, relative age and birth weight, and either DBP or SBP. Both models drew the same conclusions with

regards to both covariates being related to a nominal increase in SBP and DBP, see Table 22, with the multivariate model, in general, having more precise estimates, as shown by smaller SE's and tighter CI's.

Table 22: Covariate Comparison for Female Subjects

	DBP			SBP		
	Slope (SE)	95% CI	<i>p</i> -value	Slope (SE)	95% CI	<i>p</i> -value
Relative Age						
Univariate	0.462 (0.365)	(-0.255, 1.179)	0.2065	0.623 (0.358)	(-0.079, 1.324)	0.0821
Multivariate	0.024 (0.344)	(-0.651, 0.699)	0.9440	0.611 (0.338)	(-0.052, 1.274)	0.0708
Birth Weight						
Univariate	1.679 (1.253)	(-0.818, 4.175)	0.1843	0.664 (1.222)	(-1.770, 3.098)	0.5885
Multivariate	1.511 (1.287)	(-1.054, 4.075)	0.2442	0.814 (1.256)	(-1.687, 3.315)	0.5189

4 Conclusions and Future Work

4.1. Overview of Work

In this thesis, two methods were considered for fitting two response variables measured longitudinally, a univariate mixed-effects model and a multivariate mixed-effects model. Jointly modeling multiple response variables, rather than modeling them independently incorporates additional information into the model that may increase estimation and prediction accuracy. An exploration into the potential gain by the multivariate model was motivated by a desire to more accurately identify when in childhood individuals start to become at risk for the metabolic syndrome, which they may develop later in adulthood. In Chapter 1, a discussion of the methodology available in the literature for modeling both univariate and multivariate longitudinal data was discussed in some detail. Also in Chapter 1, the metabolic syndrome was introduced and defined along with the specific aim of the analysis. In Chapter 2 the methodology, for both the univariate mixed-effects model and the multivariate mixed-effects model was discussed. Estimation of the fixed and random effects was described, along with formal definitions of the association in the evolution (AOE) of the two response and the evolution in the associations (EOA).

In Chapter 3 the results of the analysis of the Fels Longitudinal Study data was presented. Specifically, longitudinal trends in the measures of DBP and SBP during childhood were modeled using univariate and multivariate approaches, and compared

between those who went on to develop the metabolic syndrome in adulthood and those who did not. Here, not only were both models presented for male and female subjects, but the two models were compared (i.e. the models fit for the males subjects were compared against each other, and the models fit for the female subjects were compared against each other). An emphasis placed on what was gained by a multivariate mixed-model. Gains in the ability for the multivariate model to more accurately estimate both the fixed and random effects were described. In addition, the AOE and the EOA between DBP and SBP were estimated for both male and female participants. In fact, there was no evidence in the literature of estimates for the EOA and the AOE of DBP and SBP for children ages two through eighteen.

It is common practice for researchers to model several outcomes involved in a disease process independently. The results of this analysis indicated that there is additional information that can be gained with a multivariate model of the responses that are interrelated. It is comforting for researchers to know that the same questions addressed by the univariate model could be addressed with the multivariate model. However, the multivariate model is able to address these same questions with more accuracy (smaller standard errors) while also addressing additional questions that may be of great interest to the researcher, such as the AOE and the EOA of the responses. Furthermore, researchers are able to use the same software methods such as PROC MIXED in SAS for fitting multivariate longitudinal data as they currently use for fitting

univariate longitudinal data. Other reasons researchers may avoid jointly modeling response measures is that they may be of different types or collected at varying time points. For example, in HIV patients, one response may be a time-to-event (death) measure and another may be a continuous measure of CD4 counts. In fact, responses of different types are able to be jointly modeled under the framework of generalized mixed-effects models and fit using PROC GLIMMIX in SAS. Furthermore, the mixed-effects model is able to handle missing data relatively well.

4.2. Limitations

While there are several attractions to the multivariate model, there are several limitations that researchers should be aware of. For one, an issue of dimensionality is always present. Depending on the way the variance-covariance structure for the random-effects is defined, increasing the number of response variables can very quickly increase the number of parameters that the model must estimate. This issue is typically sidestepped with ease using modern computing methods for a multivariate model in which there are only two or three response variables. However, with increasing response variables, there is an exponential increase in the amount of computing power necessary to produce estimates. One such method when dealing with more than three response variables is a pair-wise approach discussed by Verbeke, *et al.* (2001). Other concerns when using a multivariate mixed-effects model include the assumptions made on the

correlation structure and the variance-covariance structure. For example, relaxing the conditional independence assumption by allowing correlated errors may cause a discrepancy due to the inappropriate modeling of the covariance structure. It is always important to understand what constraints and requirements are being made regarding the variance-covariance structure.

5 List of References

5.1.

List of References

- Agresti, A. (2002) *Categorical Data Analysis (2nd ed.)*. John Wiley & Sons: New York.
- Airy, G.B. (1861). *On the Algebraic and Numerical Theory of Errors of Observation and the Combination of Observations*. Macmillan: London.
- Anderson, T.W. and Goodman, L.A. (1957). Statistical Inference About Markov Chains. *Annals of Mathematical Statistics*, 28; 89-110.
- Apridonidze, T., Essah, P.A., Iuorno, M.J., and Nestler, J.E. (2004) Prevalence and Characteristics of the Metabolic Syndrome in Women with Polycystic Ovary Syndrome. *Journal of Clinical Endocrinology and Metabolism*, 20; 1929-1935.
- Ashford, J.R. and Sowden, R.R. (1970). Multivariate Probit Analysis. *Biometrics*, 26; 535-546.
- Billingsley, P. (1961). Statistical Methods in Markov Chains. *Annals of Mathematical Statistics*, 32; 12-40.
- Box, G.E.P. (1950). Problems in the Analysis of Growth and Wear Data. *Biometrics*, 6; 362-389.

- Breslow, N. E. and Clayton, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 88; 9-15.
- Brown, H. and Prescott, R. (1999). *Applied Mixed Models in Medicine*. John Wiley & Sons Ltd.
- Cox, D.R. (1958). The Regression Analysis of Binary Sequences (with discussion). *Journal of the Royal Statistical Society, Series B*, 20; 215-242.
- Cox, D.R. (1972). The Analysis of Mutivariate Binary Data. *Applied Statistics*, 21; 113-120.
- Cox, D.R. and Wermuth, N. (1992). Response Models for Mixed Binary and Quantitative Variables. *Biometrika*, 79; 441-461.
- Cox, D.R. and Wermuth, N. (1994). A Note on the Quadratic Exponential Binary Distribution. *Biometrika*, 81; 403-408.
- Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall: London.
- DeGruttola, V. and Tu, X.M. (1994). Modeling Progression of CD-4 Lymphocyte Count and its Relationship to Survival Time. *Biometrics*, 50; 1003-1014.
- Diggle, P.J., Heagerty, P.J., Liang, K.Y., and Zeger, S.L. (2002). *Analysis of Longitudinal Data* (2nd ed.). Oxford Science Publications: Oxford.
- Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modeling Based on Generalized Linear Models*. Springer-Verlag: Heidelberg.

- Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modeling Based on Generalized Linear Models (2nd ed)*. Springer Series in Statistics: New York.
- Faucett, C.J. and Thomas, D.C. (1996). Simultaneously Modeling Censored Survival Data and Repeatedly Measured Covariates: A Gibbs Sampling Approach. *Statistics in Medicine*, 15; 1663-1685.
- Fieuws, S. and Verbeke, G. (2004). Joint Modeling of Multivariate Longitudinal Profiles: Pitfalls of the Random-Effects Approach. *Statistics in Medicine*, 23; 3093-3104.
- Fieuws, S. and Verbeke, G. (2006). Pair-wise Fitting of Mixed Models for the Joint Modeling of Multivariate Longitudinal Profiles. *Biometrics*, 62; 424-431.
- Fieuws, S. and Verbeke, G. (2008). Predicting Renal Graft Failure Using Multivariate Longitudinal Profiles. *Biostatistics*, 9(3); 419-431.
- Fieuws, S., Verbeke, G. and Molenberghs, G. (2007). Random-Effects Models for Multivariate Repeated Measures. *Statistical Methods in Medical Research*, 16(5); 387-397.
- Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (2009). *Longitudinal Data Analysis*. Chapman & Hall/CRC Taylor Francis Group: Boca Raton.
- Galecki, A.T. (1994). General Class of Covariance Structures for Two or More Repeated Factors in Longitudinal Data Analysis. *Communications in Statistics-Theory and Methods*, 23; 3105-3119.

- Grundy, S. M., Cleeman, J. I., Daniels, S. R., Donato, K. A., Eckel, R. H., Franklin B.A., Gordon, D. J., Krauss, R. M., Savage, P. J., Smith, Jr., S. C., Spertus, J. A., and Costa, F. (2005). Diagnosis and Management of the Metabolic Syndrome: An American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. *Circulation*, 112; 2735-2752.
- Grundy, S.M. (2006). Metabolic Syndrome and Cardiovascular and Diabetes Worlds: ASCVD Risk in Metabolic Syndrome is Greater than the Sum of its Measured Risk Factors. *Journal of the American College of Cardiology*, 47(6); 1093-1100.
- Gueorguieva, R. (2001). A Multivariate Generalized Linear Mixed Model for Joint Modeling of Clustered Outcomes in the Exponential Family. *Statistical Modeling*, 1(3); 177-193.
- Gustat, J., Srinivasan, S.R., Elkasabany, A. and Berenson, G.S. (2002). Relation of Self-Related Measures of Physical Activity to Multiple Risk Factors of Insulin Resistance Syndrome in Young Adults: the Bogalusa Heart Study. *Journal of Clinical Epidemiology*, 55; 997-1006.
- Kaplan, N.M. (1989). The Deadly Quartet. Upper-Body Obesity, Glucose Intolerance, Hypertriglyceridemia, and Hypertension. *Archives of Intern Medicine*, 149; 1514-1520.

- Korn, E.L. and Whittemore, A.S. (1979). Methods of Analyzing Panel Studies of Acute Health Effects of Air Pollution. *Biometrics*, 35; 795-802.
- Krzankowski, W.J. (1988). *Principles of Multivariate Analysis*. Clarendon Press: Oxford.
- Laird, N. M. and Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*, 38; 963-974.
- Lemieux, I., Pascot, A., Couillard, C., Lamarche B., Tchernof A., Almeras, N., Bergeron, J., Gaudet, D., Tremblay, G., Prud'homme, D., Nadeau, A, and Despres, J.P. (2000). Hypertriglyceride Waist: A Marker of the Atherogenic Metabolic Triad (hyperinsulinemia; hyperapolipoprotein B; small, dense LDL) in Men? *Circulation*, 102; 179-184.
- Liang K.Y. and Zeger, S.L. (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, 73; 13-22.
- Little, R.J.A. and Schluchter, M.D. (1985). Maximum Likelihood Estimation for Continuous and Categorical Data With Missing Values. *Biometrika*, 72; 497-512.
- Lohman, T., Roche A., and Martorel, R., eds. (1988). *Anthropometric Standardization Reference Manual*. Human Kinetic Books.
- Molenberghs, G. and Verbeke, G. (2006). *Models for Discrete Longitudinal Data*. Springer Science+Business Media, LLC: New York.
- Nedler, J.A. and Wedderburn, R.W.M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society, Series A*, 135; 370-384.

- Olkin, I. and Tate, R.F. (1961). Multivariate Correlation Models With Discrete and Continuous Variables. *Annals of Mathematical Statistics*, 32; 448-465.
- Park, Y.W., Zhu, S., Palaniappan, L., Heshka, S., Carnethon, M.R., and HeyMsfield, S.B, (2003). The Metabolic Syndrome: Prevalence and Associated Risk Factor Findings in the US Population from the Third National Health and Nutrition Examination Survey, 1988-1994. *Archives of Internal Medicine*, 136; 427-436.
- Pierce, D.A. and Sands, B.R. (1975). Extra-Bernoulli Variation in Binary Data. *Technical Report 46, Department of Statistics, Oregon State University*.
- Rao, C.R. (1958). Some Statistical Methods for Comparison of Growth Curves. *Biometrics*, 14; 1-17.
- Reaven, G.M. (1993). Role of Insulin Resistance in Human Disease (Syndrome X). *Annual Review of Medicine*, 44; 121-131.
- Roche, A. F. (1992). *Growth, Maturation and Body Composition; The Fels Longitudinal Study 1929-1991*. Cambridge University Press.
- Roche, A. F., Chumlea W. C. and Thissen, D. (1988). *Assessing the Skeletal Maturity of the Hand-Wrist: Fels Method*. Charles C. Thomas.
- Roswell, J.J. and Walters, D.E. (1976) Analyzing Data With Repeated Observations on Each Experimental Unit. *Journal of Agricultural Science*, 87; 423-432.
- Stern, M.P. (1994). The Insulin Resistance Syndrome: the Controversy is Dead, Long Live the Controversy. *Diabetologia.*, 27; 956-958.

- Sun, S. S., Grave, G. D., Siervogel R. M., Pickoff, A. A., Arslanian, S. S., and Daniels, S. (2007). Systolic Blood Pressure in Childhood Predicts Hypertension and Metabolic Syndrome Later in Life. *Pediatrics*, 119; 237-246.
- Tate, R.F. (1954). Correlation Between a Discrete and Continuous Variable. *Annals of Mathematical Statistics*, 25; 603-607.
- Tsaitis, A.A., DeGruttola, V., and Wulfsohn, M.S. (1995). Modeling the Relationship of Survival to Longitudinal Data Measured With Error: Applications to survival and CD-4 counts in patients with AIDS. *Journal of the American Statistical Association*, 90; 27-37.
- Tsaitis, A.A. and Davidian, M. (2004). Joint Modeling of Longitudinal and Time-To-Event Data: An Overview. *Statistica Sinica*, 14; 809-834.
- Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer Science+Business Media, LLC: New York.
- Ware, J.H. and Liang K.Y. (1996). *The Design and Analysis of Longitudinal Studies: A Historical Perspective*. Wiley: New York.
- Ware, J.H., Lipsitz, S., and Speizer, F.E. (1988). Issues in the Analysis of Repeated Categorical Outcomes. *Statistics in Medicine*, 7; 95-107.
- Wishart, J. (1938). Growth-Rate, Determinations in Nutrition Studies with the Bacon Pig, and Their Analysis. *Biometrika*, 30; 16-28.

Zeger, S.L., Liang, K.Y., and Self, S.G. (1985). The Analysis of Binary Longitudinal Data With Time-Independent Covariates. *Biometrika*, 72; 31-38.

6 Appendices

6.1. Details for Data Processing

Creating the data sets (CHILD), (ADULT), and (FINAL) was done using SAS v. 9.2. The code associated with this process is shown below.

Code for Data Processing

```
/* READING IN DATA */
; /*****
*****
      Import Data and extract necessary variables
*****
*****
*/;
libname library "c:\AATHESIS\format";
libname LibForm "C:\Documents and Settings\jmthorpe\Desktop\THESIS
WORK\Fels";
options fmtsearch=(LibForm);

Data Libform.test; set Libform.Vcu0608;
run;

Data LibForm.Subset; set Libform.Vcu0608;
      keep      ptno          /*PATIENT ID NUMBER*/
      ANsafhand /*FELS HAND SKELETAL AGE*/
      AGE        /*PATIENT AGE*/
      SEX        /*PATIENT SEX*/
      ANbpd5     /*DIASTOLIC B.P -5th (mmhg)*/
      ANbpsys    /*SYSTOLIC B.P. (mmhg)*/
      BChdl      /*ALPHA LIPOPROTEIN (C-HDL)*/
      BCsglucose /*SERUM GLUCOSE (mg/dl)*/
      BCtrigly   /*TRIGLYCERIDES*/
      ANcrabdomili /*ABOMINAL-iliac (cm)*/
      ANcrabdomumb /*ADOMINAL-umbilicus (cm)*/
      ANz_stature /*Standarized Z-score for Height*/
      ANstature   /*Height (cm)*/
      ANweight;   /*Weight (kg)*/

run;
```

```

; /*****
*****
Creating ADULT data set from SUBSET, flag "high" measures
create METS variables for each observation
*****
*****
Data ADULTa;      set Libform.Subset;

/*select only adult obs between 30-75 years*/
if 30 <= age <= 55;
/* Set "High" Trigly */
if BCtrigly >= 150 then Htrigly = 1;
    if . < BCtrigly < 150 then Htrigly = 0;
/* Set "Low" hdl measure, MEN */
if SEX = 1 && . < BChdl < 40 then Lhdl = 1;
    if sex = 1 and BChdl >= 40 then Lhdl = 0;
/* Set "Low" hdl measure, WOMEN */
if SEX = 2 && . < BChdl < 50 then Lhdl = 1;
    if sex = 2 && BChdl >= 50 then Lhdl = 0;

/* Set "High" BPD measure */
if ANbpd5 >= 85 then Hbpd5 = 1;
    if . < ANbpd5 < 85 then Hbpd5 = 0;
/* Set "High" BPS measure */
if ANbpsys >= 130 then Hbpsys = 1;
    if . < ANbpsys < 130 then Hbpsys = 0;

/* Set single "High" BP measure */
if Hbpd5 = 1 or Hbpsys = 1 then Hbp = 1;
    if Hbpd5 = 0 && Hbpsys = 0 then Hbp = 0;
/* Set "High" glucose measure */
if BCs glucose >= 100 then Hglucose = 1;
    if . < BCs glucose < 100 then Hglucose = 0;
/* Set Waiste measurements to one variable */
if ANcrabdomili = . then wc = ANcrabdomumb;
    else wc = ANcrabdomili;
/* Set "High" waiste measure, MEN */
if SEX = 1 && wc >= 102 then Hwc = 1;
    if sex = 1 && . < wc < 102 then Hwc = 0;

/* Set "High" waiste measure, WOMEN */
if SEX = 2 && wc >= 88 then Hwc = 1;
    if sex = 2 && . < wc < 88 then Hwc = 0;

/*Two VERY IMPORTANT variables to create are number of RF missing*/
/*(nmiss) and the sum of the RFs (numRF) These will be used to*/
/*create all of the MS variables*/

```

```

if sum(Hbp,Lhdl,Hglucose,Htrigly,Hwc)>0
    then numRF=sum(Hbp,Lhdl,Hglucose,Htrigly,Hwc);
nmiss = nmiss(Hbp,Lhdl,Hglucose,Htrigly,Hwc);

/*Exactly no, one, two, three, four, or five RFs */
XnoRF=.; XoneRF=.; XtwoRF=.;
XthreeRF=.; XfourRF=.; XfiveRF=.;
/*At least one, two, three, four, or five RFs */
ALoneRF=.; ALtwoRF=.; ALthreeRF=.;
ALfourRF=.; ALfiveRF=.;

if nmiss = 4 then do;
    if numRF = . then do;    XfiveRF = 0;
                            ALfiveRF = 0; end;
    if numRF = 1 then do;   XnoRF = 0;
                            ALoneRF = 1; end;
    end;

if nmiss = 3 then do;
    if numRF = . then do;    XfiveRF = 0; XfourRF = 0;
                            ALfourRF = 0; ALfiveRF = 0; end;
    if numRF = 1 then do;   XnoRF = 0; XfiveRF = 0;
                            ALoneRF = 1; ALfiveRF = 0; end;
    if numRF = 2 then do;   XnoRF = 0; XoneRF = 0;
                            ALoneRF = 1; ALtwoRF = 1; end;
    end;

if nmiss = 2 then do;
    if numRF = . then do;    XfiveRF = 0; XfourRF = 0;
                            XthreeRF = 0; ALthreeRF = 0;
                            ALfourRF = 0; ALfiveRF = 0; end;
    if numRF = 1 then do;   XnoRF = 0; XfiveRF = 0;
                            XfourRF = 0; ALoneRF = 1;
                            ALfourRF = 0; ALfiveRF = 0; end;
    if numRF = 2 then do;   XnoRF = 0; XoneRF = 0;
                            XfiveRF = 0; ALoneRF = 1;
                            ALtwoRF = 1; ALfiveRF = 0; end;
    if numRF = 3 then do;   XnoRF = 0; XoneRF = 0;
                            XtwoRF = 0; ALoneRF = 1;
                            ALtwoRF = 1; ALthreeRF = 1; end;
    end;

if nmiss = 1 then do;
    if numRF = . then do;    XfiveRF = 0; XfourRF = 0;
                            XthreeRF = 0; XtwoRF = 0;
                            ALtwoRF = 0; ALthreeRF = 0;
                            ALfourRF = 0; ALfiveRF = 0; end;
    if numRF = 1 then do;   XnoRF = 0; XfiveRF = 0;
                            XfourRF = 0; XthreeRF = 0;
                            ALoneRF = 1; ALthree = 0;

```

```

ALfourRF = 0; ALfiveRF = 0; end;
if numRF = 2 then do; XnoRF = 0; XoneRF = 0;
XfiveRF = 0; XfourRF = 0;
ALoneRF = 1; ALtwoRF = 1;
ALfourRF = 0; ALfiveRF = 0; end;
if numRF = 3 then do; XnoRF = 0; XoneRF = 0;
XtwoRF = 0; XfiveRF = 0;
ALoneRF = 1; ALtwoRF = 1;
ALthreeRF = 1; ALfiveRF = 0; end;
if numRF = 4 then do; XnoRF = 0; XoneRF = 0;
XtwoRF = 0; XthreeRF = 0;
ALoneRF = 1; ALtwoRF = 1;
ALthreeRF = 1; ALfourRF = 1; end;
end;

if nmiss = 0 then do;
if numRF = . then do; XfiveRF = 0; XfourRF = 0;
XthreeRF = 0; XtwoRF = 0;
XoneRF = 0; XnoRF = 1;
ALoneRF = 0; ALtwoRF = 0;
ALthreeRF = 0; ALfourRF = 0;
ALfiveRF = 0; end;
if numRF = 1 then do; XnoRF = 0; XfiveRF = 0;
XfourRF = 0; XthreeRF = 0;
XtwoRF = 0; XoneRF = 1;
ALoneRF = 1; ALtwoRF = 0;
ALthreeRF = 0; ALfourRF = 0;
ALfiveRF = 0; end;
if numRF = 2 then do; XnoRF = 0; XoneRF = 0;
XfiveRF = 0; XfourRF = 0;
XthreeRF = 0; XtwoRF = 1;
ALoneRF = 1; ALtwoRF = 1;
ALthreeRF = 0; ALfourRF = 0;
ALfiveRF = 0; end;
if numRF = 3 then do; XnoRF = 0; XoneRF = 0;
XtwoRF = 0; XfiveRF = 0;
XfourRF = 0; XthreeRF = 1;
ALoneRF = 1; ALtwoRF = 1;
ALthreeRF = 1; ALfourRF = 0;
ALfiveRF = 0; end;
if numRF = 4 then do; XnoRF = 0; XoneRF = 0;
XtwoRF = 0; XthreeRF = 0;
XfiveRF = 0; XfourRF = 1;
ALoneRF = 1; ALtwoRF = 1;
ALthreeRF = 1; ALfourRF = 1;
ALfiveRF = 0; end;
if numRF = 5 then do; XnoRF = 0; XoneRF = 0;
XtwoRF = 0; XthreeRF = 0;
XfourRF = 0; XfiveRF = 1;

```



```

ALoneRF = 1; ALtwoRF = 1;
ALthreeRF = 1; ALfourRF = 1;
ALfiveRF = 1; end;

end;

keep ptno ANsafhand AGE SEX Hbp Lhdl Hglucose Htrigly
Hwc ANstature ANweight numRF nmiss XnoRF XoneRF
XtwoRF XthreeRF XfourRF XfiveRF ALoneRF ALtwoRF
ALthreeRF ALfourRF ALfiveRF;

proc freq data = Adulta;
tables nmiss numRF numRF*nmiss
XnoRF XoneRF XtwoRF XthreeRF XfourRF XfiveRF
ALoneRF ALtwoRF ALthreeRF ALfourRF ALfiveRF;
run;
/*****
*****
Creating MS variables for each subject
*****
*****/;
proc sort data = adulta; by ptno descending ALoneRF age; run;
data adult1; set adulta;
by ptno;
if first.ptno = 1;
if ALoneRF = 1 then AgeMS1diag = age;
keep ptno AgeMS1diag ALoneRF;
run;

proc sort data = adulta; by ptno descending ALtwoRF age; run;
data adult2; set adulta;
by ptno;
if first.ptno = 1;
if ALtwoRF = 1 then AgeMS2diag = age;
keep ptno AgeMS2diag ALtwoRF;
run;

proc sort data = adulta; by ptno descending ALthreeRF age; run;
data adult3; set adulta;
by ptno;
if first.ptno = 1;
if ALthreeRF = 1 then AgeMS3diag = age;
keep ptno AgeMS3diag ALthreeRF;
run;

proc sort data = adulta; by ptno descending ALfourRF age; run;
data adult4; set adulta;
by ptno;
if first.ptno = 1;
if ALfourRF = 1 then AgeMS4diag = age;

```

```

        keep ptno AgeMS4diag ALfourRF;
run;

proc sort data = adulta;      by ptno descending ALfiveRF age;      run;
data adult5;      set adulta;
    by ptno;
    if first.ptno = 1;
    if ALfiveRF = 1 then AgeMS5diag = age;
    keep ptno AgeMS5diag ALfiveRF;
run;

data adultb;
    merge adult1 adult2 adult3 adult4 adult5;
    by ptno;
run;

proc sort data = adultb; by ptno; run;
; /*****
*****
        Creating CHILD data set from SUBSET
*****
*****/;
Data Childa;      set Libform.Subset;
    keep ptno ANweight;
    if Age <.013;
    by ptno age;
    if first.ptno = 1;
    rename ANweight = birthWT;
run;
proc means data = Childa; run;

proc sort data = Childa;      by ptno; run;
Data Childb;      set Libform.Subset;
    if 2 <= age < 19;
    if ANcrabdomili = . then WC = ANcrabdomumb;
        else WC = ANcrabdomili;
    rename ANbpd5 = DBP;
    rename ANbpsys = SBP;
    rename ANsafhand = SKELETALage;
    rename BCHDL = HDL;
    rename BCsglucose = Glucose;
    rename BCtrigly = Trig;
    rename ANstature = ht;
    rename ANweight = wt;

```

```

if SEX=1 then SBPmu=102.19768+(1.82416*(AGE-10)**1)+(0.12776*(AGE-
10)**2)+(0.00249*(AGE-10)**3)-(0.00135*(AGE-
10)**4)+(2.73157*(ANz_stature)**1)-(0.19618*(ANz_stature)**2)-
(0.04659*(ANz_stature)**3)+(0.00947*(ANz_stature)**4);

if SEX=2 then SBPmu=102.01027+(1.94397*(AGE-10)**1)+(0.00598*(AGE-
10)**2)-(0.00789*(AGE-10)**3)-(0.00059*(AGE-10)**4)
+(2.03526*(ANz_stature)**1)+(0.02534*(ANz_stature)**2)-
(0.01884*(ANz_stature)**3)+(0.00121*(ANz_stature)**4);

if SEX=1 then DBPmu=61.01217+(0.68314*(AGE-10)**1)-(0.09835*(AGE-
10)**2)+(0.01711*(AGE-10)**3)-(0.00045*(AGE-10)**4)
+(1.46993*(ANz_stature)**1)-(0.07849*(ANz_stature)**2)-
(0.03144*(ANz_stature)**3)+(0.00967*(ANz_stature)**4);

if SEX=2 then DBPmu=60.50510+(1.01301*(AGE-10)**1)+(0.01157*(AGE-
10)**2)+(0.00424*(AGE-10)**3)-(0.00137*(AGE-10)**4)
+(1.16641*(ANz_stature)**1)+(0.12795*(ANz_stature)**2)-
(0.03869*(ANz_stature)**3)-(0.00079*(ANz_stature)**4);

ZSBP=(ANbpsys-SBPmu)/10.7128;
ZDBP=(ANbpd5-DBPmu)/10.7128;
PercentileSBP=probnorm(ZSBP)*100;
PercentileDBP=probnorm(ZDBP)*100;

drop ANcrabdomili ANcrabdomumb ANz_stature SBPmu DBPmu ZSBP ZDBP;

run;
proc sort data = Childb;      by ptno age;

data Childc;
merge Childb (in=a) Childa (in=b);
by ptno;
if a;
run;

proc means data = Childc n mean std median min max;
*where DBP > .;
*where SBP > .;
*where HDL > .;
*where Glucose > .;
*where Trig > .;
*where ANcrabdomili >.;
*where ANcrabdomumb >.;
*where WC >.;
*where SkeletalAge > .;
*var age SkeletalAge;
run;

```

```

; /*****
*****
Combine CHILD and ADULT data sets
*****
***** /;
proc sort data = Childc;      by ptno age;      run;
proc sort data = Adultb;     by ptno;      run;

data Libform.Childc;
    set Childc;
    by ptno;
    FlagSub = first.ptno;
run;

proc freq data = Childc;
    tables ptno;
run;

data Final;
    merge Childc (in=a) Adultb (in=b);
    by ptno;
    if a and b;
    FlagSub = first.ptno;
    if ALoneRF = 0 then MSgroup = "None";
    if ALthreeRF = 1 then MSgroup = "3+";
run;
data LibForm.Final;
    set Final;
    if DBP ne . && SBP ne . && WC ne . then eligible = 1;
run;
proc means data = Libform.Final n mean std median min max;
where FlagSub = 1;
var ptno sex birthWT;
run;

proc freq data = Libform.Final;
    tables Flagsub;
run;

/*There are 528 subjects (from freq above)*/
ods html file = 'C:\AATHESIS\excel.xls';
proc freq data = Libform.Final;
    where FlagSub = 1;
    tables MSgroup ALoneRF ALtwoRF ALthreeRF ALfourRF ALfiveRF SEX;
run;
ods html close;

```

```

/* Percentage of Male/Female in the None MSgroup */
proc freq data=Libform.Final;
    where MSgroup = "None" && FlagSub=1;
    tables sex;
run;

/* Percentage of Male/Female in the 3+ MSgroup */
proc freq data=Libform.Final;
    where MSgroup = "3+" && FlagSub=1;
    tables sex;
run;

/* Age at diagnosis */
proc means data=Libform.Final n mean std;
    where MSgroup = "3+" && FlagSub=1;
    var AgeMS3diag AgeMS4diag AgeMS5diag;
run;

proc print data = Final; run;
/*Freq above gives "prevelance" for each MS variable*/

/* Setting SBP and DBP to an integer measure */
data Libform.Final;
    set Libform.Final;
    SBP = int(SBP);
    DBP = int(DBP);
run;

data Libform.FinalStacked;
    set Final;
    value = SBP;      resp = "SBP"; output;
    value = DBP;      resp = "DBP"; output;
keep ptno Age Sex MSGroup birthwt Skeletalage value resp ht wt FlagSub;
run;

proc freq data=Libform.FinalStacked;
    where MSgroup = "3+" && FlagSub=1;
    tables sex;
run;
proc sort data = Libform.FinalStacked;
    by resp;
run;

```

6.2. Details for Mixed-Effects Modeling

Both the univariate and multivariate mixed-effects models were done using PROC

MIXED with using SAS v. 9.2. The code associated with this process is shown below.

Code for the Mixed-Effects Models

```
;/*****  
*****  
Univariate/Multivariate Mixed-Model Procedure for Male Subjects  
*****  
*****/;  
proc mixed data = Libform.FinalStacked2 covtest;  
  Where sex = 1 ;*Males;  
  class resp ptno MSGroup;  
  model value =      resp relage birthwt resp*relage resp*birthwt  
                    age MSGroup age*MSGroup resp*age  
                    resp*MSGroup resp*age*MSGroup  
                    age*age age*age*MSGroup age*age*resp  
                    age*age*resp*MSGroup age*age*age  
                    age*age*age*MSGroup age*age*age*resp  
                    age*age*age*resp*MSGroup age*age*age*age  
                    age*age*age*age*MSGroup  
                    age*age*age*age*resp  
                    age*age*age*age*resp*MSGroup  
  
                    /s ddfm=kr residual;  
  
  *G-Matrix;  
  ***Univariate***  
  *random int age/ sub=ptno type = un group = resp g gcorr;  
  
  ***Multivariate***;  
  random resp resp*age / sub=ptno type=un g gcorr;  
  
  *R-Matrix;  
  *Uncorrelated errors;  
  repeated /type=VC group=resp subject=ptno;  
  
  *Correlated errors  
  *repeated resp /type=un subject=ptno;  
  
  *DBP Fixed Effect Tests;  
  estimate "DBP: relage" relage 1 resp*relage 1 0;  
  estimate "DBP: birthwt" birthwt 1 resp*birthwt 1 0;  
  estimate "DBP: lin age" age 1 age*MSGroup 0.5 0.5  
                    age*resp 1 0  
                    age*resp*MSGroup 0.5 0.5 0 0;  
  estimate "DBP: MS vs noMS" MSGroup 1 -1  
                    resp*MSGroup 1 -1 0 0;  
  estimate "DBP: MS vs noMS lin age"  
                    age*MSGroup 1 -1
```

```

age*resp*MSgroup 1 -1 0 0;
estimate "DBP: quad age"      age*age 1 age*age*MSgroup 0.5 0.5
                               age*age*resp 1 0
                               age*age*resp*MSgroup 0.5 0.5 0 0;
estimate "DBP: MS vs noMS quad age" age*age*MSgroup 1 -1
                               age*age*resp*MSgroup 1 -1 0 0;
estimate "DBP: cub age"      age*age*age 1
                               age*age*age*MSgroup 0.5 0.5
                               age*age*age*resp 1 0
                               age*age*age*resp*MSgroup 0.5 0.5 0 0;
estimate "DBP: MS vs noMS cub age" age*age*age*MSgroup 1 -1
                               age*age*age*resp*MSgroup 1 -1 0 0;
estimate "DBP: quar age"     age*age*age*age 1
                               age*age*age*age*MSgroup 0.5 0.5
                               age*age*age*age*resp 1 0
                               age*age*age*age*resp*MSgroup 0.5 0.5 0 0
                               /singular=1;
estimate "DBP: MS vs noMS quar age" age*age*age*age*MSgroup 1 -1
                               age*age*age*age*resp*MSgroup 1 -1 0 0
                               /singular=1;

*SBP Fixed Effect Tests;
estimate "SBP: relage" relage 1 resp*relage 0 1;
estimate "SBP: birthwt" birthwt 1 resp*birthwt 0 1;
estimate "SBP: lin age" age 1 age*MSgroup 0.5 0.5
                               age*resp 0 1
                               age*resp*MSgroup 0 0 0.5 0.5;
estimate "SBP: MS vs noMS" MSgroup 1 -1
                               resp*MSgroup 0 0 1 -1;
estimate "SBP: MS vs noMS lin age" age*MSgroup 1 -1
                               age*resp*MSgroup 0 0 1 -1;
estimate "SBP: quad age"     age*age 1 age*age*MSgroup 0.5 0.5
                               age*age*resp 0 1
                               age*age*resp*MSgroup 0 0 0.5 0.5;
estimate "SBP: MS vs noMS quad age" age*age*MSgroup 1 -1
                               age*age*resp*MSgroup 0 0 1 -1;
estimate "SBP: cub age"     age*age*age 1
                               age*age*age*MSgroup 0.5 0.5
                               age*age*age*resp 0 1
                               age*age*age*resp*MSgroup 0 0 0.5 0.5;
estimate "SBP: MS vs noMS cub age" age*age*age*MSgroup 1 -1
                               age*age*age*resp*MSgroup 0 0 1 -1;
estimate "SBP: quar age"    age*age*age*age 1
                               age*age*age*age*MSgroup 0.5 0.5
                               age*age*age*age*resp 0 1
                               age*age*age*age*resp*MSgroup 0 0 0.5 0.5
                               /singular=1;
estimate "SBP: MS vs noMS quar age" age*age*age*age*MSgroup 1 -1
                               age*age*age*age*resp*MSgroup 0 0 1 -1

```

```

/singular=1;

*Differences in BP between MS group at each age;
lsmeans resp*MSgroup / diff cl at age=2 singular = 1;
lsmeans resp*MSgroup / diff cl at age=3 singular = 1;
lsmeans resp*MSgroup / diff cl at age=4 singular = 1;
lsmeans resp*MSgroup / diff cl at age=5 singular = 1;
lsmeans resp*MSgroup / diff cl at age=6 singular = 1;
lsmeans resp*MSgroup / diff cl at age=7 singular = 1;
lsmeans resp*MSgroup / diff cl at age=8 singular = 1;
lsmeans resp*MSgroup / diff cl at age=9 singular = 1;
lsmeans resp*MSgroup / diff cl at age=10 singular = 1;
lsmeans resp*MSgroup / diff cl at age=11 singular = 1;
lsmeans resp*MSgroup / diff cl at age=12 singular = 1;
lsmeans resp*MSgroup / diff cl at age=13 singular = 1;
lsmeans resp*MSgroup / diff cl at age=14 singular = 1;
lsmeans resp*MSgroup / diff cl at age=15 singular = 1;
lsmeans resp*MSgroup / diff cl at age=16 singular = 1;
lsmeans resp*MSgroup / diff cl at age=17 singular = 1;
lsmeans resp*MSgroup / diff cl at age=18 singular = 1;

*Comparisons between DBP and SBP (not described in thesis);
estimate "DBP vs SBP: relage" resp*relage 1 -1;
estimate "DBP vs SBP: birthwt" resp*birthwt 1 -1;

run;

; /*****
*****
Univariate Mixed Model Procedure for Female Subjects
*****
*****/;
proc mixed data = Libform.FinalStacked2 covtest;
  Where sex = 1 ; *Females;
  class resp ptno MSGroup;
  model value =      resp relage birthwt resp*relage
                    resp*birthwt
                    age MSGGroup age*MSGGroup resp*age
                    resp*MSGGroup resp*age*MSGGroup
                    age*age age*age*MSGGroup age*age*resp
                    age*age*resp*MSGGroup

                    /s ddfm=kr residual;

  *G-Matrix;
  ***Univariate***
  *random int age/ sub=ptno type = un group = resp g gcorr;

```



```

***Multivariate***;
random resp resp*age / sub=ptno type=un g gcorr;

*R-Matrix;
*Uncorrelated errors;
repeated /type=VC group=resp subject=ptno;

*Correlated errors
*repeated resp /type=un subject=ptno;

*DBP Fixed Effect Tests;
estimate "DBP: relage"          relage 1
                                resp*relage 1 0;
estimate "DBP: birthwt"        birthwt 1
                                resp*birthwt 1 0;
estimate "DBP: lin age"        age 1
                                age*MSgroup 0.5 0.5
                                age*resp 1 0
                                age*resp*MSgroup 0.5 0.5 0 0;
estimate "DBP: MS vs noMS"    MSgroup 1 -1
                                resp*MSgroup 1 -1 0 0;
estimate "DBP: MS vs noMS lin age"
                                age*MSgroup 1 -1
                                age*resp*MSgroup 1 -1 0 0;
estimate "DBP: quad age"      age*age 1
                                age*age*MSgroup 0.5 0.5
                                age*age*resp 1 0
                                age*age*resp*MSgroup 0.5 0.5 0 0;
estimate "DBP: MS vs noMS quad age"
                                age*age*MSgroup 1 -1
                                age*age*resp*MSgroup 1 -1 0 0;

*SBP Fixed Effect Tests;
estimate "SBP: relage"          relage 1
                                resp*relage 0 1;
estimate "SBP: birthwt"        birthwt 1
                                resp*birthwt 0 1;
estimate "SBP: lin age"        age 1
                                age*MSgroup 0.5 0.5
                                age*resp 0 1
                                age*resp*MSgroup 0 0 0.5 0.5;
estimate "SBP: MS vs noMS"    MSgroup 1 -1
                                resp*MSgroup 0 0 1 -1;
estimate "SBP: MS vs noMS lin age"
                                age*MSgroup 1 -1
                                age*resp*MSgroup 0 0 1 -1;
estimate "SBP: quad age"      age*age 1
                                age*age*MSgroup 0.5 0.5

```

```

                                age*age*resp 0 1
                                age*age*resp*MSgroup 0 0 0.5 0.5;
estimate "SBP: MS vs noMS quad age"
                                age*age*MSgroup 1 -1
                                age*age*resp*MSgroup 0 0 1 -1;

*Differences in BP between MS group at each age;
lsmeans resp*MSgroup / diff cl at age=2 singular = 1;
lsmeans resp*MSgroup / diff cl at age=3 singular = 1;
lsmeans resp*MSgroup / diff cl at age=4 singular = 1;
lsmeans resp*MSgroup / diff cl at age=5 singular = 1;
lsmeans resp*MSgroup / diff cl at age=6 singular = 1;
lsmeans resp*MSgroup / diff cl at age=7 singular = 1;
lsmeans resp*MSgroup / diff cl at age=8 singular = 1;
lsmeans resp*MSgroup / diff cl at age=9 singular = 1;
lsmeans resp*MSgroup / diff cl at age=10 singular = 1;
lsmeans resp*MSgroup / diff cl at age=11 singular = 1;
lsmeans resp*MSgroup / diff cl at age=12 singular = 1;
lsmeans resp*MSgroup / diff cl at age=13 singular = 1;
lsmeans resp*MSgroup / diff cl at age=14 singular = 1;
lsmeans resp*MSgroup / diff cl at age=15 singular = 1;
lsmeans resp*MSgroup / diff cl at age=16 singular = 1;
lsmeans resp*MSgroup / diff cl at age=17 singular = 1;
lsmeans resp*MSgroup / diff cl at age=18 singular = 1;

*Comparisons between DBP and SBP (not described in thesis);
estimate "DBP vs SBP: relage" resp*relage 1 -1;
estimate "DBP vs SBP: birthwt" resp*birthwt 1 -1;

run;

```