# A DOSE FINDING METHOD IN JOINT MODELING OF EFFICACY AND SAFETY ENDPOINTS IN PHASE II STUDIES -AN EXTENSION OF THE MCP-MOD METHOD 

BY AIYANG TAO
A dissertation submitted to the School of Public Health

University of Medicine and Dentistry of New Jersey and the

Graduate School-New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
UMDNJ-School of Public Health
Awarded jointly by these institutions and
Written under the direction of
Yong Lin, Ph.D.
and approved by

New Brunswick, New Jersey
May, 2010
(C) Copyright 2010

AIYANG TAO

ALL RIGHTS RESERVED

# ABSTRACT OF THE DISSERTATION A DOSE FINDING METHOD IN JOINT MODELING OF EFFICACY AND SAFETY ENDPOINTS IN PHASE II STUDIES -AN EXTENSION OF THE MCP-MOD METHOD 

BY AIYANG TAO<br>Dissertation Director: Yong Lin, Ph.D.

Determination of appropriate dose(s) to advance into Phase III is one of the most challenging and important decisions made during drug development. Selecting a dose too high may result in unacceptable safety problems, while a too low dose may lead to ineffective drugs. Proper estimation of such dose-response profiles for relevant safety and efficacy endpoints allows the reliable evaluation of the risk-benefit profile of a drug at the end of Phase II, as well as the selection of appropriate doses to be brought into confirmatory Phase III trials. This dissertation will address how to select dose(s) in Phase II trials by combining information about the efficacy and safety in a joint model setting. The methods we present in the dissertation may play a key role in drug development programs, and are often the gate-keeper for large confirmatory Phase III trials with greater chance of successful approval.

The dose selection when both safety and efficacy are represented by continuous responses is discussed in Part I of the dissertation, while Part II addresses the methodology when the safety and efficacy are mixed type responses. Both scenarios involve joint modeling of safety and efficacy endpoints. The methodology will focus on the following:
(1) Joint modeling approaches; (2) Model selection; (3) Identification of minimum effective dose (MED) and maximum safety dose(MSD); (4) Selection of optimal dose(s) for the Phase III program.

## Preface

This document consists of two parts:
I. Dose finding methodology for joint continuous bivariate responses (Chapter 1 to Chapter 6).
II. Dose finding methodology for joint continuous and discrete bivariate responses (Chapter 7 to Chapter 11).

Discussion and future directions are in Chapter 12. The dissertation is organized as follows:

- Part I:

Chapter 1 describes the background for the need of developing dose finding methodology, the motivating example for the dissertation, research objectives and specific aims for Part I.

Chapter 2 discusses the existing approaches for dose finding and focuses on the MCP-Mod method. This chapter also lists the limitations of the existing approaches.

Chapter 3 develops the concept and methodology of finding maximum safety dose (MSD) by extending the MCP-Mod to safety outcome.

Chapter 4 is devoted to the methodology of joint modeling of continuous bivariate responses; In the meantime, this chapter also presents two approaches of finding minimum effect dose (MED) and MSD for combined efficacy and safety data.

Chapter 5 proposes two strategies for recommending dose(s) to carry into Phase III program development either through the joint criteria of success or utility function.

## Conclusions of Part I are presented in Chapter 6.

- Part II:

Chapter 7 reviews the examples of clinical setting under mixed type responses, the need for developing the dose finding methodology for mixed type responses, research objectives and specific aims for Part II.

Chapter 8 presents existing methods for estimating multivariate responses, limitations as well as unaddressed areas for the dose finding for mixed type responses.

In chapter 9, we devote to a complete methodology for the joint model estimation, model fitting, strategies of finding the $M E D$ and $M S D$ for mixed type responses. This chapter also discusses the relationship between the observed discrete response and possible latent continuous variable in terms of mean model and correlations. In addition full likelihood for joint mixed type bivariate responses is also derived in this chapter.

Chapter 10 proposes two strategies for recommending dose(s) to carry into Phase III program development either through the joint criteria of success or utility function. The concept is similar as what we have developed for the continuous bivariate responses in Part I.

Conclusions of Part II are presented in Chapter 11.
Chapter 12 includes the discussion as well as the potential future research directions.

Appendix includes derivations to illustrate that the two likelihood formulas we derived in Chapter 9 have different distributions. The method for the proof of concept study for discrete response is also derived in Appendix.

## Acknowledgments

I would like to express my sincere appreciation to my advisor, Dr. Yong Lin, for guiding and helping me throughout the research process. He is always available and is so helpful whenever I face challenges. He has taught me a great deal and has been a wonderful mentor. This thesis would not have been possible without Dr. Lin's tremendous help with the key instructions and critical thoughts in theoretical difficulties as well as the strategies for resolving the problems.

Deep appreciation also goes to Dr. Jose Pinheiro who offered and brought me into this exciting dose-finding field. I also want to thank him for providing new perspectives and insightful discussion. Throughout the whole research period I always received valuable insights from him in working towards meeting pharmaceutical needs.

I would also like to particularly express my gratitude to Dr. Weichung Joe Shih, who provided me critical ideas and thought-provoking discussions. I appreciated Dr. Shih's valuable comments and constructive suggestions, which widely improved the validity and clarity of the thesis. He also taught me how to shape my technical writing. His insights challenged me to think broadly and critically about my research.

Thanks also go to Dr. Pamela Ohman-Strickland who provided me valuable comments and feedback based on her experience. Furthermore I appreciate Dr. OhmanStrickland's guidance during the first couple of years when I entered the Ph.D. program.

In the end I would like to thank my family: My parents who took care of all the housework for me as always and supported me with their love; My husband who gave me a sense of humor so I would not feel the pressure during these years; My kids were helping
me in their way, staying strong and healthy so I could focus on my study and research for all these years.

## Contents

Abstract ..... ii
Preface ..... iv
Acknowledgements ..... vi
List of Figures ..... xii
List of Tables ..... xiv
I Dose finding for Joint Continuous Bivariate Responses ..... 1
1 Introduction ..... 2
1.1 Background ..... 2
1.2 Motivating example ..... 4
1.3 Research objectives ..... 6
1.3.1 Objectives ..... 6
1.3.2 Specific aims ..... 6
2 Literature Review ..... 9
2.1 Existing approaches for dose finding ..... 9
2.2 MCP-Mod ..... 11
2.2.1 Multiple comparison procedures (MCP) ..... 11
2.2.2 Modeling approaches ..... 12
2.2.3 MCP-Mod approach procedures ..... 15
2.2.4 Bootstrap confidence interval approaches ..... 34
2.2.5 Extension of the current MCP-Mod approach ..... 36
2.3 Dose-finding based on efficacy-toxicity response ..... 37
2.3.1 Dose-finding based on efficacy-toxicity trade off by in Phase I cancer studies ..... 37
2.3.2 Limitations of this trade-off methodology ..... 38
3 Extend MCP-Mod to Safety Outcome ..... 39
3.1 Definition ..... 39
3.2 Simulation ..... 42
3.3 MSD dose selection performance ..... 42
3.3.1 MSD estimation plot ..... 42
3.3.2 Bias and dispersion ..... 44
3.3.3 Coverage of bootstrap confidence interval for MSD ..... 45
4 Joint Nonlinear Continuous Mixed Model Estimation ..... 47
4.1 Joint model formulation and log-likelihood function ..... 47
4.2 Estimation and computational methods ..... 50
4.2.1 Fitting joint model with gnls in R ..... 51
4.2.2 Parameter estimates, bias and relative efficiency ..... 52
4.2.3 Simulation ..... 52
4.2.4 Simulation results for parameter estimates ..... 54
4.2.5 Procedures/Strategies of finding MED and MSD for combined effi- cacy and safety data ..... 60
4.2.6 Strategy for separate and joint model fitting ..... 61
4.2.7 Simulation results for $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ from separate and joint model fitting ..... 65
5 Suggest Dose(s) for the Phase III Program Development for Joint Bivariate Con- tinuous Responses ..... 78
5.1 Method I: Identify the dose(s) through the joint criteria of continuous effi- cacy and safety responses for Phase III program ..... 79
5.1.1 Method ..... 79
5.1.2 Simulation ..... 79
5.2 Method II: Utility function based on trade off of continuous efficacy and safety responses for Phase III. ..... 84
5.2.1 Methods ..... 84
5.2.2 Simulation ..... 85
6 Conclusions ..... 90
II Dose Finding for Joint Continuous and Discrete Bivariate Responses ..... 93
7 Introduction to Joint Continuous and Discrete Bivariate Responses ..... 94
7.1 Background ..... 94
7.2 Research objectives ..... 96
7.2.1 Objectives ..... 96
7.2.2 Specific aims ..... 96
8 Literature Review for Estimation in Multivariate Responses ..... 99
8.1 Existing estimating approaches for multivariate responses ..... 100
8.1.1 GEE2 method ..... 102
8.1.2 Iteratively reweighted generalized least squares (IRGLS), extended least squares (ELS) and quasi-extended least squares (QELS) methods ..... 104
8.2 Unaddressed areas from the existing approaches for multivariate responses ..... 109
9 Developing the Methodology to Estimate MED and MSD for Joint Continuous and Discrete Responses ..... 110
9.1 Joint nonlinear bivariate continuous and discrete model ..... 110
9.2 Likelihood for nonlinear joint continuous and discrete bivariate model ..... 112
9.2.1 Method 1: The latent variable threshold $c$ and standard deviation $\sigma_{S}$ of latent variable are known ..... 113
9.2.2 Method 2: The latent variable threshold c is unknown ..... 121
9.3 Estimation of joint nonlinear bivariate continuous and discrete model ..... 125
9.3.1 Estimation by IRGLS, ELS and QELS methods ..... 125
9.3.2 Fitting the joint nonlinear bivariate continuous and discrete model with ELS or QELS approach ..... 134
9.4 Evaluation of the nonlinear joint model fitting ..... 135
9.4.1 Parameter estimate bias and relative efficiency ..... 135
9.4.2 Evaluate the model fitting ..... 136
9.4.3 Model mis-specification ..... 137
9.5 Simulation of correlated bivariate continuous and discrete responses ..... 138
9.5.1 Simulation Procedures ..... 138
9.5.2 Evaluation of Simulation Results ..... 140
9.6 Fitting the joint nonlinear continuous and discrete responses model ..... 142
9.6.1 Comparison of parameter estimates and their standard errors by the ELS and QELS approaches ..... 143
9.6.2 The parameter estimates results from joint fitting and separate fitting ..... 148
9.6.3 The model fitting from the model mis-specification ..... 176
9.7 How to determine MED and MSD ..... 178
9.7.1 MED definition ..... 178
9.7.2 MSD definition ..... 178
9.8 Procedures/Strategies of finding $M E D$ and $M S D$ for bivariate continuous efficacy and discrete safety data ..... 179
9.9 Design for separate and joint model fitting ..... 180
9.9.1 Separate model fitting ..... 180
9.9.2 Joint model fitting ..... 180
10 Suggest Dose(s) for the Phase III Program Development Based Joint Continuous and Discrete Responses ..... 183
10.1 Identify dose(s) for Phase III program through joint criteria of continuous efficacy and discrete safety responses ..... 184
10.1.1 Methods ..... 184
10.2 Identify dose(s) for Phase III through utility function based on trade off of efficacy and safety responses ..... 186
10.2.1 Methods ..... 186
11 Conclusions ..... 187
12 Discussion and Future Directions ..... 190
12.1 Discussion ..... 190
12.2 Future Directions ..... 193
12.2.1 Joint nonlinear continuous and discrete model estimation using full likelihood approach ..... 194
12.2.2 Extension of dose finding based on bivariate discrete variable ..... 195
12.2.3 Comparison of estimating methods in Part I and Part II ..... 195
12.2.4 Quasi-likelihood criterion (QIC) for model selections ..... 196
13 Appendix ..... 197
13.1 Comparison of Likelihoods Between Method 1 and Method 2 ..... 197
13.2 Proof of Concept Study for Binary Outcomes ..... 199
13.2.1 Contrast Tests with Multiple Dose Response Models ..... 199
13.2.2 Finding Optimal Model Contrasts ..... 205
Bibliography ..... 207
Vita ..... 212

## List of Figures

2.1 Multiple comparisons and Modeling approach to find MED ..... 14
2.2 MCP-Mod flow chart ..... 16
2.3 Dose-response models ..... 20
2.4 Plot of optimal contrasts ..... 26
2.5 MED dose estimation ..... 33
3.1 MSD Estimation ..... 41
3.2 MSD boxplot ..... 43
4.1 Efficiency of parameter estimates based on separate and joint modeling - Emax for efficacy and Exponential for safety ..... 56
4.2 Percent bias of efficacy parameter estimates based on separate and joint modeling -Emax for efficacy and Exponential for safety ..... 57
4.3 Percent bias of safety parameter estimates based on separate and joint mod- eling -Emax for efficacy and Exponential for safety ..... 58
4.4 MSE of parameter estimates based on separate and joint modeling -Emax for efficacy and Exponential for safety ..... 59
4.5 Testing procedures of PoC for efficacy and safety endpoints ..... 60
4.6 Flow chart of joint model fitting-Strategy I ..... 63
4.7 Flow chart of joint model fitting-Strategy II ..... 65
4.8 Flow chart of the joint model fitting-Strategy I Example ..... 66
4.9 Flow chart of the joint model fitting-Strategy II example ..... 67
4.10 Plots of $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ based on separate or joint fitting: horizontal and vertical line represent true MED and true MSD, respectively. ..... 76
4.11 Plots of $\widehat{M E D}_{2}$ and $\widehat{M S D}_{2}$ based on separate or joint fitting: horizontal and vertical line represent true MED and true MSD, respectively. ..... 77
5.1 Plot of probability to satisfy efficacy and safety criteria ..... 82
5.2 The set of doses satisfy the joint criteria with at least $60 \%$ probability or the best dose which has the maximum probability of success ..... 83
5.3 Best dose with the maximum utility index by varying the weight of safety response ..... 87
5.4 Best dose with the maximum utility index by varying the weight of stan- dardized safety response ..... 88
9.1 Efficiency of parameter estimates based on separate and joint modeling - Emax for efficacy outcome and probit for safety outcome (QELS approach). ..... 152
9.2 Percent bias of efficacy parameter estimates based on separate and joint modeling - Emax for efficacy outcome and probit for safety outcome (QELS approach) ..... 153
9.3 Percent bias of safety parameter estimates based on separate and joint mod- eling - Emax for efficacy outcome and probit for safety outcome (QELS ap- proach) ..... 154
9.4 MSE of parameter estimates based on separate and joint modeling - Emax for efficacy outcome and probit for safety outcome (QELS approach) ..... 155
9.5 Efficiency of parameter estimates for correlations based on simple Pearson correlation coefficient estimate and joint modeling - Emax for efficacy out- come and probit for safety outcome (QELS approach) ..... 156
9.6 Percent bias of parameter estimates for correlations based on simple Pear- son correlation coefficient estimate and joint modeling - Emax for efficacy outcome and probit for safety outcome (QELS approach) ..... 157
9.7 MSE of parameter estimates for correlations based on simple Pearson corre- lation coefficient estimate and joint modeling - Emax for efficacy outcome and probit for safety outcome (QELS approach) ..... 158

## List of Tables

2.1 A selection of frequently used dose-response models ..... 18
2.2 Median relative bias and relative IQR of $M E D$ estimate under various dose- response shapes(from table 7 in Bretz et al (2005)) ..... 34
3.1 Bias and dispersion of $\widehat{M S D}_{1}$ and $\widehat{M S D}_{2}$ ..... 44
3.2 Coverage of $95 \%$ bootstrap C.I. for true $M S D$ ..... 45
3.3 Comparison of lower bound of $95 \%$ bootstrap C.I. with true MSD ..... 45
4.1 Parameter estimates based on separate and joint modeling -Emax for effi- cacy and Exponential for safety ..... 55
4.2 The $\widehat{M E D_{2}}$ and $\widehat{M S D}_{1}$ from separate fitting and joint fitting with $\rho=0$ ..... 69
4.3 The $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ from separate fitting and joint fitting with $\rho=0.4$ ..... 70
4.4 The $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ from separate fitting and joint fitting with $\rho=0.8$ ..... 71
4.5 The correlation and variance by simulation and gnls output ..... 72
$4.6 \widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ and true mean based on separate and joint fittings: ..... 74
5.1 The best set of dose(s) satisfy joint efficacy and safety criteria ..... 84
5.2 The best set of dose satisfy different weight of efficacy and safety criteria ..... 89
9.1 Simulated bivariate continuous and discrete outcomes (assuming discrete outcome follows probit regression) ..... 141
9.2 Simulated bivariate continuous and discrete outcomes (assuming discrete outcome follows logistic regression) ..... 141
9.3 Simulated bivariate continuous and discrete outcomes (assuming discrete outcome follows complementary log-log regression) ..... 142
9.4 Comparison of parameter estimates and their standard errors between the ELS and QELS approaches-Emax for efficacy outcome and Probit for safety outcome ..... 144
9.5 Comparison of parameter estimates and their standard errors between the ELS and QELS approaches-Emax for efficacy outcome and Logit for safety outcome ..... 145
9.6 Comparison of parameter estimates and their standard errors between the ELS and QELS approaches-Emax for efficacy and Cloglog for safety ..... 146
9.7 Comparison of the ELS and QELS approaches for joint nonlinear continuous and discrete outcomes ..... 147
9.8 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and probit for safety outcome (QELS approach): $\rho=0$ ..... 149
9.9 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and probit for safety outcome (QELS approach): $\rho=0.4$ ..... 150
9.10 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and probit for safety outcome (QELS approach): $\rho=0.8$ ..... 151
9.11 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and logit for safety outcome (QELS approach): $\rho=0$ ..... 159
9.12 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and logit for safety outcome (QELS approach): $\rho=0.4$ ..... 160
9.13 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and logit for safety outcome (QELS approach): $\rho=0.8$ ..... 161
9.14 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and Cloglog for safety outcome (QELS approach): $\rho=0$ ..... 162
9.15 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and Cloglog for safety outcome (QELS approach): $\rho=0.4$ ..... 163
9.16 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and Cloglog for safety outcome (QELS approach): $\rho=0.8$ ..... 164
9.17 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and probit for safety outcome (ELS approach): $\rho=0$ ..... 167
9.18 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and probit for safety outcome (ELS approach): $\rho=0.4$ ..... 168
9.19 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and probit for safety outcome (ELS approach): $\rho=0.8$ ..... 169
9.20 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and logit for safety outcome (ELS approach): $\rho=0$ ..... 170
9.21 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and logit for safety outcome (ELS approach): $\rho=0.4$ ..... 171
9.22 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and logit for safety outcome (ELS approach): $\rho=0.8$ ..... 172
9.23 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and Cloglog for safety outcome (ELS approach): $\rho=0$ ..... 173
9.24 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and Cloglog for safety outcome (ELS approach): $\rho=0.4$ ..... 174
9.25 Parameter estimates based on separate and joint modeling - Emax for effi- cacy outcome and Cloglog for safety outcome (ELS approach): $\rho=0.8$ ..... 175
9.26 Model mis-specification fitting when there is no correlation between bivari- ate responses ..... 177
9.27 Model mis-specification fitting when there is weak correlation between bi- variate responses ..... 177
9.28 Model mis-specification fitting when there is strong correlation between bivariate responses . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 178

## Part I

# Dose finding for Joint Continuous 

## Bivariate Responses

## Chapter 1

## Introduction

### 1.1 Background

Selection of appropriate dose(s) to carry into confirmatory Phase III trials is one of the most difficult decisions that need to be considered during drug development. It is believed by many that the high attrition rate currently observed in Phase III is largely driven by inadequate dose selection (FDC Report, 1991; Bornkamp et al., 2007). Most commonly used dose finding designs and methods today still focus on selection of a target dose out of a fixed, generally small, number of dose levels, via pairwise hypothesis testing, which is typically inefficient (FDA, 2004). Assessment of dose-response should be an integral component of drug development, with studies designed to assess dose-response as an inherent part of establishing the safety and effectiveness of the drug. If characterization of dose-response relationship is built into the development process, it can usually be accomplished with no loss of time and minimal extra effort compared to development plans that ignore dose-response estimation. Proper estimation of such dose-response profiles for rel-
evant safety and efficacy endpoints allows the reliable evaluation of the risk-benefit profile of a drug at the end of Phase II, as well as the selection of appropriate doses to be brought into confirmatory Phase III trials. Thus how to select dose(s) in Phase II trials by combining information about the efficacy and safety in a joint model setting may play a key role in drug development programs, and are often the gate-keeper for large confirmatory Phase III trials with greater chance of successful approval.

The primary goal of dose-response studies (Bretz et al., 2005; Dragalin, 2007) is to establish the dose-response relationship or to find the target dose, usually the minimal effective dose (MED). There are fixed and adaptive approaches for designing and analyzing dose-ranging studies (Bornkamp et al., 2007). These approaches target mostly efficacy as the goal to get the optimal dose. Most Phase II designs assume that a dose range with an acceptable toxicity has been previously determined and aim to establish treatment efficacy at some dose within this range. However, under a variety of circumstances the safety may lead to the early termination of the drug development in Phase II/III trials. Therefore it is important to address safety and efficacy simultaneously. There has been some research on how to design dose-finding studies based on efficacy-toxicity response in early phase(I/II) (Thall and Cook, 2004; Dragalin, 2005; Thall et al., 2008) for oncology trials. Part I of this dissertation will address how to combine continuous efficacy and safety responses in Phase II trials in a joint model setting.

Bretz et al. (2005) combined multiple comparison procedure and modeling (MCPMod), which has been used extensively for analyzing dose finding trials recently. It includes a PoC (proof-of-concept) assessment and a dose-selection step. The clear advantage
of this approach, compared to traditional multiple comparison dose finding methods, is its added flexibility in selecting an appropriate dose-response model for future drug development. Bretz et al. (2005) considered only efficacy to identify the minimum effective dose assuming all considered doses are within safety tolerance. However, as mentioned earlier in some cases both efficacy and safety may need to be considered for selecting optimum doses to carry into Phase III trials. Typically the two outcomes (efficacy and safety) for the same patient are usually correlated, how to determine the optimal dose(s) account for correlated efficacy and safety outcomes remains unresolved. Part I of this dissertation will extend the MCP-Mod approach to select the best joint model based on two continuous correlated efficacy and safety outcomes and to get the final optimum dose(s) from the best joint model for the Phase III study.

### 1.2 Motivating example

ACE inhibitors (inhibitors of Angiotensin-converting enzyme) are used primarily in treatment of hypertension and heart failure. ACE inhibitors are used first-line as several agents in the class have been clinically shown to be superior to other classes of drugs in the reduction of morbidity and mortality for cardiovascular disorders and hypertension (Thomas, 2000; Rossi, 2004).

Normally, angiotensin II will have the following effects (Dluhy et al., 2004; Krum et al., 2002; Flack et al., 2003; Epstein et al., 2006) :

- Vasoconstriction (narrowing of blood vessels), which may lead to increased blood pressure and hypertension. Specifically, angiotensin II constricts the efferent arteri-
oles of the kidney, leading to increased perfusion pressure in the glomeruli.
- Stimulate the adrenal cortex to release aldosterone, a hormone that acts on kidney tubules to retain sodium and chloride ions and excrete potassium. Sodium is a "water-holding" molecule, so water is also retained, which leads to increased blood volume, hence an increase in blood pressure.
- Stimulate the posterior pituitary into releasing vasopressin (also known as anti-diuretic hormone (ADH)) which also acts on the kidneys to increase water retention.

With ACE inhibitor use, the effects of angiotensin II are prevented, leading to decreased blood pressure.

Renal impairment is a significant adverse effect of all ACE inhibitors. The reason for this is still unknown. Some suggest that it is associated with their effect on angiotensin II-mediated homeostatic functions such as renal blood flow. Renal blood flow may be affected by angiotensin II because it vasoconstricts the efferent arterioles of the glomeruli of the kidney, thereby increasing glomerular filtration rate (GFR). Hence, by reducing angiotensin II levels, ACE inhibitors may reduce GFR, a marker of renal function.

In one clinical trial an ACE inhibitor (Drug A) is used to treat hypertension. The efficacy endpoint is the change of sitting blood pressure from baseline. Decreasing GFR is the undesirable effect and the main safety measure is the change of GFR from baseline. Both response variables are assumed to be normally distributed. When the bivariate continuous outcomes are considered to find a suitable dose, joint model fitting that accounts for the correlation between efficacy and safety outcomes should be superior to the separate modeling that ignores the correlation.

### 1.3 Research objectives

In Part I of this thesis, we set the following objectives:

### 1.3.1 Objectives

Objective I: To estimate the MSD (maximum safety dose) when the safety response is considered (Chapter 3).

Objective II: To formulate and select the joint model when continuous efficacy and safety responses are correlated and both need to be included for the dose selection (Chapter 4).

Objective III: To find more accurate MED and MSD estimates when the correlated efficacy and safety responses are both included in the dose selection (Chapter 4).

Objective IV: To identify the final dose(s) to carry into the Phase III program with a high probability of success (Chapter 5).

### 1.3.2 Specific aims

The specific aims for Part I of this dissertation are described as follows:

1. Identify the MSD (maximum safety dose) if the safety outcome is of interest.

In clinical trials, when the safety outcome is the main interest in the study, we need to find the maximum safety dose (MSD) for subsequent drug development. Bretz et al. (2005) developed MCP-Mod method to estimate the MED by three different rules based on the predicted response at each dose level and the corresponding confidence level. We will extend the MCP-Mod approach for the safety endpoint to identify the MSD. How to
define the MSD and estimate it accordingly will be explored in Part I.
2. Estimate the parameters for the continuous efficacy and safety outcomes corresponding to the joint modeling of the bivariate outcomes.

The MCP-Mod method was developed for dose selection based on continuous efficacy outcome. When both the safety outcome and the efficacy outcome are considered for dose selection, the use of MCP-Mod based on separate model fitting will not account for the correlation between endpoints, thus the parameter estimations based on the separate model fitting are not precise and efficient. In addition it is common that the dose-response models for both continuous efficacy and safety data are nonlinear. Therefore how to formulate the joint nonlinear continuous bivariate model and how to estimate its parameters will be addressed in this dissertation.
3. Find more precised MED and MSD estimates based on the correlated bivariate responses for dose selection.

From the above two specific aims we will be able to find the $M E D$ and $M S D$ from the joint bivariate outcomes, but they may not be the most accurate estimates. The rationale behind this statement is as follows: both efficacy and safety outcomes can have more than one significant model to fit the data. For example, for efficacy data we may find both Emax and quadratic models can fit the data well, with only slightly different AIC or t-statistics. A similar case can take place for the safety data. Then we need to develop a strategy for how to proceed to find the final joint model when there are more than one significant model for either the efficacy or safety and what criteria to use for determining the best model. After the final best fitted joint model is determined, the MED or MSD can
be computed based on the final joint model.
4. Define the criteria and identify the optimal dose(s) based on the defined criteria for the Phase III program.

After the MED and MSD are estimated, what exactly are the final dose(s) to used in the Phase III program still may not be clear. In order to identify the final optimal dose(s), first we may need to define the success criteria for the Phase III program. The next step we will explore is how to identify the final dose(s) after the success criterion is clearly defined. All these questions will be answered in this dissertation.

## Chapter 2

## Literature Review

### 2.1 Existing approaches for dose finding

The traditional process of dose selection in a clinical program for a new candidate drug begins with human studies, when single ascending doses are used, followed by multiple ascending doses in healthy volunteers. This is followed by empiric testing of doses in small patient cohorts, looking both for differences in the pharmacokinetic profile and evidence that the drug has an effect on the disease being studied, the so called proof-of-concept (PoC) study. When the drug moves into Phase II, or even Phase III programs, often only small amounts of efficacy and safety data are available to justify the dose selection. Until recently, dosing decisions were guided by the availability of prior information, which changed from the traditional approach based on a purely empirical method to the approach based on verification of a derived model.

There are a number of different dose-finding approaches to address different objectives. There also exist a number of different statistical methodologies that have been
developed to allow adaptations to take place without compromising the statistical integrity of the trial (e.g., preserving the Type I error rate). Mainly there are 3 classes of approaches for dose-finding studies based on very different demands on the logistics of planning and running a trial. The first class is fixed designs with pre-determined randomization strategy and a single analysis at the end of study, represented by ANOVA and multiple comparison-modeling (MCP-Mod). The second class is discrete adaptation design in which randomization can be adjusted (e.g., patients are enrolled in groups and an interim analysis is performed after each group has completed the study) by D-Optimal response-adaptive approach (Dopt). The third class is continuous adaptation which allows adjustment of randomization (e.g., the subject is chosen by simulating the effects of randomizing the next subject to each of possible doses and finding the one that minimizes the variance of a parameter of interest) represented by the general adaptive dose allocation approach (GADA). The detailed summary for the above approaches may refer to Bornkamp et al. (2007). In this dissertation the main focus is on fixed clinical trial designs, with the MCP-Mod approach being discussed in detail.

As described above, except for some approaches to estimate maximum tolerated dose in Phase I trials (e.g, CRM), most approaches only estimate the minimum effective dose for Phase II program. Some authors (Thall and Cook, 2004; Thall et al., 2008) have discussed how to select a target dose based on efficacy-toxicity trade off or bivariate outcomes with patient specific covariates with their research more focused on Phase I/II oncology trials for binary endpoints, which may not be suitable for other therapeutic areas. This dissertation will address how to develop an approach based on MCP-Mod to select
final optimum dose(s) using efficacy and safety data for phase II trials in a general setting. In order to introduce the proposed approach and the need for using a joint model to select target doses, a review of the MCP-Mod approach is necessary.

### 2.2 MCP-Mod

The analysis of dose finding studies can be classified into two major strategies: modeling techniques (Pinheiro et al., 2006; Bates and Watts, 1988) and multiple comparison procedures (MCP) (Hochberg and Tamhane, 1987; Hsu, 1996). In the next sections we will review the MCP and modeling techniques before we introduce MCP-Mod approach.

### 2.2.1 Multiple comparison procedures (MCP)

MCP can be considered in analysis-of-variance (ANOVA) settings, which regard the dose as a qualitative factor and make no or only few assumptions about the underlying dose-response model. This approach is easy to implement and interpret, does not require prior knowledge of dose response relationship and is less sensitive to assumptions. The inference is restricted to a set of doses under investigation. This procedure is robust with respect to the underlying dose-response shape, but not designed for the exploration of information other than the dose levels studied. There are two methods can be applied in MCP to find the minimum effective dose (MED) (Bretz et al., 2008). First method is Dunnett's (1955) method which uses the two-sample t method to compare each dose group with control group while adjusting for multiplicity. Second method is a stepdown method with pre-determined steps (Stefansson, et al., 1988; Finner and Strassburger, 2002). This
method offers some advantage of using a statistical method designed to give a contiguous set of doses as efficacious when the Dunnett's method may infer a dis-contiguous set of doses to be efficacious due to sampling variation. If the dose levels giving the higher sample responses are the ones tested early in the steps of the stepwise method, then the stepwise method will infer more doses as efficacious than Dunnett's method. Otherwise Dunnett's method will infer more doses as efficacious than the stepwise method. The details are described in Bretz et al. (2008).

### 2.2.2 Modeling approaches

The modeling approach to dose finding is based on an assumed functional relationship between the clinical endpoint and the dose, treated as a continuous variable, according to a pre-specified parametric model. The MED dose or other target dose is estimated by inverse regression techniques and confidence intervals can be provided on the estimated doses. This approach is easy to include requirements of clinical relevance and leads to better understanding of the dose-response relationship, which allows planning of future studies and simulations. The typical model-based analysis does not provide a rigid Type I error control and the conclusions rely highly on the right choice of the dose-response model. Furthermore the validity of its conclusions will highly depend on the correct choice of the dose response model, which is of course a prior unknown. Bretz et al.(2008) also discussed the alternative modeling approaches such as non-parametric models and Bayesian methods. Bornkamp et al. (2007) considered a non-parametric dose response modeling approach based on local polynomial fits (Loader, 1999). Bayesian methods can be used with the model-based methods as an extension. The bayesian dose-finding approach offer
additional flexibility and in some cases can ease of interpretation compared to frequentist approaches. But all priors for the model parameters need to be specified and computational complexity is increased (Berry et al., 2001).

Figure 2.1 provides an illustration of finding the $M E D$ by multiple comparison procedure and modeling approach. It is shown that MCP approach treats dose as discrete variable while the modeling treats the dose as continuous variable.


Figure 2.1: Multiple comparisons and Modeling approach to find MED

Bretz et al. (2005) proposed a hybrid methodology which combines aspects of MCP and modeling into a unified strategy for dose-finding studies. Typically decisions derived from dose-response studies can be divided into two main steps: establishing that the treatment has some effect on the outcome, which has evidence of dose response and is called proof-of-concept (PoC), and selecting a target dose from the best fitted model. The first step starts with a set of candidate models covering a suitable range of dose-response shapes. Each of the models in the candidate set is assessed using predefined contrast tests and applying MCP techniques to preserve the family-wise error rate (FWER). At least one of the model contrast tests needs to be significant in order to establish PoC. Otherwise, the procedure stops at this step and concludes that there is no sufficient evidence of a doseresponse relationship in the study. Once the overall dose-response relationship has been established (PoC), the next step is to select a target dose from the best fitted model. The
selection of the best model from possibly more than one statistically significant models can be based on the minimum p-value of the test statistics or some other relevant model selection criteria such as the Akaike's information criteria (AIC) (Akaike, 1973) or the Bayesian Information Criteria (BIC) (Schwarz, 1978). The target dose is then estimated using inverse regression techniques based on the selected dose-response model.

### 2.2.3 MCP-Mod approach procedures

Figure 2.2 provides the steps of MCP-Mod approach, the details are illustrated in the text of body in this Section.


Figure 2.2: MCP-Mod flow chart

## Candidate models

Several candidate parametric models are identified for the response Y and a given set of parallel groups of patients corresponding to doses $d_{2}, d_{3}, \ldots, d_{k}$, plus placebo group $d_{1}$, for a total $k$ treatment groups. For the purpose of dose estimation, a one-way layout for the model is specified as follows:

Model:

$$
\begin{gathered}
Y_{i j}=f\left(d_{i}, \boldsymbol{\theta}\right)+\epsilon_{i j}, \epsilon_{i j} \stackrel{i n d}{\sim} N\left(0, \sigma^{2}\right), \\
i=1, \ldots, k, j=1, \ldots, n_{i},
\end{gathered}
$$

where $\theta$ refers to the vector of model parameters, $i$ to the dose group ( $i=1$ corresponds to placebo), and $j$ to the patient within dose group $i$.

Following Bretz et al. (2005), most dose-response models used can be written as standardized version $f^{0}$ of the dose response model:

$$
f(d, \boldsymbol{\theta})=\theta_{0}+\theta_{1} f^{0}\left(d, \boldsymbol{\theta}^{0}\right),
$$

$\theta_{0}$ is a location parameter and $\theta_{1}$ is a scale parameter. The advantages of using the standardized model $f^{0}$ instead the full model $f$ will become clear when obtaining the optimum contrast coefficients later on. Prior estimates for the standardized model parameters $\boldsymbol{\theta}^{0}$ are typically derived from initial knowledge (or guesses) of the expected percentage $p^{*}$ of the maximum response associated with a given dose $d^{*}$. Table 2.1 provides a list of models frequently used to represent dose-response relationships, together with their respective standardized versions.

| Model | $f(d, \boldsymbol{\theta})$ | $f^{0}\left(d, \theta^{0}\right)$ | $\theta_{1}$ | $\theta^{0}$ | Example |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Emax | $E_{0}+\frac{E_{\text {max }}}{E D_{50}+d}$ | $d /\left(E D_{50}+d\right)$ | $E_{\text {max }}$ | $E D_{50}$ | . $2+.96 d /(.2+d)$ |
| Linear log-dose | $E_{0}+\delta \log (d+1)$ | $\log (d+1)$ | $\delta$ | - | $.2+1.15 \log (d+1)$ |
| Linear | $E_{0}+\delta d$ | $d$ | $\delta$ | - | $0.2+0.8 d$ |
| Exponential | $E_{0}+E_{1} \exp (d / \delta)$ | $\exp (d / \delta)$ | $E_{1}$ | $\delta$ | . $1772+.0228 \exp \left(\frac{d}{279}\right)$ |
| Quadratic | $E_{0}+\beta_{1} d+\beta_{2} d^{2}$ | $d+\frac{\beta_{2}}{\left\|\beta_{1}\right\|} d^{2}, \beta_{2}<0$ | $\beta_{1}$ | $\beta_{1}, \beta_{2}$ | . $2+2.7344 d-2.335 d^{2}$ |
|  |  | $-d+\frac{\beta_{2}}{\left\|\beta_{1}\right\|} d^{2}, \beta_{2}>0$ | $\beta_{1}$ | $\beta_{1}, \beta_{2}$ |  |
| Logistic | $E_{0}+\frac{E_{\text {max }}}{\left.1+\exp \left(E D_{50}-d\right) / \delta\right)}$ | $\frac{1}{1+\exp \mid\left(E D_{50}-d\right) / \delta}$ | $E_{\text {max }}$ | $E D_{50}, \delta$ | . $19+\frac{.811}{1+\exp [.4-d) / .091]}$ |
| Note: $\theta_{0}=E_{0}$ for all models. |  |  |  |  |  |

Figure 2.3 provides the dose-response curves for the different models shown in Table 2.1 as an example. The open dots in the figure indicate the responses at the dose levels used in the simulation in Part I of this dissertation. The model means at these dose levels are generated by the mean model designed for the simulation. All the dose-response curves have the same baseline values and maximum change from baseline values. The models used for generating the curves are indicated in Table 2.1.


Figure 2.3: Dose-response models

## Optimal contrast coefficients

Monotonicity is not required for the PoC Two extreme approaches had been advocated in the 1970s: (i) Assume the existence of monotonicity of response with dose (Williams, 1971, 1972). (ii) Assume nothing about response shape and use a procedure to compare individual dose responses with the control response, taking account of the fact that the k comparisons share a common control (Dutt et al.,1976).

Tukey et al. (1985) considered a more balanced approach which recognizes that (i) if there is a response, it will most often (but not always) be monotone, and (ii) there are many situations where the direction of any effect is far from certain. They proposed regression analysis approach to test the existence of a trend in the response variable. The test assesses the trend as the most extreme P-value observed from a candidate set of "dose carriers" (arithmetic, ordinal, and arithmetic-logarithmic as defined in Tukey et al., 1985). It is claimed that the regression approach is both more trustworthy than assuming monotonicity and more powerful than any individual comparison approach (Tukey et al., 1985).

Regression analysis uses within-group estimates of variability to detect a nonzero trend. Doing regression on dose level $d_{i}$ and comparing the result with the corresponding linear combination of estimated variances $\operatorname{var}_{e s t}[\boldsymbol{y}]$ will have a sensitivity for detection of an effect with the average responses $\eta_{i}=$ ave $[\boldsymbol{y}]$ determined by the formal correlation between the $d_{i}$ and $\eta_{i}$ (Abelson and Tukey, 1963). Here, $y=\left(y, \ldots, y_{n}\right), y_{i}=\mu_{i}+\epsilon_{i}=$ $\alpha+\beta d_{i}+\epsilon_{i}$ and the correlation is computed as follows:

$$
r=\frac{\sum\left(d_{i}-\bar{d}\right)\left(\eta_{i}-\bar{\eta}\right)}{\left[\sum\left(d_{i}-\bar{d}\right)^{2} \sum\left(\eta_{i}-\bar{\eta}\right)^{2}\right]^{1 / 2}}
$$

The validity for detection of the trend test is not dependent on assumptions about the behavior of the $\eta_{i}$. The value of $r$, and thus the sensitivity of the procedure compared to what would be possible if the $\eta_{i}$ is known, must of course depend on the degree of general similarity of the $\bar{\eta}$ 's and the $d^{\prime}$ 's. This approach can achieve very high power when the response pattern has a very high correlation with the dose level and respectable power against a wide variety of response pattern. Thus the validity of the procedure are not seriously distorted or troubled in the instances when the repose pattern is not monotone, although the sensitivity is somewhat decreased.

There were other approaches such as likelihood ratio test (LRT; Roberson, Wright, and Dykstra, 1988) and the step contrasts (Bauer and Hackl, 1985) to test PoC. LRT is known to be one of the most powerful tests for trend throughout the order restricted alternative region $\mu_{1} \leq \ldots \leq \mu_{k}$. LRT is designed for the PoC only and thus is not for finding any information about the underlying dose-response curves. The step contrasts match exactly the corner vectors of the polyhedral cone described through the relationship $\mu_{1} \leq \ldots \leq \mu_{k}$ and span the order restricted space of interest. Both approaches used assumptions of monotonicity.

Bretz et al. (2005) computed optimum weights based on the dose-response curves for the comparison of the different candidate models within a multiple hypothesis testing framework. The "best" contrast associated with the candidate model will maximizes the chance of rejecting the associated null hypothesis. There is no monotonicity required when
optimum contrast is used for testing the dose-response curves. As described in Bretz et al. (2005), the following linear model was assumed for the purpose of detecting an overall trend:

$$
\begin{gathered}
Y_{i j}=\mu_{i}+\epsilon_{i j}, \epsilon_{i j} \stackrel{i n d}{\sim} N\left(0, \sigma^{2}\right), \\
i=1, \ldots, k, j=1, \ldots n_{i}
\end{gathered}
$$

where $\mu_{i}=f\left(d_{i}, \boldsymbol{\theta}\right)$ are the unknown treatment means with $\mu=\left(\mu_{1}, \ldots, \mu_{k}\right)$. Let $\bar{Y}_{i}=$ $\sum_{j=1}^{n_{i}} Y_{i j} / n_{i}$ denote the arithmetic mean of group $i$ with $\overline{\boldsymbol{Y}}^{\prime}=\left(\bar{Y}_{1}, \ldots, \bar{Y}_{k}\right)$. Further let $S^{2}=$ $\sum_{i=1}^{k} \sum_{j=1}^{n_{i}}\left(Y_{i j}-\bar{Y}_{i}\right)^{2} / v$ denote the pooled variance estimator with $v=\sum_{i} n_{i}-k$ degree of freedom. The testing of PoC requires the estimation of the parameter $\boldsymbol{\theta}$. Under the assumption of independent, identically distributed errors, ordinary least squares (OLS) estimates that minimize the residual sum of squares for dose-response models can be used. Non-linear squares algorithms are needed for estimating $\boldsymbol{\theta}$. The most oftenly used is the Gauss-Newton algorithm (Bates and Watts, 1988; Seber and Wild, 1989), which is an iterative procedure consisting of solving, until convergence, a sequence of linear least squares problems based on a local approximation of the nonlinear model. Such iterative algorithms typically require a starting point and methods for deriving the initial estimates for nonlinear models are discussed in Bates and Watts (1988).

If $M$ candidate models are identified with $\Omega=\left\{M_{m}, m=1, \ldots, M\right\}$. These models generate the mean response vector $\boldsymbol{\mu}_{m}=\left(\mu_{m 1}, \ldots, \mu_{m k}\right)^{\prime}$, where $\mu_{m i}=f_{m}\left(d_{i}, \boldsymbol{\theta}_{m}\right)$. Each of the dose-response shapes in the candidate set is tested using a single contrast test, with coefficients chosen to maximize the power of the test when the true underlying mean response equals $\boldsymbol{\mu}_{m}$. Abelson and Tukey (1963) gave an early introduction of contrast tests in
the context of dose-response analyses. A linear contrast test is defined as the difference of any two adjacent contrast coefficients which is a constant. Assuming that the standard linear model has been included in the candidate set, the linear contrast test is then a powerful test to detect the linear trend. The following hypotheses are tested:

$$
H_{0}^{m}: \boldsymbol{c}_{m}^{\prime} \boldsymbol{\mu}=0 \text { vs. } H_{1}^{m}: \boldsymbol{c}_{m}^{\prime} \boldsymbol{\mu}>0 .
$$

To determine the "best" contrast associated with a given model function $f(d, \boldsymbol{\theta})$, when the model is correct, it will maximize the chance of rejecting the associated null hypothesis, that is, it maximizes the non-centrality parameter $\tau=\tau(c)$. Thus we should choose $\mathbf{C}_{\text {opt }}(f)$ such that

$$
\mathbf{C}_{o p t}(f)=\underset{\mathbf{c}}{\operatorname{argmax}} g(\boldsymbol{c}, \boldsymbol{\mu})
$$

where

$$
g(\boldsymbol{c}, \boldsymbol{\mu})=\frac{\left(\boldsymbol{c}^{\prime} \boldsymbol{\mu}\right)^{2}}{\sum_{i=1}^{k} c_{i}^{2} / n_{i}}=\sigma^{2}[\tau(\boldsymbol{c})]^{2}
$$

Without loss of generality, we assume that the contrast vectors $\mathbf{c}_{m}=\left(c_{m 1}, \ldots, c_{m k}\right)^{\prime}$ follow the regularity conditions $\sum_{i=1}^{k} c_{m i}=0$ and $\sum_{i=1}^{k} c_{m i}^{2}=1$.

Assuming that there exists a standardized version $f^{0}$ of $f$ :

$$
\mu=\mu(f)=\theta_{0}+\theta_{1} \boldsymbol{\mu}\left(f^{0}\right)=\theta_{0}+\theta_{1} \boldsymbol{\mu}^{0}
$$

Due to the shift and scale invariance properties, the optimum contrast coefficients depend only on the standardized model $f^{0}$ as follows:

$$
\mathbf{c}^{\prime} \boldsymbol{\mu}=\boldsymbol{\theta}_{1} \mathbf{c}^{\prime} \boldsymbol{\mu}^{0}, g(\mathbf{c}, \boldsymbol{\mu})=\theta_{1}^{2} g\left(\mathbf{c}, \boldsymbol{\mu}^{0}\right) .
$$

Thus,

$$
C_{o p t}(f)=C_{o p t}\left(f^{0}\right)
$$

Under balanced sample size allocation the optimum contrast coefficients are found as follows:

$$
C_{o p t}(\boldsymbol{\mu})=\frac{\boldsymbol{\mu}-\bar{\mu} \mathbf{1}}{\|\boldsymbol{\mu}-\bar{\mu} \mathbf{1}\|}=\frac{\boldsymbol{\mu}^{0}-\bar{\mu}^{0} \mathbf{1}}{\left\|\boldsymbol{\mu}^{0}-\bar{\mu}^{0} \mathbf{1}\right\|} .
$$

Details on the computation of optimum contrast coefficients for model testing can be found in Bretz et al.(2005). Figure 2.4 shows the optimal contrasts for various models to be tested. Similarly to Figure 2.3, the open dots in the figure indicate the contrast for the dose levels used in the simulation. The contrast coefficients are determined by the model means at these dose levels, which are generated by the mean model designed for the simulation.


Figure 2.4: Plot of optimal contrasts

## Selection of significant models while controlling FWER

A common approach in situations when one has to select the model that ultimately is used to fit the data is to use information criteria based on a reasonable discrepancy measure to assess the lack of fit. A number of criteria are available for model selection. Commonly used criteria is the ratio $R^{2}$ of the sum of squares for the regression to the total sum of squares. The problem with the $R^{2}$ is that the sum of squares for the regression, and hence by construction $R^{2}$ itself, increases with the number of parameters and thus leads to over-fitting. There are some alternatives measures that have been proposed such as the Akaike information criterion (AIC) or Bayesian information criterion (BIC). All these and other measures are generally not suitable in dose-response analyses as they do not incorporate potential constraints, such as the simple order restriction $\mu_{1} \leq \ldots \leq \mu_{k}$. The theory of order restricted inference that the maximum likelihood estimates for the mean level responses subject to a given order restriction are different from the unrestricted maximum likelihood estimates (Robertson et al., 1988). Anraku (1999) thus proposed to use an order restricted information criterion (ORIC) based on monotonic regression theory.

However, any of the above measures of fit (AIC, BIC, ORIC or any other criterion) has the inherent drawback of missing family-wise error control. For example, we would have no conclusion on the validity of the decision if we simply selected the model by the best ORIC criteria. In addition, the application of the ORIC or AIC will always lead to the selection of one single model, irrespective the goodness of fit given the observed data.

Buckland et al. (1997) proposed a philosophy for weighting contending models in preference to selecting between the models that can be done within a Bayesian frame-
work. The model selection is replaced by estimated probabilities that models are correct. Buckland et al. (1997) introduced the weighted estimate approach to compute the parameter estimates. Let $\Omega=\left\{M_{1, \ldots,}, M_{L}\right\}$ denote a set of $L$ candidate models, the weights are determined by the common information criterion IC described above applied to each models. The weighted estimate is defined as

$$
\widehat{\boldsymbol{\mu}}=\sum_{l} w_{l} \widehat{\boldsymbol{\mu}}_{l}
$$

where $\widehat{\mu}_{l}$ is the estimate of $\mu$ under model for given weights $w_{l}$ and

$$
w_{l}=\frac{e^{-\frac{I C_{l}}{2}}}{\sum_{j=1}^{L} e^{-\frac{I C_{j}}{2}}}, l=1, \ldots, L .
$$

In addition, Bayesian model averaging techniques are discussed (Hoteing et al., (1999) and Clyde and George (2004)).

$$
P(\boldsymbol{\mu} \mid X)=\sum_{j=1}^{L} P\left(\boldsymbol{\mu} \mid X, M_{l}\right) P\left(M_{l} \mid X\right)
$$

where $X$ is the observed data and $P\left(M_{l} \mid X\right)$ represents posterior model probability for the investigated model.

Note although these approaches provide a simple and intuitive way to overcome some of the model uncertainty problems, one is still left with the open problem of how to ultimately choose the final model as a multiple hypotheses testing problem.

When performing multiple pairwise tests, familywise error rate (FWER) is the probability of making one or more false discoveries, or type I errors among all the hypotheses. Typical model-based analyses do not provide a rigid error control as it is provided, for example, by multiple comparison procedures (Hochberg and Tamhane, 1987; Hsu, 1996).

Selection of a specific model while controlling the familywise error rate at a pre-specified level $\alpha$ is described in Shimodaira (1998). A reference set of good models is constructed rather than choosing a single model. For example, denote the set of candidate models by $\mathcal{M}=\left\{M_{l} \mid l \in M\right\}$, for each $l \in M$, consider testing a set of hypotheses:

$$
\begin{gathered}
H_{l}: E\left(A I C_{M_{l}}\right) \leq \min _{M_{j} \in \mathcal{M} \backslash M_{l}} E\left(A I C_{M_{j}}\right) \text { vs. } \\
K_{l}: E\left(A I C_{M_{l}}\right)>\min _{M_{j} \in \mathcal{M} \backslash M_{l}} E\left(A I C_{M_{j}}\right),
\end{gathered}
$$

and include $M_{l}$ in the confidence set unless $H_{l}$ is rejected at a prescribed significance level. This construction leads to the following formula:

$$
\operatorname{Pr}\left\{l^{*} \in T\right\}=\operatorname{Pr}\left\{H_{l^{*}} \text { is not rejected }\right\} \geq 1-\alpha
$$

where $T$ is the confidence set and $l^{*}$ is the minimum $E(A I C)$ model and $E\left(A I C_{M_{l}}\right)$ is the expected AIC value for model $M_{l}$. Multiple comparison techniques (Gupta and Panchapakesan (1979), Hochberg and Tamhane (1987)) are then used to test $H_{l}$. The proposed multiple test procedure uses the standardized difference of any two AIC values within a variant of Gupta's subset selection procedure using bootstrap techniques to assess the joint distribution of the test statistics. The final confidence set at a given significance level is obtained as

$$
T=\left\{l \mid M_{l} \in M, P_{M_{l}} \geq \alpha\right\}
$$

where $P_{M_{l}}$ is the p -value associated with the $l$ th model. By construction, if $P_{M_{l}}<\alpha$, it has been shown that the AIC for the $l$ th model $M_{l}$ is significantly larger than the minimum AIC of the remaining set $\mathcal{M} \backslash M_{l}$. Thus, the present approach includes all models at the
beginning and only removes those models shown to behave inferiorly to other models. This approach never leads to $T=\varnothing$ and may contain more than one model at the end.

Bretz et al. (2005) selected the "best" model (if any), while controlling the FWER by the use of multiple comparison procedures similar in spirit to the ideas expressed in Schimodaira (1998). As described previously, each single contrast test thus tests whether a selected dose-response curve is significant given the observed data, while controlling the type I error rate at level $\alpha$.

The following hypotheses are tested:

$$
\begin{gathered}
H_{0}^{m}: \boldsymbol{c}_{m}^{\prime} \boldsymbol{\mu}=0 \\
\text { vs. } H_{1}^{m}: \boldsymbol{c}_{m}^{\prime} \boldsymbol{\mu}>0,
\end{gathered}
$$

where contrast vectors $\boldsymbol{c}_{m}^{\prime}=\left(c_{m 1}, \ldots, c_{m k}\right)$ are known constants subject to $\boldsymbol{c}_{m}^{\prime} \mathbf{1}=0, m=$ $1, \ldots, M$.

The single contrast tests are defined as

$$
T_{m}=\frac{\boldsymbol{c}_{m}^{\prime} \bar{\gamma}}{\sqrt{S^{2} \sum_{i=1}^{k} c_{m i}^{2} / n_{i}}}, m=1, \ldots, M .
$$

Each single contrast test can be translated into a decision procedure to determine whether a given dose-response shape is statistically significant. Under the assumptions above, the joint distribution of the vector $\boldsymbol{T}=\left(T_{1}, \ldots, T_{M}\right)^{T} \sim$ multivariate $\boldsymbol{T}_{M}(v ; \mathbf{0}, \boldsymbol{R})$ is M-variate T-distributed with $v$ degrees of freedom and correlation matrix $\boldsymbol{R}=\left(\rho_{i j}\right)$, where

$$
\rho_{i j}=\frac{\sum_{l=1}^{k} c_{i l} c_{j l} / n_{l}}{\left(\sum_{l=1}^{k} c_{i l}^{2} / n_{l} \sum_{l=1}^{k} c_{j l}^{2} / n_{l}\right)^{\frac{1}{2}}} .
$$

The computation of the critical value $q_{1-\alpha}$ should account for the multiplicity to control the FWER at a pre-specified level $\alpha$ (Hochberg and Tamhane, 1987). The associated probabilities are computed by numerical integration methods (Genz and Bretz,2002).

PoC is established if at least one contrast is significant i.e. $\max _{m}\left(T_{m}\right)>q_{1-\alpha}$ and all models with $T_{m}>q_{1-\alpha}$ are kept for possible use in the dose-response modeling.

## Model selection

Among the statistically significant models in the candidate set, a most adequate dose-response model is selected for dose estimation. Different criteria may be used to choose among models passing the PoC filter, e.g., max t-statistic, min AIC or min BIC. Target doses of interest are estimated using the selected model.

## Dose estimation and selection

Target doses can be selected out of the discrete dose set $D=\left\{d_{1}, \ldots, d_{k}\right\}$ under investigation or from the entire dose range $\left(d_{1}, d_{k}\right]$. To maintain confidentiality, as also described in Bretz et al. (2005), the actual doses have been rescaled to lie within $[0,1]$ interval. For illustration purpose this rescaled dose will be used in this dissertation for the dose ranges.

The MED is defined as the smallest dose which shows a clinically relevant effect. The absolute clinically relevant difference $\Delta$ with respect to the smallest dose $d_{1}$ (often placebo) is typically obtained from the guidelines/clinicians. Estimation of the $M E D \in D$ is conducted by applying appropriate multiple testing procedures (Tamhane, Dunnett, and Hochberg, 1996). Estimation of the model-based approaches allow $M E D \in\left(d_{1}, d_{k}\right]$. Given
a model $f(., \boldsymbol{\theta})$,

$$
M E D=\operatorname{argmin}_{d \in\left(d_{1}, d_{k}\right]}\left\{f(d, \boldsymbol{\theta})>f\left(d_{1}, \boldsymbol{\theta}\right)+\Delta\right\} .
$$

Let $p(d)=f(d, \widehat{\boldsymbol{\theta}})$ denote the predicted response at dose $d$ based on the model $f(., \boldsymbol{\theta})$, with corresponding $1-2 \gamma$ confidence interval $\left[L_{d}, U_{d}\right]$. Three different rules are proposed for estimating MED (Figure 2.5):

$$
\begin{aligned}
& \widehat{M E D_{1}}=\operatorname{argmin}_{d \in\left(d_{1}, d_{k}\right]}\left\{U_{d}>p\left(d_{1}\right)+\Delta, L_{d}>p\left(d_{1}\right)\right\}, \\
& \widehat{M E D_{2}}=\operatorname{argmin}_{d \in\left(d_{1}, d_{k}\right]}\left\{p(d)>p\left(d_{1}\right)+\Delta, L_{d}>p\left(d_{1}\right)\right\}, \\
& \widehat{M E D_{3}}=\operatorname{argmin}_{d \in\left(d_{1}, d_{k}\right]}\left\{L_{d}>p\left(d_{1}\right)+\Delta\right\} .
\end{aligned}
$$

By construction, $\widehat{M E D_{1}} \leq \widehat{M E D_{2}} \leq \widehat{M E D_{3}}$.


Figure 2.5: $M E D$ dose estimation

Table 2.2 (from table 7 in Bretz et al. (2005)) shows the examples of dose estimation bias and precision under various dose-response shapes. The simulation results from their paper showed that the estimated $\widehat{M E D}_{1}$ tends to underestimate the target dose, $\widehat{M E D}_{3}$ tends to overestimate it, and $\widehat{M E D}_{2}$ estimates the target dose more consistently.

Table 2.2: Median relative bias and relative IQR of MED estimate under various doseresponse shapes(from table 7 in Bretz et al (2005))

|  |  | Median relative bias (\%) |  |  | Relative IQR (\%) |  |  |
| :--- | ---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Model | n | $\overline{M E D 1}$ | $\overline{M E D 2}$ | $\widehat{M E D 3}$ | $\widehat{M E D 1}$ | $\overline{M E D 2}$ | $\overline{M E D 3}$ |
| Emax | 50 | -36.3 | 5.0 | 38.7 | 82.5 | 93.7 | 108.7 |
|  | 150 | -25.0 | 1.2 | 31.2 | 48.7 | 60.0 | 75.0 |
| Linear in log-dose | 50 | -24.0 | -0.1 | 23.8 | 56.5 | 63.0 | 71.7 |
|  | 150 | -13.1 | -0.1 | 17.3 | 28.2 | 36.9 | 43.4 |
| Linear | 50 | -22.0 | -7.0 | 9.5 | 28.5 | 33.0 | 34.5 |
|  | 150 | -11.5 | -2.5 | 8.0 | 15.0 | 18.0 | 19.5 |
| Exponential | 50 | -17.0 | -2.2 | 4.7 | 21.6 | 19.3 | 14.8 |
|  | 150 | -6.7 | 0.1 | 4.7 | 12.5 | 11.4 | 10.2 |
| Quadratic | 50 | -23.3 | 1.0 | 25.2 | 24.2 | 36.4 | 48.5 |
|  | 150 | -11.1 | 1.0 | 21.2 | 16.2 | 20.2 | 32.3 |
| Logistic | 50 | -11.8 | 5.4 | 29.0 | 38.7 | 40.9 | 45.4 |
|  | 150 | -7.5 | 3.2 | 20.4 | 30.1 | 30.1 | 32.3 |

Relative IQR (\%) is defined as the interquartile range of relative bias (\%).
Relative bias (\%) is calculated as the $100^{*}\left(\widehat{M E D}_{i}-M E D\right) / M E D$.

### 2.2.4 Bootstrap confidence interval approaches

Bootstrap is a data driven resampling method for statistical inference. The basic idea of the bootstrap is to use the sample data to compute a statistic and to estimate its sampling distribution based on resampling of observed data. Efron(1979) has considered two types of bootstrap procedures: nonparametric and parametric. The nonparametric bootstrap relies on consideration of the discrete empirical distribution generated by a random sample of size $n$ from an unknown distribution $F$. This empirical distribution $\widehat{F}_{n}$ assigns
equal probability to each sample item.
Efron and Tibisahrani (1993) described bootstrap confidence intervals. The basic idea is as follows:

Let $\hat{\theta}_{n}^{*}$ be a bootstrap estimate of $\theta$ based on a resample of size $n$ from the original data $X_{1}, \ldots, X_{n}$, and let $G^{*}$ be its distribution function given the observed value

$$
G^{*}(x)=\operatorname{prob}\left\{\widehat{\theta}_{n}^{*} \leq x \mid X_{1}=x_{1}, \ldots, X_{n}=x_{n}\right\} .
$$

The bootstrap percentiles method gives $G^{*-1}(\alpha)$ and $G^{*-1}(1-\alpha)$ as, respectively, lower and upper bounds for $1-2 \alpha$ confidence interval for $\widehat{\theta}_{n}$. In practice $G^{*-1}(\alpha)$ and $G^{*-1}(1-\alpha)$ are approximated by generating $B$ pseudo-sequences $\left(X_{1}, \ldots, X_{n}\right)$, calculating the corresponding values of $\widehat{\theta}_{n}^{*}(b)$ for $b=1, \ldots, B$, and then finding the empirical percentiles. The number of resamples of $B$ usually need to be quite large, in most cases it is recommended that $B \geqslant 1000$.

The percentile interval may not have the nominal coverage when the sampling distribution is skewed. Efron's (1982) first improvement of the percentile interval is called the bias-corrected (BC) percentile interval. This is based on observing that the estimator under consideration may not be median-unbiased. Efron's (1987) second improvement to the percentile interval is called the bias-corrected, accelerated interval, $\mathrm{BC}_{\alpha}$ interval. Computing the accelerated bootstrap confidence interval requires estimating a bias coefficient, $\mathrm{z}_{0}$, and an acceleration coefficient, $\alpha$. Both coefficients can be estimated nonparametrically from the data. Confidence interval endpoints are obtained by inverting percentiles of the bootstrap distribution. Adjusting for bias and acceleration shifts the percentiles used to find the confidence interval endpoints. Because endpoints of the confidence interval are
obtained by inverting the bootstrap distribution, both the percentiles and accelerated bootstraps preserve the range of the parameter. When the acceleration coefficient, $\alpha=0, \mathrm{BC}_{\alpha}$ interval reduces to the $B C$ interval.

For this dissertation, patients within each dose group are re-sampled with replacement, and the whole MCP-Mod procedure is repeated $B$ times, which generates $B$ bootstrap sample values of $M E D$ or $M S D$. Appropriate confidence interval for the $M E D$ or MSD is derived from the $B M E D$ or MSD bootstrap sample values.

### 2.2.5 Extension of the current MCP-Mod approach

The issues and potential work to extend the current MCP-Mod approach are listed as follows:

1. Dose selection based on safety dose-response models will be different from dose selection from the efficacy dose-response model due to the different dose-response profiles. In addition, the MCP-Mod approach tends to select the minimum efficacy dose which can provide treatment effect. It hasn't yet implemented a method for how to select the maximum safety dose under acceptable toxicity profile once the model is selected.
2. For some clinical trials when both efficacy and safety are needed to be combined for selecting appropriate dose for drug development, extended research is needed. The following cases will be considered: a). Both efficacy and safety responses are continuous; b). Efficacy response is continuous and safety response is discrete or vice versa.

The above extensions will be addressed in the remaining of this dissertation.

### 2.3 Dose-finding based on efficacy-toxicity response

### 2.3.1 Dose-finding based on efficacy-toxicity trade off by in Phase I cancer studies

Several research papers have proposed methods for dose-finding based on both efficacy and toxicity. Gooley et al. (1994) were perhaps the first to consider two doseoutcome curves. O'Quigley, Hughes, and Fenton (2001) proposed a two-stage dose-finding design, assuming continual reassessment method (CRM) models for efficacy and toxicity outcomes. An acceptable level of toxicity is determined in the first stage, starting with a low toxicity target that later may be increased, and a sequential probability ratio test is used in the second stage to compare null and alternative values of probability for efficacy and toxicity outcomes. Thall and Russell (TR, 1998) designed various dose-finding trials in oncology. Thall and Russell require efficacy (E) and toxicity (T) to be disjoint. The method in Thall and Cook (2004) differs from those earlier approaches in terms of both the underlying model and the dose-finding algorithm. They assume a more flexible model having more parameters than the models used in the previous approaches and provide an algorithm for establishing priors based on elicited mean outcome probabilities. The difference is that their dose-finding algorithm is based on explicit trade-offs between probability of efficacy and toxicity. They use both efficacy (E) and toxicity (T) to choose doses for successive cohorts of patients in early phase clinical trials. The details are not illustrated here. There are 3 components in the methodology: A bayesian model for joint probabilities of efficacy and toxicity outcomes; consisting of criteria for deciding acceptable high efficacy and low toxicity and several elicited (efficacy, toxicity) probability pairs that are equally
desirable pairs provided by physicians. Thall et al. (2008) further developed a bayesian sequential dose-finding procedure based on bivariate outcomes that accounts for patient covariates and dose-covariate interactions. Because the dose selection criteria are covariate specific, different patients may receive different doses at the same point in the trial, and the set of eligible patients may change adaptively during the trial.

### 2.3.2 Limitations of this trade-off methodology

Although this methodology accommodates the bivariate outcomes, it is more structured than most dose-finding methods and requires reliable, user-friendly computer programs for the simulations during the design process and trial conduct. The method proposed by Thall et al. (2008) is very complex, and it requires a substantial effort on both the statistician and the physicians planning the trial. Their approach requires binary endpoints for both efficacy and safety endpoints. Furthermore, due to the design algorithm, the patients or cohorts are enrolled sequentially. This is more limited to dose finding studies in phase I/II cancer trials. It is not easy to generalized to dose finding studies in other phase II/III trials.

## Chapter 3

## Extend MCP-Mod to Safety Outcome

### 3.1 Definition

The MCP-Mod from Bretz et al. (2005) can be extended to safety outcomes. The $M E D$ (minimum effective dose) is replaced by the MSD (maximum safety dose). The $M S D$ is defined as the maximum dose which shows clinically acceptable toxicity. Let $\Delta$ denote the clinically acceptable difference, that is, the largest safety acceptable difference, by which we expect a dose to be not too worse than placebo. For the purpose of dose estimation, the following model $g\left(d_{i}, \theta_{Z}\right)$ specification for the safety responses $Z_{i j}$ is considered:

$$
Z_{i j}=g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)+\epsilon_{i j}^{Z},
$$

where

$$
\epsilon_{i j}^{Z} \sim N\left(0, \Psi_{Z Z}\right),
$$

for the $j$ th patient in the $i$ th dose group where $i=1, \ldots, k$, and $j=1, \ldots, n_{i}$.

Two definitions of the MSD are possible, depending on whether the target dose is selected out of the discrete dose set $D=d_{1}, \ldots, d_{k}$ under investigation or from the entire dose range $\left(d_{1}, d_{k}\right]$. Model-based approaches allow $M S D \in\left(d_{1}, d_{k}\right]$. In the following sections we model the mean of the efficacy variable as $f\left(d_{i}, \boldsymbol{\theta}_{Y}\right)$ and the mean of the safety variable as $g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)$. Given a model $g\left(., \boldsymbol{\theta}_{Z}\right)$, define

$$
M S D=\underset{d \in\left(d_{1}, d_{k}\right]}{\operatorname{argmax}}\left\{g\left(d, \boldsymbol{\theta}_{Z}\right) \leq g\left(d_{1}, \boldsymbol{\theta}_{Z}\right)+\triangle\right\} .
$$

Following the above definition, two different rules are proposed to estimate the true MSD. Denote $U_{d}$ the upper $1-2 \gamma$ confidence limit of predicted mean value $p(d)=g\left(d, \widehat{\boldsymbol{\theta}}_{Z}\right)$ at dose $d$ based on the model $g\left(., \theta_{Z}\right)$ (Figure 3.1).

$$
\begin{aligned}
& \widehat{M S D_{1}}=\operatorname{argmax}_{d \in\left(d_{1}, d_{k}\right\}}\left\{U_{d} \leq g\left(d_{1}, \widehat{\boldsymbol{\theta}}_{Z}\right)+\Delta\right\}, \\
& \widehat{M S D_{2}}=\operatorname{argmax}_{d \in\left(d_{1}, d_{k}\right]}\left\{p(d) \leq g\left(d_{1}, \widehat{\boldsymbol{\theta}}_{Z}\right)+\Delta\right\},
\end{aligned}
$$

where $\widehat{\boldsymbol{\theta}}_{Z}$ can be obtained based on the model $g\left(., \boldsymbol{\theta}_{Z}\right)$ by the $n l$ function in R (Bates and Chambers, 1992).


Figure 3.1: MSD Estimation

### 3.2 Simulation

Select mean dose-response models $g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)$ for the safety responses $Z_{i j}$. For simulation purposes, we assumed decreased GFR from baseline follows an exponential model with mean $0.163+0.037 \exp (.5912$ dose $)$. The standard deviation for decreased GFR is 8 ml $\min -11.73 \mathrm{~m}^{-2}$, and the clinical relevance for $M S D$ is the decrease of GFR from baseline less than 5 ml min- $11.73 \mathrm{~m}^{-2}$. The placebo effect for mean change of decreased GFR from baseline is assumed as $0.2 \mathrm{ml} \mathrm{min}-11.73 \mathrm{~m}^{-2}$. The true $M S D$ from the safety mean model is 0.829 . The basis for the data simulation is from the motivating example of the clinical trial in Section 1.2.

$$
\text { Let } \Psi_{Z Z}=64, n=100 / \text { dose group and dose }=0,0.05,0.2,0.4,0.6,0.8 \text { and } 1 .
$$

The data are simulated 1000 times from the above safety models. The data are generated by the rnorm function in R (Becker et al., 1988).

### 3.3 MSD dose selection performance

### 3.3.1 MSD estimation plot

A graphical view of the dose-selection performance of the MSD estimators is given by the boxplots of estimated $\widehat{M S D}_{1}$ and $\widehat{M S D}_{2}$ (Figure 3.2).

## Boxplot of simulated MSD estimators



Figure 3.2: MSD boxplot

### 3.3.2 Bias and dispersion

The dose-selection performance based on the MSD is measured in terms of its proximity to the target dose (the dose producing an acceptable clinical safety difference over placebo and dispersion around the target dose). Mean and Median relative bias (\%) $R_{b}$ will be provided by the following formula:

$$
R_{b}=\frac{100 *(\widehat{M S D}-M S D)}{M S D} .
$$

where $M S D$ is the true value and $\widehat{M S D}$ is the estimated value.

## Dispersion will be calculated by interquartile range (IQR).

IQR is a measure of statistical dispersion. The median and IQR of $R_{b}$ characterize the relative bias and variability of the $M S D$.

As shown in Table 3.1, mean relative bias and median relative bias for both $\widehat{M S D_{1}}$ and $\widehat{M S D}_{2}$ are in a reasonable range. Mean and median $R_{b}$ for $\widehat{M S D}_{1}$ are $5.78 \%$ and $4.70 \%$ respectively. $\widehat{M S D}_{1}$ has slightly more bias and dispersion than $\widehat{M S D}_{2}$ but is still within reasonable range.

Table 3.1: Bias and dispersion of $\widehat{M S D}_{1}$ and $\widehat{M S D}_{2}$

|  | True <br> mean | Mean | Median | Mean <br> bias (\%) | Median <br> bias (\%) | Relative <br> IQR(\%) | MSE |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\overline{M S D}_{1}$ | 0.829 | 0.783 | 0.79 | $5.78 \%$ | $4.70 \%$ | $4.83 \%$ | 2.1198 |
| $\overline{M S D}_{2}$ | 0.829 | 0.826 | 0.83 | $2.94 \%$ | $0.12 \%$ | $3.62 \%$ | 0.0095 |

### 3.3.3 Coverage of bootstrap confidence interval for $M S D$

The precision of the estimated MSD dose can be assessed using the nonparametric bootstrap method. The patients within each dose group are re-sampled with replacement and the whole MCP-Mod procedure (described in Chapter 2) extended for MSD is repeated 100 times. The bootstrap confidence intervals are derived. The percentage of the bootstrap confidence intervals that cover the true MSD from the mean safety doseresponse model is calculated.

Table 3.2: Coverage of $95 \%$ bootstrap C.I. for true MSD

|  | Coverage of <br> $95 \%$ bootstrap <br> percentile C.I. | Coverage of 95\% |  |
| :--- | :--- | :--- | :--- |
| bootstrap BC C.I. | Coverage of 95\% <br> bootstrap BC $\alpha$ <br> C.I. |  |  |
| $\widehat{M S D}_{1}$ | $69 \%$ | $51 \%$ | $56 \%$ |
| $\widehat{M S D}_{2}$ | $97 \%$ | $86 \%$ | $89 \%$ |

Table 3.3: Comparison of lower bound of $95 \%$ bootstrap C.I. with true MSD

|  | Lower bound of <br> $95 \%$ bootstrap <br> percentile C.I. <br> $\leq$ True MSD | Lower bound of <br> $95 \%$ bootstrap BC <br> C.I. $\leq$ True MSD | Lower bound of <br> $95 \%$ bootstrap <br> BC <br> MSD. $\quad \leq$ True |
| :--- | :--- | :--- | :--- |
| $\widehat{M S D}_{1}$ | $100 \%$ | $90 \%$ | $95 \%$ |
| $\overline{M S D}_{2}$ | $100 \%$ | $89 \%$ | $92 \%$ |

As expected, the $\widehat{M S D_{1}} 95 \%$ bootstrap C.I has less coverage than the $\widehat{M S D_{2}}$, this is consistent with the result in Table 3.1 and Figure 3.2. In addition, Table 3.3 showed that the lower bound of $95 \%$ C.I bootstrap are smaller than the true $M S D(89 \%-100 \%) . \widehat{M S D}_{1}$ coverage $(69 \%)$ is reasonable and is more clinically appropriate to be used to estimate the
maximum safety dose as a conservative approach since we want to select a dose which has less chance to have toxicity.

## Chapter 4

## Joint Nonlinear Continuous Mixed

## Model Estimation

### 4.1 Joint model formulation and log-likelihood function

$$
\text { For a bivariate vector }\left[\begin{array}{l}
Y_{i j} \\
Z_{i j}
\end{array}\right] \text {, representing the efficacy variable } Y_{i j} \text { and the safety }
$$ variable $Z_{i j}$, the correlation between $Y_{i j}$ and $Z_{i j}$ is determined by the within-patient variancecovariance structure and may be different across treatment doses. Here, the within-patient errors are assumed to be heteroscedastic and correlated. This formulation results in a simplified version of the extended nonlinear joint regression model as follows:

$$
\left[\begin{array}{c}
Y_{i j}  \tag{4.1}\\
Z_{i j}
\end{array}\right]=\left[\begin{array}{l}
f\left(d_{i}, \boldsymbol{\theta}_{Y}\right) \\
g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)
\end{array}\right]+\left[\begin{array}{c}
\epsilon_{i j}^{Y} \\
\epsilon_{i j}^{Z}
\end{array}\right], \mathrm{i}=1, \ldots \mathrm{k}, \mathrm{j}=1, \ldots . \mathrm{n}_{i}
$$

where

$$
\boldsymbol{\epsilon}_{i j}=\left[\begin{array}{c}
\epsilon_{i j}^{Y} \\
\epsilon_{i j}^{Z}
\end{array}\right] \sim B V N\left(\left[\begin{array}{l}
0 \\
0
\end{array}\right],\left[\begin{array}{ll}
\Psi_{Y Y}^{i} & \Psi_{Y Z}^{i} \\
\Psi_{Y Z}^{i} & \Psi_{Z Z}^{i}
\end{array}\right]\right)
$$

and $\left[\begin{array}{c}\boldsymbol{\theta}^{Y} \\ \boldsymbol{\theta}^{Z}\end{array}\right]$ refers to the vector of model parameters, $i$ to the dose group ( $i=1$ corresponds to placebo), and $j$ to the patient within dose group $i$.

Further, the response variables $Y_{i j}$ or $Z_{i j}$ can be combined into a single variable $V_{i j}$ with the indicator variable $t$, i.e.,

$$
V_{i j t}=\left\{\begin{array}{l}
Y_{i j} \text { if } t=0 \\
Z_{i j} \text { if } t=1
\end{array}\right.
$$

then the model can be reformulated as

$$
\begin{equation*}
V_{i j t}=f\left(d_{i}, \boldsymbol{\theta}_{Y}\right)^{1-t} g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)^{t}+(1-t, t) \epsilon_{i j}, \tag{4.2}
\end{equation*}
$$

where

$$
\boldsymbol{\epsilon}_{i j} \sim N\left(0, \sigma^{2} \boldsymbol{\Lambda}_{i}\right)
$$

and

$$
\sigma^{2} \boldsymbol{\Lambda}_{i}=\left[\begin{array}{ll}
\Psi_{Y Y}^{i} & \Psi_{Y Z}^{i} \\
\Psi_{Y Z}^{i} & \Psi_{Z Z}^{i}
\end{array}\right]
$$

$\boldsymbol{\Lambda}_{i}$ is a positive-definite matrix. The $\boldsymbol{\Lambda}_{i}$ matrices are determined by fixed, generally a small set of parameters $\boldsymbol{\lambda}$. Let $\boldsymbol{\Lambda}_{i}=\left(\boldsymbol{\Lambda}_{i}^{1 / 2}\right)^{T} \boldsymbol{\Lambda}_{i}^{1 / 2}, \boldsymbol{\Lambda}_{i}^{-1}=\boldsymbol{\Lambda}_{i}^{-1 / 2}\left(\boldsymbol{\Lambda}_{i}^{-1 / 2}\right)^{T}, \boldsymbol{V}_{i j}=\left[\begin{array}{c}Y_{i j} \\ Z_{i j}\end{array}\right]$, $\boldsymbol{V}_{i j}^{*}=\left(\boldsymbol{\Lambda}_{i}^{-1 / 2}\right)^{T} \boldsymbol{V}_{i j}$,

$$
\binom{f^{*}\left(d_{i}, \boldsymbol{\theta}_{Y}\right)}{g^{*}\left(d_{i}, \boldsymbol{\theta}_{Z}\right)}=\left(\boldsymbol{\Lambda}_{i}^{1 / 2}\right)^{T}\binom{f\left(d_{i}, \boldsymbol{\theta}_{Y}\right)}{g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)}
$$

and

$$
\boldsymbol{\epsilon}_{i j}^{*}=\left(\boldsymbol{\Lambda}_{i}^{-1 / 2}\right)^{T} \boldsymbol{\epsilon}_{i j} .
$$

Estimation and inference under a single response nonlinear model have been studied in Pinheiro and Bates (2000). When the $\boldsymbol{\Lambda}_{i}$ matrices are known, this is referred to the generalized nonlinear least-squares (GNLS) model. The model 4.1 can be re-expressed as a "classic" nonlinear model:

$$
V_{i j t}^{*}=f^{*}\left(d_{i}, \boldsymbol{\theta}_{Y}\right)^{1-t} g^{*}\left(d_{i}, \boldsymbol{\theta}_{Z}\right)^{t}+(1-t, t) \boldsymbol{\epsilon}_{i j}^{*},
$$

where

$$
\boldsymbol{\epsilon}_{i j}^{*} \sim N\left(0, \sigma^{2} \boldsymbol{I}\right) .
$$

The log-likelihood function for the GNLS model in (4.2) can be written as:

$$
\begin{align*}
l\left(\boldsymbol{\theta}_{Y}, \boldsymbol{\theta}_{Z}, \boldsymbol{\lambda}, \sigma^{2} \mid \boldsymbol{V}\right)= & -\frac{1}{2} N \log (2 \pi) \\
& -\frac{1}{2}\left\{\sum_{i=1}^{k} \sum_{j=1}^{n_{i}}\left[\sum_{t=0}^{1} \frac{\left\|V_{i j t}^{*}-f^{*}\left(d_{i}, \boldsymbol{\theta}_{Y}\right)^{1-t} g^{*}\left(d_{i}, \boldsymbol{\theta}_{Z}\right)^{t}\right\|^{2}}{\sigma^{2}}+\log \left|\boldsymbol{\Lambda}_{i}\right|\right]\right\} \tag{4.3}
\end{align*}
$$

where $\mathrm{N}=\sum_{i=1}^{k} \mathrm{n}_{i}, \mathrm{~N}$ represents the total number of observations.
The estimation and inference method discussed in Pinheiro and Bates (2000) can be applied to find the MLE estimates based on the log-likelihood function in (4.3).

### 4.2 Estimation and computational methods

For fixed $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$, the maximum likelihood estimator of $\sigma^{2}$ for model in (4.2) is:

$$
\widehat{\sigma}^{2}\left(\boldsymbol{\theta}_{Y}, \boldsymbol{\theta}_{Z}, \boldsymbol{\lambda}\right)=\sum_{i=1}^{k} \sum_{j=1}^{n_{i}} \sum_{t=0}^{1}\left\|V_{i j t}^{*}-f^{*}\left(d_{i}, \boldsymbol{\theta}_{Y}\right)^{1-t} g^{*}\left(d_{i}, \boldsymbol{\theta}_{Z}\right)^{t}\right\|^{2} / N .
$$

thus by replacing $\sigma^{2}$ with $\widehat{\sigma}^{2}\left(\boldsymbol{\theta}_{Y}, \boldsymbol{\theta}_{Z}, \boldsymbol{\lambda}\right)$, the profile log-likelihood can be written as:

$$
\begin{aligned}
& l\left(\boldsymbol{\theta}_{Y}, \boldsymbol{\theta}_{Z}, \lambda \mid V\right)= \\
& -\frac{1}{2}\left\{N\left[\log \left(\frac{2 \pi}{N}\right)+1\right]+\log \left(\sum_{i=1}^{k} \sum_{j=1}^{n_{i}} \sum_{t=0}^{1}| | V_{i j t}^{*}-f^{*}\left(d_{i}, \boldsymbol{\theta}_{Y}\right)^{1-t} g^{*}\left(d_{i}, \boldsymbol{\theta}_{Z}\right)^{t}| |^{2}\right)+\sum_{i=1}^{k} \sum_{j=1}^{n_{i}} \log \left|\boldsymbol{\Lambda}_{i}\right|\right\} .
\end{aligned}
$$

Following the derivations in Pinheiro and Bates (2000), a Gauss-Seidel algorithm was used with the profile log-likelihood to obtain the maximum likelihood estimates of $\boldsymbol{\theta}_{Y}, \boldsymbol{\theta}_{Z}$ and $\lambda$ :

1) Given current estimate $\hat{\lambda}^{(m)}$ of $\boldsymbol{\lambda}$, a new estimate $\left(\widehat{\boldsymbol{\theta}}_{Y}^{m+1}, \widehat{\boldsymbol{\theta}}_{Z}^{m+1}\right)$ for $\left(\boldsymbol{\theta}_{Y}, \boldsymbol{\theta}_{Z}\right)$ is produced.
2) Given current estimate $\left(\widehat{\boldsymbol{\theta}}_{Y}^{m+1}, \widehat{\boldsymbol{\theta}}_{Z}^{m+1}\right)$, a new estimate $\widehat{\boldsymbol{\lambda}}^{(m+1)}$ is produced.
3) Repeat 1), 2) until a convergence criterion is met.

The asymptotic distributions of MLEs in the GNLS model, which are used for constructing confidence intervals and hypothesis tests, are:

$$
\begin{gathered}
\binom{\widehat{\boldsymbol{\theta}}_{Y}}{\widehat{\boldsymbol{\theta}}_{Z}} \sim N\left(\binom{\boldsymbol{\theta}_{Y}}{\boldsymbol{\theta}_{Z}}, \sigma^{2}\left[\sum_{i=1}^{k} \widehat{\boldsymbol{X}}_{i}^{T} \boldsymbol{\Lambda}_{i}^{-1} \widehat{\boldsymbol{X}}_{i}\right]^{-1}\right) \\
{\left[\begin{array}{c}
\hat{\lambda} \\
\log \widehat{\sigma}
\end{array}\right]}
\end{gathered}
$$

and

$$
I(\lambda, \sigma)=-\left[\begin{array}{cc}
\frac{\partial^{2} l}{\partial \lambda \partial \lambda^{T}} & \frac{\partial^{2} l}{\partial \log \sigma \partial \lambda^{T}} \\
\frac{\partial^{2} l}{\partial \lambda \partial \log \sigma} & \frac{\partial^{2} l}{\partial^{2} \log \sigma}
\end{array}\right]
$$

where $\widehat{\boldsymbol{X}_{i}}=\binom{\frac{\partial f\left(d_{i}, \theta_{Y}\right)}{\partial \theta_{Y}^{T}}}{\frac{\partial g\left(d_{i}, \theta_{Z}\right)}{\partial \theta_{Z}^{T}}}$ is the derivative matrix evaluated at the true parameter values for efficacy or safety model parameter values. $\log \sigma$ is used in place of $\sigma^{2}$ to give unrestricted parameterization for which the normal approximations tend to be more accurate.

As described in Pinheiro and Bates (2000), to reduce the bias associated with the maximum likelihood estimation of $\sigma^{2}$, the following modified version of $\widehat{\sigma}^{2}$ is used:

$$
\widehat{\sigma}^{2}=\frac{\sum_{i=1}^{k} \sum_{j=1}^{n_{i}}\left\|\left(\widehat{\boldsymbol{\Lambda}}_{i}^{-1 / 2}\right)^{T}\left[\boldsymbol{V}_{i j}^{*}-\binom{f^{*}\left(d_{i}, \boldsymbol{\theta}_{Y}\right)}{g^{*}\left(d_{i}, \boldsymbol{\theta}_{Z}\right)}\right]\right\|^{2}}{(N-p)}
$$

with $p$ denoting the length of $\boldsymbol{\theta},(N-p) \widehat{\sigma}^{2}$ is asymptotically distributed as a $\sigma^{2} \chi_{N-p}^{2}$ random variable and is asymptotically independent of $\widehat{\boldsymbol{\theta}}$.

### 4.2.1 Fitting joint model with gnls in R

The gnls (Pinheiro and Bates, 2000) function in R can be used to fit the extended nonlinear regression model using maximum likelihood. It can either be viewed as a special case of the nlme function (Pinheiro and Bates, 2000) without using random effects or as a version of the nls function with the arguments weights and correlation which can account for heteroscedastic and correlation within patient. The gnls (Pinheiro and Bates, 2000) function in $R$ can be used to fit the reformulated model in (4.2) which is equivalent to the model in (4.1). Efficacy or safety data for the same patient are correlated and variances
are usually unequal across dose group, the two components can be reflected in the withinpatient error variance structure by using the weight and correlation arguments in gnls (Pinheiro and Bates, 2000) in R.

### 4.2.2 Parameter estimates, bias and relative efficiency

The median, mean(std) from separate models and joint model are computed from 1000 simulations of data having bivariate efficacy and safety responses. The percent of bias is calculated as $100 \times$ (mean-truemean)/truemean. The smaller the standard error of a statistic, the more efficient it is. The relative efficiency of two statistics is the ratio of the squares of their standard errors. The parameter estimates for the joint model and separate models are compared by the percent of bias and relative efficiency.

In addition, estimates of the correlation between bivariate responses and the variances of the efficacy and safety responses from the gnls (Pinheiro and Bates, 2000) output in R are used to verify the correctness of simulation and goodness of the covariance matrix estimation of the model.

### 4.2.3 Simulation

Select mean dose-response models $f\left(d_{i}, \boldsymbol{\theta}_{Y}\right)$ and $g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)$ for the efficacy and safety responses $\left[\begin{array}{l}Y_{i j} \\ Z_{i j}\end{array}\right]$, assume $\rho$ is the correlation between efficacy and safety responses, i.e,

$$
\left[\begin{array}{c}
Y_{i j} \\
Z_{i j}
\end{array}\right]=\left[\begin{array}{c}
f\left(d_{i}, \boldsymbol{\theta}_{Y}\right) \\
g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)
\end{array}\right]+\left[\begin{array}{c}
\epsilon_{i j}^{Y} \\
\epsilon_{i j}^{Z}
\end{array}\right],
$$

where,

$$
\left[\begin{array}{c}
\epsilon_{i j}^{Y} \\
\epsilon_{i j}^{Z}
\end{array}\right] \sim \operatorname{BVN}\left(\left[\begin{array}{l}
0 \\
0
\end{array}\right],\left[\begin{array}{cc}
\Psi_{Y Y} & \rho \Psi_{Y} \Psi_{Z} \\
\rho \Psi_{Z} \Psi_{Y} & \Psi_{Z Z}
\end{array}\right]\right)
$$

for $j$ th patient in $i$ th dose group where $i=1, \ldots k$, and $j=1, \ldots . n_{i}$, and $\Psi_{Y}=\sqrt{\Psi_{Y Y}} \& \Psi_{Z}=$ $\sqrt{\Psi_{Z Z}}$.

For the simulation purpose, we assumed the decreased diastolic blood pressure (DBP) from baseline follows the Emax model $f\left(d_{i}, \boldsymbol{\theta}_{Y}\right)=2.5+14.5 *$ dose $/(.2+$ dose $)$; and the decreased GFR from baseline follows the exponential model $g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)=0.163+0.037 *$ $\exp (3.3 *$ dose $* \log (6))$. The standard deviation for decreased diastolic blood pressure and GFR are 7 mmHg and $8 \mathrm{ml} \mathrm{min}-1.73 \mathrm{~m}^{-2}$, the clinical relevance for $M E D$ is more than 3 mmHg of the difference in decreased DBP from baseline between treatment and placebo group and for $M S D$ is less than 5 ml min- $11.73 \mathrm{~m}^{-2}$ of the difference in the decrease of GFR from baseline between treatment and placebo group respectively. The placebo effect for the mean change of decreased DBP from baseline is usually around 2.5 mmHg and the mean change of decreased GFR from baseline is $0.2 \mathrm{ml} / \mathrm{min} / 1.73 \mathrm{~m}^{2}$. The true $M E D$ from the efficacy mean model is 0.056 and true $M S D$ from the safety mean model is 0.829 . Though the values for $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ are well within the range of the dose selection and is a good window for the example we selected. But the final dose to be selected for the Phase III will rely on the joint criteria that both efficacy and safety can be satisfied. This will be discussed in Chapter 5.

Let $\Psi_{Y Y}=49, \Psi_{Z Z}=64, n=100 /$ dose group and dose $=0,0.05,0.2,0.4,0.6$, 0.8 and 1. The data are simulated 1000 times each for $\rho=0,0.4,0.8$, respectively from the above joint efficacy and safety models. The data are generated using the rmvnorm function in the mvtnorm library (Hothorn et al., 2001) in R.

After the data are simulated, models are fitted separately and jointly with Emax model for efficacy response and Exponential model for safety response. Joint model fitting is performed by gnls (Pinheiro and Bates, 2000) function in $R$ described in Section 4.2.1. The mean (Std), median, bias, percent of bias, MSE and relative efficiency are computed.

### 4.2.4 Simulation results for parameter estimates

Parameter estimates for Emax and exponential models either from the separate or the joint fitting are computed based on the simulated data. When the efficacy and safety responses are independent, the parameter estimates of Emax and exponential models are similar between separate and joint model fitting. This can be seen from the mean bias, percent of bias, MSE and relative efficiency. When data are correlated ( $\rho=0.4$, or 0.8 ), the mean bias, percent of bias, MSE are smaller for joint fitting than for separate fitting (Figure 4.2 to Figure 4.4). In addition, the efficiencies increase when the correlation between the efficacy and safety responses increases (Table 4.1 and Figure 4.1). As seen in Table 4.1 and Figure 4.1, efficiencies increase from 0.98-1.0 to 1.2-3.8 for all parameter estimates when the correlation increases from 0 to 0.8 .
Table 4.1: Parameter estimates based on separate and joint modeling -Emax for efficacy and Exponential for safety

| Model correlation |  | Parameter | True value | Separate Fitting |  |  |  | Joint Fitting |  |  |  | RE.* <br> Joint <br> vs. <br> Sep. |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| True | Simu. |  |  | Median | Mean (Std) | Bias(\%) | MSE | Median | Mean (Std) | Bias(\%) | MSE |  |
| 0 | -0.001 | e0 | 2.5 | 2.420 | 2.409 (0.9234) | -0.09 (-3.7) | 1.720 | 2.340 | 2.388 (0.9208) | -0.11 (-4.49) | 1.396 | 1.006 |
|  |  | emax | 14.5 | 14.850 | 15.081 (2.1319) | 0.58 (4.0) | 9.756 | 14.586 | 15.126 (2.424) | 0.63 (4.3) | 8.081 | 0.990 |
|  |  | ed50 | 0.2 | 0.202 | 0.231 (0.1415) | 0.03 (15.3) | 0.042 | 0.203 | 0.231 (0.1464) | 0.03 (15.6) | 0.036 | 0.934 |
|  |  | e1 | 0.037 | 0.072 | 0.180 (0.3303) | 0.14 (386.1) | 0.212 | 0.072 | 0.182 (0.3358) | 0.15 (391.6) | 0.217 | 0.967 |
|  |  | es0 | 0.163 | -0.049 | -0.128 (0.8488) | -0.29 (-178.5) | 1.321 | -0.052 | -0.128 (0.8601) | -0.29 (-178.6) | 1.337 | 0.974 |
|  |  | delta | 0.169 | 0.191 | 0.199 (0.0741) | 0.03 (17.4) | 0.010 | 0.199 | 0.199 (.0746) | 0.03 (17.6) | 0.010 | 0.987 |
| 0.4 | 0.397 | e0 | 2.5 | 2.450 | 2.411 (0.9176) | -0.09 (-3.5) | 1.698 | 2.320 | 2.340 (0.8641) | -0.15 (-6.4) | 1.260 | 1.127 |
|  |  | emax | 14.5 | 14.934 | 15.064 (2.1493) | 0.56 (3.9) | 9.866 | 14.984 | 15.132 (2.000) | 0.63 (4.4) | 7.182 | 1.155 |
|  |  | ed50 | 0.2 | 0.203 | 0.230 (0.1387) | 0.03 (14.9) | 0.040 | 0.197 | 0.223 (0.1245) | 0.03 (11.4) | 0.026 | 1.241 |
|  |  | e1 | 0.037 | 0.073 | 0.183 (0.3267) | 0.15 (393.4) | 0.212 | 0.066 | 0.165 (0.2999) | 0.13 (346.2) | 0.174 | 1.187 |
|  |  | es0 | 0.163 | -0.054 | -0.135 (0.8448) | -0.30 (-182.8) | 1.327 | -0.043 | -0.113 (0.8339) | -0.27 (-169.4) | 1.260 | 1.026 |
|  |  | delta | 0.169 | 0.190 | 0.199 (0.0746) | 0.03 (17.6) | 0.011 | 0.187 | 0.195 (.0699) | 0.03 (15.5) | 0.009 | 1.141 |
| 0.8 | 0.797 | e0 | 2.5 | 2.477 | 2.451 (0.9381) | -0.05 (-2.0) | 1.759 | 2.424 | 2.405 (0.7238) | -0.09 (-3.8) | 0.940 | 1.680 |
|  |  | emax | 14.5 | 14.748 | 14.959 (2.0296) | 0.46 (3.2) | 8.635 | 14.757 | 14.936 (1.6763) | 0.44 (3.0) | 5.292 | 1.466 |
|  |  | ed50 | 0.2 | 0.200 | 0.231 (0.1332) | 0.03 (15.3) | 0.037 | 0.204 | 0.215 (0.0702) | 0.01 (7.3) | 0.009 | 3.604 |
|  |  | e1 | 0.037 | 0.066 | 0.170 (0.2742) | 0.15 (358.8) | 0.156 | 0.055 | 0.105 (0.1406) | 0.07 (182.6) | 0.043 | 3.801 |
|  |  | es0 | 0.163 | -0.039 | -0.119 (0.7943) | -0.28(-172.8) | 1.195 | -0.044 | -0.013 (0.7239) | -0.17 (-107.8) | 0.979 | 1.204 |
|  |  | delta | 0.169 | 0.188 | 0.196 (0.0725) | 0.03 (15.8) | 0.010 | 0.181 | 0.181 (0.0536) | 0.01 (6.9) | 0.005 | 1.833 |



Figure 4.1: Efficiency of parameter estimates based on separate and joint modeling -Emax for efficacy and Exponential for safety


Figure 4.2: Percent bias of efficacy parameter estimates based on separate and joint modeling -Emax for efficacy and Exponential for safety


Figure 4.3: Percent bias of safety parameter estimates based on separate and joint modeling -Emax for efficacy and Exponential for safety


Figure 4.4: MSE of parameter estimates based on separate and joint modeling -Emax for efficacy and Exponential for safety

### 4.2.5 Procedures/Strategies of finding $M E D$ and $M S D$ for combined efficacy and safety data

In clinical trials, efficacy is the key factor for the approval of whether to proceed with the development of the drug. If a drug does not have the PoC for efficacy, then there is no continued assessment for the safety endpoint. The significance level alpha for safety dose-response testing can be relaxed. The following figure provides the testing sequence when both efficacy and safety endpoints need to be considered for the drug development.


Figure 4.5: Testing procedures of PoC for efficacy and safety endpoints

Denote the discrete dose set $D=d_{1}, \ldots, d_{k}$ under investigation or the entire dose range ( $d_{1}, d_{k}$ ]. Let $d_{1}<d_{2}<\ldots<d_{k}$. As shown in Figure 4.5, the first step is to confirm whether PoC exists for the efficacy (alpha=0.05). If there is no PoC established for efficacy, which indicates no dose-response relationship exists for efficacy or $\widehat{M E D} 2>d_{k}$, then there will be no dose-finding continued for this drug, i.e., the drug will not be carried to Phase III development. When the PoC of efficacy is established and also $\widehat{M E D} 2 \leq d_{k}$ is satisfied, the second step is to test PoC of the safety response (alpha=0.2). Only the efficacy response is studied further to identify the MED for the Phase III program when there is no established PoC for the safety response or $\widehat{M S D}_{1} \leq \widehat{M E D}_{2}$. This means that the dose-response curve for safety is flat and no dose-response relationship is present for safety, or when $\widehat{M S D}_{1}$ is not higher than $\widehat{M E D}_{2}$, then we only need to focus on the dose finding for efficacy. On the other hand, joint modeling for efficacy and safety responses are performed when both PoC for efficacy and safety responses exist and $\widehat{M S D_{1}}>\widehat{M E D}_{2}$.

### 4.2.6 Strategy for separate and joint model fitting

## Separate model fitting

The next paragraph describe how to estimate the $M E D$ and $M S D$ from separate model fitting by ignoring the correlation:

1. The following set of candidate efficacy models are chosen for fitting the data by MCPMod: Linlog, Emax, exponential, quadratic; While the Linglog, linear, Emax and exponential models are selected for fitting safety data.
2. Fit the efficacy data separately to choose the best model with alpha= 0.05 and lowest

AIC. If PoC of efficacy is established, fit safety data separately and choose best model with alpha $=0.2$ and lowest AIC.
3. Get the $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ based on the models from Step 2.

## Joint model fitting

The following two strategies can be used to obtain the $M E D$ and $M S D$ from the joint model fitting after PoC for both efficacy and safety are established:


Figure 4.6: Flow chart of joint model fitting-Strategy I

Strategy I: Keep most significant model from separate efficacy and safety model fitting

1. First get all the significant efficacy and safety models, of which $t$-statistics are bigger than the critical values $\left(q^{Y}, q^{Z}\right)$ while controlling FWER from separate fittings. Alpha $=0.05$ is pre-specifed for PoC of efficacy and alpha $=0.2$ is pre-specifed for PoC of safety. Then select the most significant efficacy and safety models based on the lowest AIC criteria, keep the parameter estimates which will be used as start values for
joint model fitting. Please note if there is no significant model for efficacy endpoint, this indicates there is no PoC for efficacy, then we will not proceed further. If there is no significant model for safety endpoint, which means there is no PoC for safety, then we can fit the model for efficacy only.
2. Joint model fitting by generalized nonlinear least squares model.
3. Estimate $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ based on updated efficacy and safety model from joint fitting (Figure 4.6).

## Strategy II: Keep all significant models from separate efficacy and safety models

1. Keep all the significant efficacy and safety models of which $t$-statistics are bigger than the critical values $\left(q^{Y}, q^{Z}\right)$ while controlling FWER from separate fittings. Alpha= $=0.05$ is pre-specified for PoC of efficacy and alpha=0.2 is pre-specifed for PoC of safety. The parameter estimates are kept and will be used as initial values for the joint model fitting.
2. Joint model fitting for all the combinations of efficacy and safety models selected from separate model fitting in Step 1, choose the best combination based on the lowest AIC from all the joint models fitting.
3. Obtain the estimated $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ based on the updated efficacy and safety models from the best joint model (Figure 4.7).


Figure 4.7: Flow chart of joint model fitting-Strategy II
4.2.7 Simulation results for $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ from separate and joint model fitting

The data are simulated as described in Section 4.2.3 and the joint model fitting is conducted as described in Figures 4.8 and 4.9.

Set of efficacy Candidate models
For example: Emax,Linlog Exponential, Quardratic


For example: Emax:
$E_{0}^{e}+E_{\text {max }}^{e} d /\left(E D_{50}^{e}+d\right)$

## Set of safety Candidate models

For example: Emax,Linear


> Most significant safety model (AIC)

For example: Linear:
$E_{0}^{S}+\delta^{s} * d$
Fit joint model: Emax and Linear

Final joint model:
(2-type) $\left(E_{0}^{e}+E_{\text {max }}^{e} d /\left(E D_{50}^{e}+d\right)\right)$ $+(\text { type }-1)^{*}\left(E_{0}^{S}+\delta^{s} * d\right)$
type: 1=efficacy; $2=$ safety

Figure 4.8: Flow chart of the joint model fitting-Strategy I Example


For example: Emax,Linlog Exponential, Quardratic


For example: Emax:
$\left.E_{0}^{e}+E_{\max }^{e} d\right\rangle\left(E D_{50}^{e}+d\right)$
Quadratic:
$E_{0}^{e}+\beta_{1}^{e} * d+\beta_{2}^{e} * d^{2}$

Set of safety
Candidate models
For example: Emax,Linear
Exponential, Linlog


All significant safety models (AIC)

For example: Exponential $E_{0}^{s}+E_{1}^{s} * \exp \left(d / \delta^{s}\right)$

## Linear:

$E_{0}^{s}+\delta^{s} * d$


Fit all combinations of joint models
Best joint model: Emax and exponential by AIC
For example: (2-type) $\left.E_{0}^{e}+E_{\text {max }}^{e} d /\left(E D_{50}^{e}+d\right)\right)$
(type-1) $\left(E_{0}^{S}+\delta^{s} * d\right)$
type: 1=efficacy; 2=safety

Figure 4.9: Flow chart of the joint model fitting-Strategy II example

Table 4.2 shows the results of $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ from joint and separate fitting when the efficacy and safety responses are independent. Percent of bias and MSE for $\widehat{M E D}_{2}$ are similar between separate model fitting and joint model fitting methods described in Figures 4.8 and 4.9; MSE and percent bias for $\widehat{M S D_{1}}$ are slightly better for the joint model fitting (using the approach in Figure 4.8) than separate fitting. MSE is decreased from . 00323 to .00183 (Table 4.2).

When the correlation between efficacy and safety increases, percent of bias and MSE decrease more for both approaches of joint fitting than separate fitting in $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ estimation (Tables 4.2-4.4). The $\widehat{M E D}_{2}$ is better estimated when the data are correlated, $\widehat{M S D}_{1}$ estimates remain the similar. In general joint fitting (2) with the Strategy II approach is better than joint fitting (1) with the Strategy I approach in terms of percent of bias and MSE.

Table 4.5 shows the consistency of correlation and variance structure for both the efficacy and safety responses.

Table 4.2: The $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ from separate fitting and joint fitting with $\rho=0$

| Efficacy <br> Model | $\widehat{M E D}_{2}$ |  |  |  |  |  | $\begin{aligned} & \text { (\%) } \\ & \text { bias } \end{aligned}$ | MSE |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | True MED | Fitting | Mean | sd | Median | bias |  |  |
| Emax | 0.052 | Separate | 0.062 | 0.019 | 0.060 | 0.011 | 20.8 | 0.00047 |
|  |  | Joint (1) | 0.062 | 0.019 | 0.060 | 0.010 | 19.7 | 0.00045 |
|  |  | Joint (2) | 0.063 | 0.020 | 0.060 | 0.011 | 21.2 | 0.00050 |
|  |  |  |  |  |  |  |  |  |
| Safety Model | $\widehat{M S D}_{1}$ |  |  |  |  |  | $\begin{aligned} & (\%) \\ & \text { bias } \end{aligned}$ | MSE |
|  | True MSD | Fitting | Mean | sd | Median | bias |  |  |
| Exp. | 0.829 | Separate | 0.783 | 0.033 | 0.790 | -0.040 | -5.6 | 0.00323 |
|  |  | Joint (1) | 0.803 | 0.035 | 0.810 | -0.026 | -3.2 | 0.00183 |
|  |  | Joint (2) | 0.798 | 0.055 | 0.810 | -0.037 | -3.7 | 0.00401 |

Notes: a. Simulated correlation is 0 . Outputs are from 500 simulations and $n=100$ per dose group.
b. Joint (1) for $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ are based on the joint model of most significant model from separate fitting of efficacy and safety data by lowest AIC criteria. The results consist of 460 emax-exponential, and 29 quardratic-exponential final joint models.
c. Joint (2) for $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ are based on all joint models by combinations of significantmodels from separate fitting of efficacy and safety data, the final model is selected from all joint models by lowest AIC criteria. The results consist of 460 emaxexponential, 34 quadratic-exponential, 5 emax-linear and 1 quardratic-linear final joint models.

Table 4.3: The $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ from separate fitting and joint fitting with $\rho=0.4$

| Efficacy <br> Model | $\widehat{M E D}_{2}$ |  |  |  |  |  | $\begin{aligned} & \hline \text { (\%) } \\ & \text { bias } \end{aligned}$ | MSE |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | True MED | Fitting | Mean | sd | Median | bias |  |  |
| Emax | 0.052 | Separate | 0.063 | 0.020 | 0.060 | 0.011 | 21.1 | 0.00054 |
|  |  | Joint (1) | 0.063 | 0.019 | 0.060 | 0.010 | 19.9 | 0.00047 |
|  |  | Joint (2) | 0.063 | 0.019 | 0.060 | 0.010 | 20.2 | 0.00049 |
| Safety Model | $\widehat{M S D}_{1}$ |  |  |  |  |  | (\%) |  |
|  | True MSD | Fitting | Mean | sd | Median | bias | bias | MSE |
| Exp. | 0.829 | Separate | 0.782 | 0.034 | 0.790 | -0.047 | -5.6 | 0.0033 |
|  |  | Joint (1) | 0.801 | 0.035 | 0.810 | -0.028 | -3.4 | 0.0019 |
|  |  | Joint (2) | 0.801 | 0.035 | 0.800 | -0.028 | -3.4 | 0.0020 |

Notes: a. Simulated correlation is 0 . Outputs are from 500 simulations and $\mathrm{n}=100$ per dose group.
b. Joint (1) for $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ are based on the joint model of most significant model from separate fitting of efficacy and safety data by lowest AIC criteria. The results consist of 453 emax-exponential, and 37 quardratic-exponential and 6 of linlog-exponential final joint models.
c. Joint (2) for $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ are based on all joint models by combinations of significantmodels from separate fitting of efficacy and safety data, the final model is selected from all joint models by lowest AIC criteria. The results consist of 464 emaxexponential, 24 quadratic-exponential and 12 linlog-exponential final joint models.

Table 4.4: The $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ from separate fitting and joint fitting with $\rho=0.8$

| Efficacy Model | $\widehat{M E D}_{2}$ |  |  |  |  |  | $\begin{aligned} & \hline \text { (\%) } \\ & \text { bias } \end{aligned}$ | MSE |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | True MED | Fitting | Mean | sd | Median | bias |  |  |
| Emax | 0.052 | Separate | 0.063 | 0.020 | 0.060 | 0.011 | 20.8 | 0.00052 |
|  |  | Joint (1) | 0.061 | 0.016 | 0.060 | 0.009 | 17.8 | 0.00032 |
|  |  | Joint (2) | 0.060 | 0.014 | 0.060 | 0.008 | 16.1 | 0.00028 |
|  |  |  |  |  |  |  |  |  |
| Safety <br> Model | $\widehat{M S D}_{1}$ |  |  |  |  |  | (\%) <br> bias | MSE |
|  | True MSD | Fitting | Mean | sd | Median | bias |  |  |
| Exp. | 0.829 | Separate | 0.782 | 0.034 | 0.780 | -0.047 | -5.7 | 0.0034 |
|  |  | Joint (1) | 0.799 | 0.033 | 0.800 | -0.029 | -3.6 | 0.0019 |
|  |  | Joint (2) | 0.800 | 0.031 | 0.800 | -0.029 | -3.4 | 0.0019 |

Notes: a. Simulated correlation is 0 . Outputs are from 500 simulations and $\mathrm{n}=100$ per dose group.
b. Joint (1) for $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ are based on the joint model of most significant model from separate fitting of efficacy and safety data by lowest AIC criteria. The results consist of 448 emax-exponential, and 35 quardratic-exponential final joint models.
c. Joint (2) for $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ are based on all joint models by combinations of significantmodels from separate fitting of efficacy and safety data, the final model is selected from all joint models by lowest AIC criteria. The results consist of 476 emaxexponential, 24 quadratic-exponential final joint models.

Table 4.5: The correlation and variance by simulation and gnls output

| Correlation and variance between efficacy and safety data |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| True values |  |  | gnls output |  |  |
| $\rho$ | $\Psi_{Y Y}$ | $\Psi_{Z Z}$ | $\widehat{\rho}$ | $\widehat{\Psi}_{Y Y}$ | $\widehat{\Psi}_{Z Z}$ |
| 0 | 49 | 64 | -0.001 | 49.13 | 64.13 |
| 0.4 | 49 | 64 | 0.399 | 49.13 | 63.96 |
| 0.8 | 49 | 64 | 0.799 | 49.14 | 63.84 |

## Evaluate bias and precision of the target dose estimate

The $\widehat{M E D}$ and $\widehat{M S D}$ from the joint or separate fitting are compared with the truemean from the selected dose-response models, the mean/median bias, percent bias, bootstrap $95 \%$ confidence interval coverage for $M E D$ and $M S D$ are computed as described in Sections 2.2.1 and 3.1. Also percent of

- $\widehat{M E D} \leqslant \widehat{M S D}$
- true $M E D \leqslant \widehat{M E D} \leqslant$ true $M S D$
- $\widehat{M E D}>$ true $M S D$ and
- $0<\widehat{M E D} \leqslant$ true $M E D$
are calculated.
Table 4.6 displays the outputs of the relationship among $\widehat{M E D}_{2}, \widehat{M S D}_{1}$, true values of efficacy and safety means. In general it shows the consistency of the estimated $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ for both the joint model and separate model fitting. All the estimated
$\widehat{M E D}_{2}$ are lower than the estimated $\widehat{M S D}_{1}$ value for this simulated example, $60 \%$ of $\widehat{M E D}_{2}$ are lower than True safety mean and higher than true efficacy mean.


Figures 4.10-4.11 show the distributions of $\widehat{M E D}_{2}$ and $\widehat{M E D}_{2}$ relative to the true $M E D$; distributions of $\widehat{M S D_{1}}, \widehat{M S D}_{2}$ and the true $M S D$. When the data are more correlated, the joint model fitting seems less spread, and there are higher proportions of $\widehat{M S D_{2}}$ above the true $M S D$ mean than that of $\widehat{M S D}_{1}$.


Figure 4.10: Plots of $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ based on separate or joint fitting: horizontal and vertical line represent true MED and true MSD, respectively.


Figure 4.11: Plots of $\widehat{M E D}_{2}$ and $\widehat{M S D}_{2}$ based on separate or joint fitting: horizontal and vertical line represent true MED and true MSD, respectively.

## Chapter 5

## Suggest Dose(s) for the Phase III

## Program Development for Joint

## Bivariate Continuous Responses

After the $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ are estimated, the optimal dose(s) to be carried into the Phase III program remain unresolved. If the $\widehat{M S D}_{1}$ is smaller than the $\widehat{M E D_{2}}$, there will be no appropriate dose for the Phase III program. In the following sections, we propose two different methods to select an optimal dose or a dose range for Phase III. The first method will focus on the joint success criteria for the efficacy and safety in Phase III; The second method will use the utility function to identify the final dose(s) to carry into Phase III program.

# 5.1 Method I: Identify the dose(s) through the joint criteria of continuous efficacy and safety responses for Phase III pro- <br> gram 

### 5.1.1 Method

The recommended dose(s) will be determined by

$$
\underset{d \in[M E D, M S D]}{\operatorname{argmax}} P(Y>a, Z<b \mid d) \geq c \text {, or }
$$

a dose range in $[M E D, M S D]$ such that $P(Y>a, Z<b \mid d) \geq c$,
where the joint density is

$$
f(y, z)=\frac{\exp \left(-\frac{1}{2\left(1-\rho^{2}\right)}\left(\frac{\left(y-\mu_{y}\right)^{2}}{\sigma_{y}^{2}}+\frac{\left(z-\mu_{z}\right)^{2}}{\sigma_{Z}^{2}}-\frac{2 \rho\left(y-\mu_{\gamma}\right)\left(z-\mu_{z}\right)}{\sigma_{\gamma} \sigma_{Z}}\right)\right)}{2 \pi \sigma_{y} \sigma_{z} \sqrt{1-\rho^{2}}}
$$

Y is the efficacy variable and Z is the safety variable, $a$ and $b$ are the criteria for Phase III success and $c$ is the success probability. The joint bivariate density is estimated from the joint fitted model with estimated values of $\sigma_{y}, \sigma_{z}, \rho, \mu_{y}, \mu_{z}$ from the joint modeling fitting in 4.2.4.

### 5.1.2 Simulation

## Procedure:

1. $\sigma_{y}, \sigma_{z}, \rho$ are from the previous joint model fitting (These values will be retained from the previous described Strategy II, which select the most significant joint model from all combinations). $\mu_{y^{\prime}} \mu_{z}$ will be derived from the different dose levels and the
mean models for efficacy and safety through the simulation results based on the joint model fitting.
2. Criteria values $a$ and $b$ are given such that the decreased DBP change from baseline is more than certain value (for example, 3 mmHg ) and decreased GFR change from baseline is less than certain value (for example, 6 ml min- $11.73 \mathrm{~m}^{-2}$ ) in order to claim the success of the Phase III program. $d_{e}$, and $d_{s}$ are the doses which fit the mean efficacy and safety model to satisfy $a$ and $b$ respectively.
3. $d_{e}$ and $d_{s}$ are 0.056 and 0.824 respectively, which are derived from the updated mean efficacy and safety model from the joint fitting to satisfy the mean efficacy change of $a$ and mean safety change of $b$.

Results Figures 5.1 and 5.2 plot the joint probability for efficacy response (change from baseline) $>3$ and safety response (change from baseline) $<6$. The simulated parameter estimates for efficacy and safety response models are based on the Emax and exponential model simulation results respectively. There are 4 lines in each plot. The upper and lower lines represent upper $97.5 \%$ and lower $2.5 \%$ of 500 values of the probabilities based on the simulated parameter estimates. The middle 2 lines are the true probability based on the true parameter value and mean of simulated probability, respectively. The simulated probability is very close to the true probability, which results in the similar final best dose satisfying the joint efficacy and safety criteria. The best dose that satisfies the joint criteria based on the true parameter value is 0.47 while the best simulated dose is 0.45 . As shown in Table 5.1 and Figure 5.2, the set of doses with probability $>0.6$ of satisfying the
joint criteria are 0.21-0.71 from true parameter and 0.21-0.69 from the simulated parameter estimates.

Eff. response> 3


Safety response $<6$


Eff. response> 3 and Safety response $<6$


Figure 5.1: Plot of probability to satisfy efficacy and safety criteria

## Probability that satisfy efficacy and safety criteria $>=0.6$



Figure 5.2: The set of doses satisfy the joint criteria with at least $60 \%$ probability or the best dose which has the maximum probability of success

Table 5.1: The best set of dose(s) satisfy joint efficacy and safety criteria

|  | Minimum dose | Maximum dose | Best dose (prob) |
| :--- | :--- | :--- | :--- |
| True | 0.21 | 0.71 | $0.47(66.03 \%)$ |
| Simulated | 0.21 | 0.69 | $0.45(65.65 \%)$ |
| Note: The correlation between efficacy and safety response is 0.8. Best dose provide <br> maximum probability to satisfy joint criteria while the minimum and maximum dose <br> can provide the probability to satisfy the joint criteria at least $60 \%$. |  |  |  |

### 5.2 Method II: Utility function based on trade off of continuous efficacy and safety responses for Phase III.

### 5.2.1 Methods

Another method to determine the recommended dose is based on the following utility function.

$$
F(d)=e f f(d)-k * \operatorname{saf}(d) .
$$

The dose is determined by maximizing the utility function $\mathrm{F}(\mathrm{d})$, i.e., $\underset{d \in[M E D, M S D]}{\operatorname{argmax}} F(d)$.
Here, $k>0$ is some weight for the discounted safety from efficacy; $\operatorname{eff}(d)$ and $\operatorname{saf}(d)$ can represent

1) $P(Y>a \mid d)$ and $P(Z>b \mid d)$ respectively, with estimated marginal density function derived from either the separate model fitting or the joint model fitting. Y and Z are efficacy and safety variables, $a$ and $b$ are criteria for Phase III success.
2) Standardized response eff $(d)=\widehat{Y}(d)$ and $\operatorname{saf}(d)=\widehat{Z}(d) \cdot \widehat{Y}(d)=\frac{E(Y(d))}{\sqrt{\operatorname{Var}(Y(d))}}=$ $\frac{f\left(d, \theta_{Y}\right)}{\sigma_{Y}}$ and $\widehat{Z}(d)=\frac{E(Z(d))}{\sqrt{\operatorname{Var}(Z(d))}}=\frac{g\left(d, \theta_{Z}\right)}{\sigma_{Z}} . \theta_{Y}, \theta_{Z}, \sigma_{Y}$, and $\sigma_{Z}$ are calculated from the estimated efficacy and safety models with the parameter estimates based on the joint or the separate
efficacy model fitting.

### 5.2.2 Simulation

## Procedure:

1. $\sigma_{y}, \sigma_{z}$ are obtained from the separate or marginal of joint model fitting. $\mu_{y}, \mu_{z}$ will be derived based on the different dose levels and the mean models for efficacy and safety from the simulation results by the separate or marginal model fitting.
2. Criteria $a$ and $b$ are given such as the decreased DBP change from baseline by more than a certain value(for example, 3 mmHg ) and decreased GFR change from baseline of less than certain value(For example, 6 ml min- $11.73 \mathrm{~m}^{-2}$ ) in order to claim the success of the Phase III program. $d_{e}$, and $d_{s}$ are the doses which fit the mean efficacy and safety model to satisfy $a$ and $b$ respectively.
3. $d_{e}$ and $d_{s}$ are 0.056 and 0.824 respectively, which are derived from the updated mean efficacy and safety models from the separate or marginal of joint model fitting to satisfy the mean efficacy change of $a$ and mean safety change of $b$.
4. Obtain the utility index based on separate probabilities of efficacy and safety response or the utility index based on the standardized efficacy and safety responses.

Results Figure 5.3 plots the utility index by probability of efficacy response (change from baseline) $>3$, safety response (change from baseline) $<6$ and some discounted $(k)$ value for the safety response. The simulated parameter estimates for efficacy and safety response models are based on the Emax and exponential model simulation results respectively.

There are 4 lines in each plot. The upper and lower lines represent upper $97.5 \%$ and lower $2.5 \%$ of 500 probabilities based simulated parameter estimates. The middle 2 lines are the true probability based on the true parameter value and mean simulated probability, respectively. The simulated probability is very close to true probability which result in the similar final best dose satisfy the joint efficacy and safety criteria. As seen in Figure 5.3 and Table 5.2, when $k$ increases from 0.2 to 0.8 , which means the safety has more discounted weight from the efficacy response, the maximum utility index increases from 1.074 to 1.508 and the best dose to satisfy this criteria decreases from 0.63 to 0.48 .

Similarly, the simulated best dose for standardized efficacy and safety response is very similar to the true dose based on the true parameter value (Figure 5.4). Furthermore, when $k$ increases from 0.2 to 0.8 , which means the safety has more discounted weight from the efficacy response, the maximum utility index decreases from 1.919 to 1.757 and the best dose to satisfy this criteria decreases from 0.76 to 0.58 (Table 5.2).

Based on probability: Efficacy>3 and Safety < 6


Figure 5.3: Best dose with the maximum utility index by varying the weight of safety response

Standardized Efficacy and Safety Eesponses


Figure 5.4: Best dose with the maximum utility index by varying the weight of standardized safety response

Table 5.2: The best set of dose satisfy different weight of efficacy and safety criteria

| $k$ |  | Best dose <br> (utility index by prob) | Best dose <br> (utility index by standardized <br> response) |
| :--- | :--- | :--- | :--- |
| 0.2 | true | $0.63(1.073)$ | $0.75(1.910)$ |
|  | simu. | $0.63(1.074)$ | $0.76(1.919)$ |
| 0.4 | true | $0.56(1.218)$ | $0.66(1.847)$ |
|  | simu. | $0.55(1.215)$ | $0.67(1.847)$ |
| 0.6 | true | $0.52(1.364)$ | $0.62(1.803)$ |
|  | simu. | $0.51(1.361)$ | $0.61(1.797)$ |
| 0.8 | true | $0.49(1.512)$ | $0.58(1.767)$ |
|  | simu. | $0.48(1.508)$ | $0.58(1.757)$ |

Note:The correlation between efficacy and safety response is 0.8 . Best dose provide maximum utility index to satisfy efficacy and safety criteria

## Chapter 6

## Conclusions

In Part I of this dissertation, first we extend the MCP-Mod to the safety outcome; then we develop the methodology for how to identify the final dose for the Phase III program based on nonlinear models for continuous bivariate endpoints. The methodology include several procedures and are summarized as follows:

Extension of MCP-Mod to safety outcome: MCP-Mod, developed by Bretz et al. (2005), was used to select the minimum effective dose which can provide the acceptable efficacy treatment effect. In Part I of the dissertation, we extend the MCP-Mod to select the maximum safety dose (MSD) that shows clinically acceptable toxicity. This definition has a different concept from the $M E D$ definition. $\widehat{M S D}_{1}$ is considered conservative for dose selection based on a safety outcome. The statistical performance of $\widehat{M S D}_{1}$ and bootstrap C.I coverage for $\widehat{M S D}_{1}$ are reasonable and confirm that the use of $\widehat{M S D}_{1}$ is valid and applicable.

Joint nonlinear continuous bivariate model estimation: In Part I of this dis-
sertation, we utilize the generalized nonlinear least squares method to estimate the joint nonlinear continuous bivariate outcomes. This method uses a Gauss-Seidel algorithm with profiled likelihood to obtain the maximum likelihood estimates for the joint model parameters. As expected, the parameter estimates from the joint model fitting are more efficient than the parameter estimates from the separate model fitting. When the correlation between the bivariate endpoints is stronger, the relative efficiency increases and results in more advantage to use the joint model estimation.

Strategy to find the best joint model: Before we make efforts to estimate the best joint model, we realize that the efficacy is the key for the approval of the drug development program. Hence, if there is no PoC for the efficacy, then there is no need to continue to Phase III which means we won't continue to estimate the joint model for the bivariate data. When there is PoC for efficacy but PoC for the safety response does not exist, then we need only to find the $\widehat{M E D_{2}}$ since there is no toxicity effect found for the safety doseresponse curves under this scenario. On the other hand we will proceed to find the best joint model if there is PoC for both bivariate outcomes.

Therefore when both efficacy and safety responses are considered for the dose selection, there are two strategies we developed to find the final best joint model. Since both efficacy and safety responses can be fitted by different dose-response curves and it is possible more than one model can fit the efficacy or safety data. The most significant fitted model based on the separated model fitting for efficacy or safety data may not be the most significant model when we take the correlation between the bivariate endpoints into account. Therefore the first strategy starts with identifying the most significant model
based on the separate efficacy or safety data model fitting by the lowest AIC criterion and t-statistics are larger than the critical values $\left(q^{Y}, q^{Z}\right)$ that control the FWERs. The final model will be the joint model from the two identified most significant models from the separate model fitting. The second strategy will identify all the significant models from the separate model fitting based on the t-statistics and critical values $\left(q^{Y}, q^{Z}\right)$ that control the FWERs. The final joint model is then selected by the lowest AIC criterion based on all the joint model combinations. The simulation results imply that both approaches provide more efficient parameter estimates than the separate model fitting. However, the second strategy seems to provide slightly more efficient parameter estimates based on the MSE and percent bias under stronger correlation between the bivariate outcomes.

Identifying the final dose(s) to carry into the Phase III program: After we find the $\widehat{M E D_{2}}$ and $\widehat{M S D}_{1}$ based on the joint modeling, how to identify the final dose(s) is still unresolved. In Part I of the dissertation, we consider two criteria for the success of Phase III program. The first criteria is the joint success criteria defined by the continuous efficacy and safety responses. The dose(s) will be identified by the probability of satisfying the joint criteria which is computed from the joint density of bivariate outcomes. The second criteria is the utility function which is the trade off of the efficacy and safety, this will need more inputs from clinicians. Overall both criteria largely rely on the individual profiles of the drug.

## Part II

## Dose Finding for Joint Continuous

## and Discrete Bivariate Responses

## Chapter 7

## Introduction to Joint Continuous and

## Discrete Bivariate Responses

### 7.1 Background

Part I addressed the methodology to estimate the joint continuous bivariate responses and to identify the final dose for the Phase III program. In clinical trials there are numerous cases that continuous and discrete endpoints are observed. To illustrate the case of mixed discrete and continuous endpoints, we may use the data from the Lipid Research Clinics Coronary Primary Prevention Trial (Freedman et al., 1992). These data concern the effect of the drug cholestramine on serum cholesterol levels as efficacy endpoint at one year and on cardiovascular events defined as either death from coronary heart disease or occurrence of myocardial infarction as safety endpoint. The true safety endpoint is binary (cardiovascular event or not), while the surrogate endpoint is cholesterol level.

Hence the surrogate can be considered as a continuous variable. Another simple yet real example is encountered in a randomized clinical trial (Pharmacological Therapy for Macular Degeneration Study Group) with mixed discrete-continuous endpoints. The primary endpoint of the trial was the loss of at least three lines of vision at one year, compared to their baseline performance (a binary endpoint). The secondary endpoint of this trial was the visual acuity at one year (treated as a continuous endpoint). Thus discrete outcomes may be thought as indicators of continuous variables that are either difficult or impossible to measure directly.

There are also cases that no direct underlying continuous variable exist for the observed discrete variable. For example patients with myocardial infarction may have a discrete outcome as mild, moderate or severe event, but there is no direct latent continuous variable to categorize this discrete variable.

As described above, it is very common that we need to consider joint continuous and discrete responses for the dose finding. Both efficacy and safety variables can be continuous, discrete or mixed-type outcomes. Though in the Part II of the thesis we consider only the joint continuous efficacy and binary safety outcomes for recommending doses of drug for the clinical development. This is only for the purpose of the illustration in this dissertation. In reality the efficacy can be discrete outcome and safety can be continuous outcome. Indeed the methodology presented here also applies to multiple efficacy endpoints or multiple safety endpoints, and the discrete variable is not restricted to binary endpoint only. The multivariate variable may be the same type or mixed type outcomes. The method presented here can also be applied to the case discussed in Part I.

### 7.2 Research objectives

In Part II, we will develop methodologies to identify the final dose to carry into Phase III program based on joint nonlinear models for mixed continuous and discrete responses. The followings are the objectives and specific aims of Part II.

### 7.2.1 Objectives

Objective I: To establish and estimate the joint nonlinear model when mixed continuous and discrete responses (efficacy and safety endpoints) are correlated (Chapter 9.3).

Objective II: To evaluate and select the best joint nonlinear model for mixed continuous and discrete responses (Chapter 9.4, 9.9).

Objective III: To define and determine more accurate MED and MSD estimates for the correlated nonlinear mixed continuous efficacy and discrete safety responses (Chapter 9.7, 9.8).

Objective IV: To identify the final dose(s) to carry into Phase III program with high probability of success when bivariate outcomes are of mixed type (Chapter 9.10).

Objective V: To simulate bivariate correlated mixed type responses data in order to evaluate the dose finding methodology for the joint nonlinear models for mixed type responses (Chapter 9.5).

### 7.2.2 Specific aims

The specific aims for Part II of this dissertation are summarized in the following:

1. To formulate bivariate nonlinear joint model for continuous and discrete responses,
and to develop methods for estimating the parameters in the joint model that account for the correlation between the continuous and discrete responses.

As mentioned above, the estimation methods for joint modeling of mixed type outcomes are not obvious. Some studies proposed generalized estimating equations (GEE2) (Prentice and Zhao, 1991) or extended least squares (ELS) (Vonesh et al., 2001) approaches to estimate the parameters for the models with mixed type outcomes. No applications have been illustrated for the efficiency of the parameter estimates. Most published work are for linear continuous response instead of nonlinear continuous response. We will explore the strategy for estimating the joint monlinear modeling for continuous and discrete outcomes. Furthermore, since the correlation between the continuous and discrete outcomes varies and depends on the dose level, estimation of the different correlation for different dose level adds further challenge.
2. To develop strategies to find the best joint nonlinear model for mixed type outcomes.

Since the joint model we discussed here is nonlinear regression of mixed type of responses, we need to explore the method for assessing the model fitting.

## 3. To define the MSD based on the discrete safety outcome.

The maximum safety dose is the highest dose that its mean toxicity does not exceed the mean toxicity of the zero dose by a specified threshold. If the safety outcome is a discrete endpoint, the parameter estimate from the joint model for the safety endpoint used certain links in the modeling. Furthermore the specific threshold is not directly measured by the binary endpoint, but by the probability of the binary endpoint.
4. To determine the final dose(s) to carry into the Phase III program.

In studying the dose response relationship of a new drug, we need to consider the nature of the dose response relationship for both continuous and discrete endpoints. In order to identify the optimal dose, the focus is on how to apply the dose response relationship to efficacy, safety, and the benefit/risk ratio. All these aspects need to be evaluated.
5. To develop the strategy to simulate bivariate correlated mixed type responses data to evaluate the dose finding methodology proposed.

In order to evaluate the dose finding methodology, including the joint nonlinear model for joint mixed type responses, we need to simulate bivariate correlated mixed type outcomes in this study. Since one of the research goals is to evaluate the estimates from different nonlinear joint models, we need to find the relationships of parameters, the mean model parameters and the correlations of the continuous and discrete variables for different joint model formation. We will explore the strategy for the simulation, including the relationship between the observed discrete outcome and the underlying continuous outcome regarding the mean model and correlations.

## Chapter 8

## Literature Review for Estimation in

## Multivariate Responses

In addition to the existing issues in the dose finding field reviewed in the Part I, in reality not all bivariate outcomes are both continuous responses. Statistical problems where various outcomes of a mixed nature are observed and are rather common at present in clinical trials. Perhaps the most common situations in clinical trials are that the safety responses maybe observed as discrete responses. Safety responses can be observed as adverse events or the discrete lab outcomes. In most situations the observed discrete outcome is derived from a latent continuous outcome. Thus under these scenarios the observed discrete outcome may be assumed to arise from an unobservable continuous random variable. This discrete outcome represents an indicator of whether this underlying variable exceeds some threshold. On the other hand we also realize that some discrete outcomes do not have a direct underlying continuous variable. It is also true that we may need to
consider multiple efficacy or multiple safety outcomes for the drug development instead of combined efficacy and safety outcomes as clinical interests.

Methods for jointly modeling continuous and discrete responses are not obvious, especially with nonlinear models for both outcomes. It is clear that the literature on joint modeling of outcomes of various nature is diverse and growing. The following section will discuss the current existing approaches for estimating the parameters in models for the multi-responses.

### 8.1 Existing estimating approaches for multivariate responses

For the problem sketched above, one approach is to factor the joint distribution into the product of a marginal and a conditional distribution. There are two versions of a conditional models, depending on whether the conditionaling is done on the continuous or the discrete outcome (Cox 1972; Krzanowski 1988; Cox and Wermuth 1992; Fitzaurice and Laird 1997). The choice of conditioning is mostly for statistical convenience rather than biological rationale, as relatively little is understood about the biologic mechanisms of developmental toxicity. This approach focus on estimating the mean-dose parameters while accounting for the bivariate correlation. Conditional models have some drawbacks in this setting, including lack of a marginal dose-response interpretation in models that condition on continuous variable and difficulty in quantitative assessment. Although such models may take into account the dependence between the bivariate variable, the correlation itself may not be directly estimated and so the joint probability of the event of the bivariate outcome is unavailable. Furthermore when the continuous outcome follows nonlinear
regression model then the estimation adds further challenges.
Another model family is that a joint model is formulated for the two outcomes. In this context, one often starts from a bivariate continuous variable, one component of which is explicitly observed and the other one observed in dichotomized, or generally discretized, version only. Regan and Catalano (1999) proposed likelihood models that fully specify the joint distribution of the continuous and discrete outcomes which directly estimate all mean dose-response and correlation parameters. These approaches focus on estimating the mean dose-response parameters. Difficulty then arises in nonlinear continuous outcome in the combined continuous and discrete outcomes because the likelihood is not in a closed form. In addition, since the key parameters usually must be interpreted conditional on the latent variables, distributional assumptions on the latent variables are necessarily somewhat arbitrary and untestable, and computational aspects of model fitting may be difficult.

GEE methodology has been proposed for multivariate continuous and discrete outcomes (Prentice and Zhao, 1991; Zeger and Liang, 1991; Rochon 1996). A set of secondorder generalized estimating equations (GEE2) for jointly estimating the mean and second moment parameters for a multivariate response vector are within the context of linear regression for continuous variable. Furthermore, mis-specified second moments will lead to poor estimation of the mean model parameters. Vonesh et al. (2001) focused on estimating the parameters of a general multivariate nonlinear regression model. Estimation is based on iteratively reweighted generalized least squares (IRGLS) and is carried out through repeated application of Taylor series linearization and estimated generalized least squares
(EGLS). Consequently, this IRGLS procedure may be viewed as performing a series of linear regressions using weighted least squares. An advantage of the procedure is that it only requires specification of the first two moments, while it does require that the first four moments exist in order for the estimators to be consistent. When the covariance parameter is unknown and need to be estimated with the location parameter in the mean vector, this leads to the joint estimation both location parameter and covariance parameter which is called extended least squares (ELS). However, the paper only showed and estimated the marginal multivariate nonlinear regression model in which the marginal mean and variance-covariance matrix share a common set of parameters.

### 8.1.1 GEE2 method

The GEE2 method has been shown as equivalent to the extended least squares (ELS) method under normality in Vonesh et al. (2001). Given a sample of $K$ independent random observations $\boldsymbol{y}_{k}^{T}=\left(y_{k 1}, \ldots, y_{k n_{k}}\right), k=1, \ldots, K$, of a general multivariate response vector. Liang and Zeger (1986) and Liang (1986) proposed that a parameter vector $\beta$ in the mean response $\boldsymbol{\mu}_{k}^{T}=\boldsymbol{\mu}_{k}^{T}(\boldsymbol{\beta})=\left\{E\left(y_{k 1}\right), E\left(y_{k 1}\right), \ldots\right\}$ can be estimated as the solution to the following equation:

$$
\begin{equation*}
K^{-1 / 2} \sum_{k=1}^{K} \boldsymbol{D}_{k 11}^{T} \boldsymbol{V}_{k 11}^{T}\left(\boldsymbol{y}_{k}-\boldsymbol{\mu}_{k}\right)=0, \tag{8.1}
\end{equation*}
$$

where $\mathbf{D}_{k 11}=\frac{\partial \mu_{k}}{\partial \boldsymbol{\beta}^{T}}$ and $\boldsymbol{V}_{k 11}$ is the variance matrix for $\boldsymbol{y}_{k}$.
Prentice (1988) introduced a second set of estimating equations (8.2) to estimate $\alpha$ which is used to characterize $\boldsymbol{V}_{k 11}$. This set of estimating equations is similar to (8.1) but with $\boldsymbol{y}_{k}-\boldsymbol{\mu}_{k}$ replaced by the vector of differences between the empirical and true pairwise
correlations.

$$
\begin{equation*}
K^{-1 / 2} \sum_{k=1}^{K} \boldsymbol{D}_{k 22}^{T} \boldsymbol{V}_{k 22}^{T}\left(\boldsymbol{s}_{k}-\boldsymbol{\sigma}_{k}\right)=0, \tag{8.2}
\end{equation*}
$$

where $\sigma_{k}$ and $s_{k}$ denote the covariance matrix $V_{k 11}$ and empirical covariance in vector forms. $\mathbf{D}_{k 22}=\frac{\partial \sigma_{k}}{\partial \alpha^{T}}$ and $V_{k 22}$ is a "working" variance matrix for the vector of empirical covariances $s_{k}$.

In order to specify the asymptotic distribution for $(\widehat{\boldsymbol{\beta}}, \widehat{\alpha})$ that solve (8.1) and (8.2), denote

$$
\begin{gathered}
\boldsymbol{D}_{k}=\left[\begin{array}{cc}
\frac{\partial \boldsymbol{u}_{k}}{\partial \boldsymbol{\beta}^{T}} & 0 \\
0 & \frac{\partial \sigma_{k}}{\partial \boldsymbol{\alpha}^{T}}
\end{array}\right]=\left[\begin{array}{cc}
\mathbf{D}_{k 11} & \mathbf{0} \\
\mathbf{0} & \mathbf{D}_{k 22}
\end{array}\right], \\
\boldsymbol{V}_{k}=\left[\begin{array}{cc}
V_{k 11} & 0 \\
0 & \boldsymbol{V}_{k 22}
\end{array}\right], f_{k}=\left[\begin{array}{c}
\boldsymbol{y}_{k}-\boldsymbol{u}_{k} \\
\boldsymbol{s}_{k}-\boldsymbol{\sigma}_{k}
\end{array}\right], \\
\boldsymbol{\Sigma}=K^{-1} \sum_{k=1}^{K} \boldsymbol{D}_{k}^{T} \boldsymbol{V}_{k}^{T} \boldsymbol{D}_{k}
\end{gathered}
$$

we have

$$
\begin{equation*}
\left[K^{-1 / 2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^{T}, K^{-1 / 2}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})\right] \sim A N\left(\mathbf{0}, K^{-1} \boldsymbol{\Sigma}^{-1}\left(\sum_{k=1}^{K} \boldsymbol{D}_{k}^{T} \boldsymbol{V}_{k}^{-1} \boldsymbol{f}_{k} \boldsymbol{f}_{k}^{T} \boldsymbol{V}_{k}^{-1} \boldsymbol{D}_{k}\right) \boldsymbol{\Sigma}^{-1}\right) \tag{8.3}
\end{equation*}
$$

$$
\left.\begin{array}{l}
(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^{T} \sim \\
A N\left(\begin{array}{c}
\mathbf{0},\left[\sum_{k=1}^{K} \boldsymbol{D}_{k 11}^{T} \boldsymbol{V}_{k 11}^{-1} \boldsymbol{D}_{k 11}\right]^{-1}
\end{array} \quad\left[\sum_{k=1}^{K} \boldsymbol{D}_{k 11}^{T} \boldsymbol{V}_{k 11}^{-1}\left(\boldsymbol{y}_{k}-\boldsymbol{u}_{k}\right)\left(\boldsymbol{y}_{k}-\boldsymbol{u}_{k}\right)^{T} \boldsymbol{V}_{k 11}^{-1} \boldsymbol{D}_{k 11}\right]\right)  \tag{8.4}\\
\\
\times\left[\sum_{k=1}^{K} \boldsymbol{D}_{k 11}^{T} \boldsymbol{V}_{k 11}^{-1} \boldsymbol{D}_{k 11}\right]^{-1}
\end{array}\right) .
$$

These features suggest that simple special cases of (8.1) and (8.2) will be attractive if interest resides primary in the mean parameter $\beta$. One can specify, or build, a model for the variance matrix $\boldsymbol{V}_{k 11}(\boldsymbol{\beta}, \boldsymbol{\alpha})$ that may imply good efficiency for $\widehat{\boldsymbol{\alpha}}$ estimation without being overly concerned about the choice of weight matrices $V_{k 22}$ for $s_{k}$. On the other hand, if interest resides in both the mean and covariance parameters, a systematic means of generating estimating equations would be desirable.

### 8.1.2 Iteratively reweighted generalized least squares (IRGLS), extended least squares (ELS) and quasi-extended least squares (QELS) methods

In order to adapt the ELS and QELS methods to dose-finding for joint nonlinear continuous and discrete data, the detailed estimating procedures are introduced.

As described in Vonesh et al. (2001), assume a given sample of $n$ individuals with a $p_{i}$ measurements on the $i$ th subject. Let $\boldsymbol{y}_{i}=\left[y_{i 1}, \ldots y_{i p_{i}}\right]^{\prime} . \boldsymbol{y}_{i}$ are independently distributed with mean and variance-covariance given by

$$
\begin{aligned}
E\left(\boldsymbol{Y}_{i}\right) & =\boldsymbol{\mu}_{i}(\boldsymbol{\beta}), \\
\operatorname{var}\left(\boldsymbol{Y}_{i}\right) & =\Sigma_{i}(\boldsymbol{\beta}, \boldsymbol{\theta}) .
\end{aligned}
$$

Denote

$$
\begin{aligned}
\boldsymbol{S}_{i}(\boldsymbol{\beta}) & =\operatorname{Vec}\left(\left(\boldsymbol{Y}_{i}-\boldsymbol{\mu}_{i}(\boldsymbol{\beta})\left(\boldsymbol{Y}_{i}-\boldsymbol{\mu}_{i}(\boldsymbol{\beta})\right)^{T}\right)\right. \\
\sigma_{i}(\boldsymbol{\beta}, \boldsymbol{\theta}) & =E\left(\boldsymbol{S}_{i}(\boldsymbol{\beta})\right) \\
\boldsymbol{\omega}_{i} & =\operatorname{Vech}\left(\boldsymbol{Y}_{i} \boldsymbol{\Upsilon}_{i}^{T}\right) .
\end{aligned}
$$

Vech(.) is the matrix operator that creates a column vector from a square matrix by stacking
the diagonal and lower diagonal elements below one another. Vec(.) is the matrix operator that creates a column vector from a matrix A by simply stacking the column vectors of A below one another.

The third- and fourth-order moments are as follows:

$$
\Theta_{i}=\operatorname{cov}\left(\boldsymbol{\varphi}_{i}, \boldsymbol{\omega}_{i}\right), \Gamma_{i}=\operatorname{var}\left(\boldsymbol{\omega}_{i}\right)
$$

Let $\tau^{T}=\left(\boldsymbol{\beta}^{T}, \boldsymbol{\theta}^{T}\right)$, and define

$$
\boldsymbol{Y}_{i}=\binom{\boldsymbol{y}_{i}}{\boldsymbol{w}_{i}}, \boldsymbol{f}_{i}(\boldsymbol{\tau})=\binom{\boldsymbol{\mu}_{i}(\beta)}{\gamma_{i}(\beta, \boldsymbol{\theta})}, \boldsymbol{\psi}_{i}(\boldsymbol{\tau})=\left[\begin{array}{cc}
\boldsymbol{\Sigma}_{i}(\boldsymbol{\beta}, \boldsymbol{\theta}) & \boldsymbol{\Theta}_{i}(\beta, \boldsymbol{\theta}) \\
\boldsymbol{\Theta}_{i}(\beta, \boldsymbol{\theta}) & \boldsymbol{\Gamma}_{i}(\boldsymbol{\beta}, \boldsymbol{\theta})
\end{array}\right] .
$$

The joint estimation of $\boldsymbol{\tau}^{T}=\left(\boldsymbol{\beta}^{T}, \boldsymbol{\theta}^{T}\right)$ can be based on the non-linear model

$$
\boldsymbol{\Upsilon}_{i}=f_{i}(\boldsymbol{\tau})+\boldsymbol{\xi}_{i,} \quad E\left(\boldsymbol{\xi}_{i}\right)=\mathbf{0}, \operatorname{var}\left(\boldsymbol{\xi}_{i}\right)=\boldsymbol{\Psi}_{i}(\boldsymbol{\tau}) .
$$

Let $\boldsymbol{t}^{T}=\widehat{\boldsymbol{\tau}}^{0^{T}}=\left(\widehat{\boldsymbol{\beta}}^{0 T}, \widehat{\boldsymbol{\theta}}^{0 T}\right)$ be an initial estimate of $\boldsymbol{\tau}^{T}=\left(\boldsymbol{\beta}^{T}, \boldsymbol{\theta}^{T}\right)$ and assume $\boldsymbol{t}$ is in the interior of parameter space of $\tau$. By applying the usual Gauss-Newton algorithm for the nonlinear regression, estimation may be carried out by taking a first-order Taylor series expansion of $\boldsymbol{f}_{i}(\boldsymbol{\tau})$ about $\boldsymbol{\tau}=\boldsymbol{t}$ yielding the approximation $\boldsymbol{Y}_{i}=\boldsymbol{f}_{i}(\boldsymbol{\tau})+\boldsymbol{X}_{i t}(\boldsymbol{\tau}-\boldsymbol{t})+\boldsymbol{\xi}_{i}$, which can be rewritten in terms of the linear model

$$
\begin{equation*}
\boldsymbol{\Upsilon}_{i}^{*}=\boldsymbol{X}_{i t} \boldsymbol{\tau}+\boldsymbol{\xi}_{i,} \tag{8.5}
\end{equation*}
$$

where $\boldsymbol{X}_{i t}=\partial f_{i}(\boldsymbol{\tau}) /\left.\partial \tau^{T}\right|_{\tau=t}$ and $\boldsymbol{Y}_{i}^{*}=\boldsymbol{Y}_{i}-f_{i}(\boldsymbol{t})+\boldsymbol{X}_{i t} \boldsymbol{t}$, based on this linear model, an estimated generalized least squares estimate (EGLS) of $\tau$ is given by

$$
\begin{equation*}
\widehat{\boldsymbol{\tau}}=\left(\sum_{i=1}^{n} \boldsymbol{X}_{i t}^{T} \boldsymbol{\Psi}_{i}(\boldsymbol{t})^{-1} \boldsymbol{X}_{i t}\right)^{-1} \sum_{i=1}^{n} \boldsymbol{X}_{i t}^{T} \boldsymbol{\Psi}_{i}(\boldsymbol{t})^{-1} \boldsymbol{Y}_{i,}^{*} \tag{8.6}
\end{equation*}
$$

where $\Psi(t)$ is the assumed covariance matrix of $\boldsymbol{\xi}_{i}$ evaluated at $\tau=t$.

Iteratively reweighted generalized least squares (IRGLS) entails iterating between (8.5) and (8.6) by setting $\tau=t$ in (8.5) and then using (8.6) to obtain an updated estimate of $\boldsymbol{\tau}$. By repeating this process, obtain a sequence of one-step Gauss-Newton estimators, $\left\{\widehat{\tau}^{k} \mid k=1,2, \ldots\right\}$, which, as $k \rightarrow \infty$, yields a solution to the set on nonlinear "normal" estimating equations,

$$
\begin{equation*}
\boldsymbol{U}(\boldsymbol{\tau})=\sum_{i=1}^{n}\left\{\boldsymbol{X}_{i}(\boldsymbol{\tau})^{T} \mathbf{\Psi}_{i}(\boldsymbol{\tau})^{-1}\left(\boldsymbol{Y}_{i}-\boldsymbol{f}_{i}(\boldsymbol{\tau})\right)\right\}=0 \tag{8.7}
\end{equation*}
$$

where $\boldsymbol{X}_{i}(\boldsymbol{\tau})=\partial f_{i}(\boldsymbol{\tau}) / \partial \boldsymbol{\tau}^{T}$. When fully iterated, the model-based asympotic variancecovariance matrix of the IRGLS estimate, $\widehat{\tau}$ is estimated by the inverse of the expected information matrix evaluated at the final estimate, i.e,

$$
\widehat{\Omega}(\widehat{\boldsymbol{\tau}})=\left(\sum_{i=1}^{n} \boldsymbol{X}_{i}(\widehat{\boldsymbol{\tau}})^{T} \boldsymbol{\Psi}_{i}(\widehat{\boldsymbol{\tau}})^{-1} \boldsymbol{X}_{i}(\widehat{\boldsymbol{\tau}})\right)^{-1}
$$

To safeguard against mis-specification of $\boldsymbol{\Psi}_{i}(\boldsymbol{\tau})$, one can use a robust estimator of the variance of $\widehat{\tau}$, which is given by the empirical "sandwich" estimator as follows:

$$
\begin{equation*}
\widehat{\boldsymbol{\Omega}}_{R}(\widehat{\boldsymbol{\tau}})=\widehat{\boldsymbol{\Omega}}(\widehat{\boldsymbol{\tau}})\left(\sum_{i=1}^{n} \boldsymbol{X}_{i}(\widehat{\boldsymbol{\tau}})^{T} \boldsymbol{\Psi}_{i}(\widehat{\boldsymbol{\tau}})^{-1} \boldsymbol{e}_{i} \boldsymbol{e}_{i}^{T} \boldsymbol{\Psi}_{i}(\widehat{\boldsymbol{\tau}})^{-1} \boldsymbol{X}_{i}(\widehat{\boldsymbol{\tau}})\right) \widehat{\boldsymbol{\Omega}}(\widehat{\boldsymbol{\tau}}), \tag{8.8}
\end{equation*}
$$

where $\boldsymbol{e}_{i}=\boldsymbol{Y}_{i}-\boldsymbol{f}_{i}(\widehat{\boldsymbol{\tau}})$.
It was shown by Vonesh et al. (2001) that the IRGLS procedure for estimating $\tau$, which consists simply of repeated application of Taylor series linearization and estimated generalized least squares, is equivalent to maximum likelihood estimation assuming normality.

In addition, Voensh et al. (2001) also discussed a least squares approach to estimate $\beta$ assuming $y_{i}$ has mean $\boldsymbol{\mu}_{i}(\beta)$ and variance-covariance $\Sigma_{i}(\beta, \theta)$. For convenience
suppose that $\theta=\boldsymbol{\theta}^{*}$ is known. The weighted least squares named as generalized least squares (GLS) which minimize the GLS objective function

$$
\begin{equation*}
\boldsymbol{Q}_{G L S}\left(\boldsymbol{\beta} \mid \boldsymbol{\theta}^{*}\right)=\sum_{i=1}^{n}\left\{\left(\boldsymbol{\gamma}_{i}-\boldsymbol{\mu}_{i}(\boldsymbol{\beta})\right)^{T} \boldsymbol{\Sigma}_{i}\left(\boldsymbol{\beta}, \boldsymbol{\theta}^{*}\right)^{-1}\left(\boldsymbol{\gamma}_{i}-\boldsymbol{\mu}_{i}(\boldsymbol{\beta})\right)\right\} . \tag{8.9}
\end{equation*}
$$

Since $\Sigma_{i}\left(\beta, \theta^{*}\right)$ depends on $\beta$, the estimation based on minimizing (8.9) may produce biased estimates of $\beta$. To avoid the biased estimating equations in GLS, the corrected generalized least squares (CGLS) estimating equation is used to provide unbiased estimating equations correspond to minimizing the bias-corrected generalized least squares objective function

$$
\begin{equation*}
Q_{C G L S}\left(\beta \mid \boldsymbol{\theta}^{*}\right)=\sum_{i=1}^{n}\left\{\left(\boldsymbol{Y}_{i}-\boldsymbol{\mu}_{i}(\boldsymbol{\beta})\right)^{T} \boldsymbol{\Sigma}_{i}\left(\boldsymbol{\beta}, \boldsymbol{\theta}^{*}\right)^{-1}\left(\boldsymbol{Y}_{i}-\boldsymbol{\mu}_{i}(\boldsymbol{\beta})\right)+\log \left|\boldsymbol{\Sigma}_{i}\left(\boldsymbol{\beta}, \boldsymbol{\theta}^{*}\right)\right|\right\} \tag{8.10}
\end{equation*}
$$

where it is understood that minimization is with respect to $\beta$ only.
From the above well-defined CGLS objective function( 8.10), it does require that $\boldsymbol{\theta}=\boldsymbol{\theta}^{*}$ is known. By replacing $\boldsymbol{\theta}^{*}$ with any consistent estimate $\widehat{\boldsymbol{\theta}}$, this ends up with a conditional objective function $Q_{C G L S}(\beta \mid \widehat{\boldsymbol{\theta}})$ and is subject to small sample bias. An alternative to independently estimating $\boldsymbol{\theta}$ prior to optimizing $\boldsymbol{Q}_{C G L S}(\boldsymbol{\beta} \mid \widehat{\boldsymbol{\theta}})$, Vonesh et al. (2001) proposed to jointly estimate both and $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ by simply including $\boldsymbol{\theta}$ as an unknown parameter in the objective function (8.10) as follows.

$$
\begin{equation*}
\boldsymbol{Q}_{E L S}(\boldsymbol{\beta}, \boldsymbol{\theta})=\sum_{i=1}^{n}\left\{\left(\boldsymbol{Y}_{i}-\boldsymbol{\mu}_{i}(\boldsymbol{\beta})\right)^{T} \boldsymbol{\Sigma}_{i}(\boldsymbol{\beta}, \boldsymbol{\theta})^{-1}\left(\boldsymbol{Y}_{i}-\boldsymbol{\mu}_{i}(\boldsymbol{\beta})\right)+\log \left|\boldsymbol{\Sigma}_{i}(\boldsymbol{\beta}, \boldsymbol{\theta})\right|\right\} \tag{8.11}
\end{equation*}
$$

Formula (8.11) denotes the extended least squares objective function associated with the joint estimation of both $\beta$ and $\boldsymbol{\theta}$. The corresponding joint estimating equations for
maximizing $Q_{E L S}(\beta, \theta)$ are given as follows:
$\boldsymbol{U}_{E L S}(\boldsymbol{\beta}, \boldsymbol{\theta})=\sum_{i=1}^{n}\left[\begin{array}{cc}\boldsymbol{D}_{i}(\boldsymbol{\beta}) & \mathbf{0} \\ \boldsymbol{E}_{i}(\boldsymbol{\beta}) & \boldsymbol{E}_{i}(\boldsymbol{\theta})\end{array}\right]^{T}\left[\begin{array}{cc}\boldsymbol{\Sigma}_{i}(\boldsymbol{\beta}, \boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{V}_{i}(\boldsymbol{\beta}, \boldsymbol{\theta})\end{array}\right]^{-1}\left[\begin{array}{c}\boldsymbol{y}_{i}-\boldsymbol{u}_{i}(\boldsymbol{\beta}) \\ S_{i}(\boldsymbol{\beta})-\sigma_{i}(\boldsymbol{\beta}, \boldsymbol{\theta})\end{array}\right]=0$,
where $\boldsymbol{E}_{i}(\boldsymbol{\theta})=\partial \sigma_{i}(\boldsymbol{\beta}, \boldsymbol{\theta}) / \partial \boldsymbol{\theta}^{T}$.
It is straightforward to show that minimizing the ELS objective function $Q_{E L S}(\beta, \theta)$, is equivalent to maximizing the joint log-likelihood function of $\boldsymbol{y}_{i}, i=1, \ldots, n$, assuming the $\boldsymbol{y}_{i}$ are independently normal distributed with mean $\boldsymbol{u}_{i}(\beta)$ and covariance matrix $\boldsymbol{\Sigma}_{i}(\beta, \boldsymbol{\theta})$.

The formula (8.12) is equivalent to the second-order generalized estimating equations (GEE2) and same to the set of nonlinear "normal" estimating equations given in (8.7) under normality assumption which is described in Vonesh et al. (2001). The advantage of using the ELS approach is that the 3rd and 4th order moments are automatically generated as through from a multivariate normal distribution even though data are not multivariate normal. The empirical sandwich estimator can be used to achieve asymptotically valid inference provided that mean and variance structure have been correctly specified.

If $\boldsymbol{E}_{i}(\boldsymbol{\beta})=\partial \sigma_{i}(\boldsymbol{\beta}, \boldsymbol{\theta}) / \partial \boldsymbol{\beta}^{T}=\mathbf{0}$ in formula (8.12), one gets the set of quasi-extended least squares (QELS) estimating equations described in Vonesh et al. (2001). The advantage of using QELS estimating equations is that under certain conditions, one can still achieve a consistent estimate of $\beta$ even if the variance-covariance structure is mis-specified.

### 8.2 Unaddressed areas from the existing approaches for multivariate responses

As summarized above, most papers either use factoring the joint distribution into a conditional model approach or use direct joint distribution approach. These methods only applied to linear model for continuous and discrete outcomes. For nonlinear regression models for mixed type responses, the approaches developed in the these papers can't be applied directly.

Vonesh et al. (2001) considered the estimation of multivariate nonlinear regression for the joint continuous and discrete variable. However, the application described in the paper is in the setting of the marginal mean and variance-covariance structure share a common set of regression parameters. For dose finding studies, both joint continuous and discrete outcomes usually follow nonlinear regression and do not share common set of regression parameters. Furthermore, the correlation between the continuous and discrete outcomes will be different for each dose level. These haven't been explored in the currently published papers.

How to adapt the extended least squares (ELS) or quasi-extended least squares (QELS) methods into the dose finding studies data will be explored and addressed in this dissertation.

Furthermore, how to develop the methodology to identify the final dose(s) to carry into Phase III program based on nonlinear mixed type efficacy and safety have not been explored so far in the clinical drug development. To address the theoretical and practical issues mentioned above will be quite worthwhile.

## Chapter 9

## Developing the Methodology to

## Estimate $M E D$ and $M S D$ for Joint

## Continuous and Discrete Responses

### 9.1 Joint nonlinear bivariate continuous and discrete model

 sponse $Z_{i j}$. Without loss of generality, both the $Y_{i j}$ and $Z_{i j}$ responses can be either continuous or discrete. The generalized nonlinear joint model can be written as follows:

$$
\left[\begin{array}{l}
h_{1}\left(u_{i}^{Y}\right) \\
h_{2}\left(u_{i}^{Z}\right)
\end{array}\right]=\left[\begin{array}{l}
h_{1}\left(E\left(Y_{i j}\right)\right) \\
h_{2}\left(E\left(Z_{i j}\right)\right)
\end{array}\right]=\left[\begin{array}{c}
f^{*}\left(d_{i}, \boldsymbol{\theta}_{Y}\right) \\
g^{*}\left(d_{i}, \boldsymbol{\theta}_{Z}\right)
\end{array}\right], \mathrm{i}=1, \ldots \mathrm{k}, \mathrm{j}=1, \ldots . \mathrm{n}_{i}
$$

where $\left[\begin{array}{c}\boldsymbol{\theta}^{Y} \\ \boldsymbol{\theta}^{Z}\end{array}\right]$ refers to the vector of model parameters, $i$ to the dose group ( $i=1$ corresponds to placebo), and $j$ to the patient within dose group $i$.

There are the following cases depend on the outcomes measures of efficacy variable $Y_{i j}$ and the safety variable $Z_{i j}$ :

1. If both $Y_{i j}$ and $Z_{i j}$ are continuous outcomes and are normally distributed, then both $h_{1}($.$) and h_{2}($.$) equal the identity links.$
2. If $Y_{i j}$ is continuous outcome and $Z_{i j}$ is discrete outcome, then $h_{1}($.$) equals to the$ identity link and and $h_{2}($.$) equals to some link function such as logit link.$
3. If both $Y_{i j}$ and $Z_{i j}$ are discrete outcomes, then both $h_{1}($.$) and h_{2}($.$) equal to some link$ functions such as logit link for binary outcome.

Case (1) is already discussed in the previous chapters and case (2) is the focus of Part II of the dissertation. Vonesh et al. (2001) developed the method of estimating parameters for the same type of outcomes in multivariate nonlinear regression and the outcomes share a common set of parameters which assumed the outcomes follow the same distributions.

In this chapter focus will be on the nonlinear mixed outcomes of continuous and discrete binary responses for simplicity. The continuous efficacy response may follow the nonlinear regression and binary response may be seen often to follow the logistic regression, probit or complementary log-log regression models.

The nonlinear continuous and discrete variable can be written as follows:

$$
\left[\begin{array}{c}
u_{i}^{Y} \\
h_{2}\left(u_{i}^{Z}\right)
\end{array}\right]=\left[\begin{array}{c}
E\left(Y_{i j}\right) \\
h_{2}\left(E\left(Z_{i j}\right)\right)
\end{array}\right]=\left[\begin{array}{c}
f\left(d_{i}, \boldsymbol{\theta}_{Y}\right) \\
g^{*}\left(d_{i}, \boldsymbol{\theta}_{Z}\right)
\end{array}\right], i=1, \ldots k, j=1, \ldots . n_{i},
$$

or

$$
\left[\begin{array}{c}
u_{i}^{Y}  \tag{9.1}\\
u_{i}^{Z}
\end{array}\right]=\left[\begin{array}{c}
E\left(Y_{i j}\right) \\
E\left(Z_{i j}\right)
\end{array}\right]=\left[\begin{array}{c}
f\left(d_{i}, \boldsymbol{\theta}_{Y}\right) \\
g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)
\end{array}\right], \mathrm{i}=1, \ldots \mathrm{k}, \mathrm{j}=1, \ldots . \mathrm{n}_{i^{\prime}}
$$

where $\boldsymbol{\theta}^{T}=\left(\boldsymbol{\theta}_{\gamma}^{T}, \boldsymbol{\theta}_{Z}^{T}\right)$, which refers to the vector of model parameters, $i$ to the dose group ( $i=1$ corresponds to placebo), and $j$ to the patient within dose group $i$, and $g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)=$ $h_{2}^{-1}\left(g^{*}\left(d_{i}, \boldsymbol{\theta}_{Z}\right)\right)$.

The joint distribution of $\left(Y_{i j}, Z_{i j}\right)$ is

$$
f\left(y_{i j}, z_{i j}\right)=f_{Y_{i j}}\left(y_{i j}\right) f_{Z_{i j} \mid Y_{i j}}\left(z_{i j} \mid y_{i j}\right)=f_{Z_{i j}}\left(z_{i j}\right) f_{Y_{i j} \mid Z_{i j}}\left(y_{i j} \mid z_{i j}\right) .
$$

### 9.2 Likelihood for nonlinear joint continuous and discrete bivariate model

The continuous efficacy variable $Y_{i j}$ and binary safety variable $Z_{i j}$ are described Section 9.1. The likelihood based on the nonlinear joint model of continuous and binary responses can be written as below:

$$
L=\sum_{i=1}^{k} \sum_{j=1}^{n_{i}} \log \left\{f\left(y_{i j}, z_{i j}\right)\right\}
$$

where $f\left(y_{i j}, z_{i j}\right)$ is the joint distribution of $\left(Y_{i j}, Z_{i j}\right)$ and $\left(Y_{i j}, Z_{i j}\right)$ are independent.
The following two methods illustrate how to derive the likelihood for the joint nonlinear continuous and discrete variable based on whether we have the information for the latent continuous safety variable.

### 9.2.1 Method 1: The latent variable threshold $c$ and standard deviation $\sigma_{S}$ of latent variable are known

Under the clinical setting, the observed binary safety outcome $Z_{i j}$ may result from dichotomizing latent continuous safety random variable $S_{i j}$ and $Z_{i j}$ represents an observable indicator based on some chosen threshold $c$. For example, ALT, AST are continuous lab measurements which can be dichotomized by some threshold $c$.

Assume that

$$
\left[\begin{array}{c}
Y_{i j} \\
S_{i j}
\end{array}\right] \sim N\left(\left[\begin{array}{c}
u_{i}^{Y} \\
u_{i}^{S}
\end{array}\right],\left[\begin{array}{cc}
\sigma_{Y}^{2} & \rho_{i} \sigma_{Y} \sigma_{S} \\
\rho_{i} \sigma_{Y} \sigma_{S} & \sigma_{S}^{2}
\end{array}\right]\right)
$$

and

$$
Z_{i j}=\left\{\begin{array}{l}
1 \text { if } S_{i j}>c  \tag{9.2}\\
0 \text { if } S_{i j} \leq c
\end{array} .\right.
$$

The joint distribution of $\left(Y_{i j}, Z_{i j}\right)$ is

$$
f\left(y_{i j}, z_{i j}\right)=f\left(y_{i j}\right) f\left(z_{i j} \mid y_{i j}\right),
$$

where $f\left(y_{i j}\right)$ is the pdf of the continuous efficacy variable $Y_{i j} \sim N\left(u_{i}^{Y}, \sigma_{Y}^{2}\right)$ and

$$
f\left(z_{i j} \mid y_{i j}\right)=\left(P\left(Z_{i j}=1 \mid Y_{i j}=y_{i j}\right)\right)^{z_{i j}}\left(1-P\left(Z_{i j}=1 \mid Y_{i j}=y_{i j}\right)\right)^{1-z_{i j}},
$$

where

$$
\begin{aligned}
P\left(Z_{i j}=1 \mid Y_{i j}=y_{i j}\right) & =P\left(S_{i j}>c \mid Y_{i j}=y_{i j}\right) \\
& =1-P\left(S_{i j} \leq c \mid Y_{i j}=y_{i j}\right)
\end{aligned}
$$

From the general multivariate normal theory, the conditional distribution yields the following:

$$
\begin{aligned}
S_{i j} \mid Y_{i j} & \sim N\left(u_{i}^{S}+\rho_{i} \sigma_{S} / \sigma_{Y}\left(Y_{i j}-u_{i}^{Y}\right), \sigma_{S}^{2}-\left(\rho_{i} \sigma_{Y} \sigma_{S}\right)^{2} / \sigma_{Y}^{2}\right) \\
& \sim N\left(u_{i}^{S}+\rho_{i} \sigma_{S} / \sigma_{Y}\left(Y_{i j}-u_{i}^{Y}\right), \sigma_{S}^{2}\left(1-\rho_{i}^{2}\right)\right)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
P\left(S_{i j}\right. & \left.\leq c \mid Y_{i j}=y_{i j}\right) \\
& =P\left(\left.\frac{S_{i j}-\left(u_{i}^{S}+\rho_{i} \sigma_{S} / \sigma_{Y}\left(Y_{i j}-u_{i}^{Y}\right)\right)}{\sigma_{S} \sqrt{\left(1-\rho_{i}^{2}\right)}} \leq \frac{c-\left(u_{i}^{S}+\rho_{i} \sigma_{S} / \sigma_{Y}\left(Y_{i j}-u_{i}^{Y}\right)\right)}{\sigma_{S} \sqrt{\left(1-\rho_{i}^{2}\right)}} \right\rvert\, Y_{i j}=y_{i j}\right) \\
& =P\left(\left.\frac{S_{i j}-\left(u_{i}^{S}+\rho_{i} \sigma_{S} / \sigma_{Y}\left(y_{i j}-u_{i}^{Y}\right)\right)}{\sigma_{S} \sqrt{\left(1-\rho_{i}^{2}\right)}} \leq \frac{c-\left(u_{i}^{S}+\rho_{i} \sigma_{S} / \sigma_{Y}\left(y_{i j}-u_{i}^{Y}\right)\right)}{\sigma_{S} \sqrt{\left(1-\rho_{i}^{2}\right)}} \right\rvert\, Y_{i j}=y_{i j}\right) \\
& =\Phi\left(\frac{c-\left(u_{i}^{S}+\rho_{i} \sigma_{S} / \sigma_{Y}\left(y_{i j}-u_{i}^{Y}\right)\right)}{\sigma_{S} \sqrt{\left(1-\rho_{i}^{2}\right)}}\right) .
\end{aligned}
$$

The log-likelihood can be written as follows:

$$
\begin{align*}
L= & \sum_{i=1}^{k} \sum_{j=1}^{n_{i}} \log \left\{f\left(y_{i j}, z_{i j}\right)\right\} \\
= & \sum_{i=1}^{k} \sum_{j=1}^{n_{i}} \log \left\{f\left(y_{i j}\right)\left(P\left(Z_{i j}=1 \mid Y_{i j}=y_{i j}\right)\right)^{z_{i j}}\left(1-P\left(Z_{i j}=1 \mid Y_{i j}=y_{i j}\right)\right)^{1-z_{i j}}\right\} \\
= & \sum_{i=1}^{k} \sum_{j=1}^{n_{i}} \log \left\{f\left(y_{i j}\right)\left(1-\Phi\left(\frac{c-\left(u_{i}^{S}+\rho_{i} \frac{\sigma_{S}}{\sigma_{Y}}\left(y_{i j}-u_{i}^{Y}\right)\right)}{\sigma_{S} \sqrt{\left(1-\rho_{i}^{2}\right)}}\right)\right)^{z_{i j}}\right\} \\
& +\sum_{i=1}^{k} \sum_{j=1}^{n_{i}} \log \left\{\left(\Phi\left(\frac{c-\left(u_{i}^{S}+\rho_{i \frac{\sigma_{S}}{\sigma_{Y}}}\left(y_{i j}-u_{i}^{Y}\right)\right)}{\sigma_{S} \sqrt{\left(1-\rho_{i}^{2}\right)}}\right)\right)^{1-z_{i j}}\right\} \\
= & \sum_{i=1}^{k} \sum_{j=1}^{n_{i}}\left\{\log \left(\frac{1}{\left.\sqrt{2 \pi \sigma_{Y}^{2}}\right)}\left[-\frac{1}{2 \sigma_{Y}^{2}}\left(y_{i j}-u_{i}^{Y}\right)\right]\right\}\right. \\
& +\sum_{i=1}^{k} \sum_{j=1}^{n_{i}}\left\{z_{i j} \log \left[1-\Phi\left(\frac{c-\left(u_{i}^{S}+\rho_{i} \frac{\sigma_{S}}{\sigma_{Y}}\left(y_{i j}-u_{i}^{Y}\right)\right)}{\sigma_{S} \sqrt{\left(1-\rho_{i}^{2}\right)}}\right)\right]\right\} \\
& +\sum_{i=1}^{k} \sum_{j=1}^{n_{i}}\left\{\left(1-z_{i j}\right) \log \left[\Phi\left(\frac{c-\left(u_{i}^{S}+\rho_{i} \frac{\sigma_{S}}{\sigma_{Y}}\left(y_{i j}-u_{i}^{Y}\right)\right)}{\sigma_{S} \sqrt{\left(1-\rho_{i}^{2}\right)}}\right)\right]\right\} \tag{9.3}
\end{align*}
$$

where $u_{i}^{S}$ depends on the discrete safety response model assuming that the threshold value $c$ and standard deviation $\sigma_{S}$ are known. Furthermore $\rho_{i}$ can be derived from the $\widetilde{\rho}_{i}$ which is the correlation between variable $Y_{i j}$ and $Z_{i j}$ at each dose level. The correlations between observed bivariate mixed type responses and latent bivariate continuous responses in terms of mean model and correlation are discussed in the following section.

## Derive the relationship between observed bivariate mixed type responses and latent

## bivariate continuous responses:

Mean Model: A bivariate variable $\left[\begin{array}{l}Y_{i j} \\ Z_{i j}\end{array}\right]$, represents the continuous efficacy variable $Y_{i j}$ and the discrete safety variable $Z_{i j}$. Generalized nonlinear joint model is described in (9.1).

Assume $Z_{i j}$ represent observed discrete binary safety variable, $S_{i j}$ is latent continuous safety response variable and $c$ is the threshold value, where the association between indicator $Z_{i j}$ and latent variable $S_{i j}$ is indicated in (9.2).

Let $E\left(Z_{i j}\right)=g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)$, we need to find $u_{i}^{S}=E\left(S_{i j}\right)=h\left(d_{i}, \boldsymbol{\theta}_{Z}\right)$ in terms of $\boldsymbol{\theta}_{Z}$.

$$
\begin{aligned}
P\left(Z_{i j}=1 \mid d_{i}\right) & =P\left(S_{i j}>c \mid d_{i}\right) \\
& =P\left(\left.\frac{S_{i j}-h\left(d_{i}, \boldsymbol{\theta}_{Z}\right)}{\sigma_{S}}>\frac{c-h\left(d_{i}, \boldsymbol{\theta}_{Z}\right)}{\sigma_{S}} \right\rvert\, d_{i}\right) \\
& =P\left(\left.\varphi>\frac{c-h\left(d_{i}, \boldsymbol{\theta}_{Z}\right)}{\sigma_{S}} \right\rvert\, d_{i}\right),
\end{aligned}
$$

where $\varphi=\frac{s_{i j} h\left(d_{i}, \theta_{z}\right)}{\sigma_{S}}$, therefore,

$$
\begin{gather*}
1-\Phi\left(\frac{c-h\left(d_{i}, \boldsymbol{\theta}_{Z}\right)}{\sigma_{S}}\right)=g\left(d_{i}, \boldsymbol{\theta}_{Z}\right) \\
\Phi\left(\left.\frac{c-h\left(d_{i}, \boldsymbol{\theta}_{S}\right)}{\sigma_{S}} \right\rvert\, d_{i}\right)=1-g\left(d_{i}, \boldsymbol{\theta}_{Z}\right) \\
c-h\left(d_{i}, \boldsymbol{\theta}_{Z}\right)=\sigma_{S} \Phi^{-1}\left(1-g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)\right) \\
u_{i}^{S}=E\left(S_{i j}\right)=h\left(d_{i}, \boldsymbol{\theta}_{Z}\right)=c-\sigma_{S} \Phi^{-1}\left(1-g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)\right) \tag{9.4}
\end{gather*}
$$

For simplicity, we only consider the linear regression for the binary variable after certain link function is applied. The nonlinear regression model such as Emax model can also be used for $g^{*}\left(d_{i}, \boldsymbol{\theta}_{Z}\right)$ in the discrete outcome.

If discrete safety response follows logistic regression, then

$$
\begin{aligned}
u_{i}^{S} & =E\left(S_{i j}\right) \\
& =h\left(d_{i}, \theta_{Z}\right) \\
& =c-\sigma_{S} \Phi^{-1}\left(\frac{1}{1+\exp \left(\alpha+\beta d_{i}\right)}\right) .
\end{aligned}
$$

If discrete safety response follows probit regression, then

$$
\begin{aligned}
u_{i}^{S} & =E\left(S_{i j}\right) \\
& =h\left(d_{i}, \theta_{Z}\right) \\
& =c-\sigma_{S} \Phi^{-1}\left(1-\Phi\left(\alpha+\beta d_{i}\right)\right) \\
& =c+\sigma_{S}\left(\alpha+\beta d_{i}\right)
\end{aligned}
$$

If discrete safety response follows complementary log-log link regression, then

$$
\begin{aligned}
u_{i}^{S} & =E\left(S_{i j}\right) \\
& =h\left(d_{i}, \theta_{Z}\right) \\
& =c-\sigma_{S} \Phi^{-1}\left(1-\left(1-\exp \left(-\exp \left(\alpha+\beta d_{i}\right)\right)\right.\right. \\
& =c-\sigma_{S} \Phi^{-1} \exp \left(-\exp \left(\alpha+\beta d_{i}\right)\right)
\end{aligned}
$$

Correlation between continuous efficacy variable and discrete safety variable: Next, we will try to find the association of correlation $\widetilde{\rho}_{i}$ between bivariate mixed type responses and correlation $\rho_{i}$ between bivariate continuous responses:

$$
\begin{aligned}
\operatorname{cov}\left(Y_{i j}, Z_{i j}\right) & =E\left(Y_{i j} Z_{i j}\right)-E Y_{i j} E Z_{i j} \\
& =E\left(Y_{i j} Z_{i j}\right)-u_{i}^{Y} u_{i}^{Z}
\end{aligned}
$$

Since

$$
\begin{aligned}
E\left(Y_{i j} Z_{i j}\right) & =E\left(Y_{i j} I\left(S_{i j}>c\right)\right) \\
& =\int_{c} y I(s>c) d F_{Y_{i j} S_{i j}}(y, s) \\
& =\int_{c}^{\infty} \int_{-\infty}^{\infty} y d F_{Y_{i j} s_{i j}}(y, s) \\
& =\int_{c}^{\infty} \int_{-\infty}^{\infty} y f_{Y_{i j} \mid S_{i j}}(y \mid s) f_{S_{i j}}(s) d y d s \\
& =\int_{c}^{\infty} f_{S_{i j}}(s) \int_{-\infty}^{\infty} y f_{Y_{i j} \mid S_{i j}}(y \mid s) d y d s \\
& =\int_{c}^{\infty} f_{S_{i j}}(s) E\left(Y_{i j} \mid s\right) d s \\
& =\int_{c}^{\infty} f_{S_{i j}}(s)\left(u_{i}^{Y}+\frac{\rho_{i} \sigma_{Y}\left(s-u_{i}^{S}\right)}{\sigma_{S}}\right) d s \\
& =u_{i}^{Y} \int_{c}^{\infty} f_{S_{i j}}(s) d s+\rho_{i} \sigma_{Y} \int_{c} f_{S_{i j}}(s) \frac{s-u_{i}^{S}}{\sigma_{S}} d s \\
& =u_{i}^{Y} u_{i}^{Z}+\rho_{i} \sigma_{Y} \int_{c}^{\infty} \frac{1}{\sqrt{2 \pi} \sigma_{S}} \exp \left[\frac{\left(s-u_{i}^{S}\right)^{2}}{2 \sigma_{S}^{2}}\right] \frac{s-u_{i}^{S}}{\sigma_{S}} d s \\
& =u_{i}^{Y} u_{i}^{Z}+\rho_{i} \sigma_{Y} \frac{1}{\sqrt{2 \pi}} \exp \left(-\frac{\left(c-u_{i}^{S}\right)^{2}}{2 \sigma_{S}^{2}}\right),
\end{aligned}
$$

$$
\begin{align*}
\widetilde{\rho}_{i} & =\rho_{i\left(Y_{i j}, Z_{i j}\right)} \\
& =\frac{\operatorname{cov}\left(Y_{i j}, Z_{i j}\right)}{\sqrt{\operatorname{var}\left(Y_{i j}\right) \operatorname{var}\left(Z_{i j}\right)}} \\
& =\frac{\rho_{i} \sigma_{Y} \frac{1}{\sqrt{2 \pi}} \exp \left(-\frac{\left(c-u_{i}^{S}\right)^{2}}{2 \sigma_{S}^{2}}\right)}{\sigma_{Y} \sqrt{u_{i}^{Z}\left(1-u_{i}^{Z}\right)}} \\
& =\frac{\rho_{i} \frac{1}{\sqrt{2 \pi}} \exp \left(-\frac{\left(c-u_{i}^{S}\right)^{2}}{2 \sigma_{S}^{2}}\right)}{\sqrt{u_{i}^{Z}\left(1-u_{i}^{Z}\right)}} . \tag{9.5}
\end{align*}
$$

From the formula (9.4) we will have:

$$
u_{i}^{S}=c-\sigma_{S} \Phi^{-1}\left(1-g\left(d_{i}, \theta_{Z}\right)\right)=c-\sigma_{S} \Phi^{-1}\left(1-u_{i}^{Z}\right)
$$

and we will get

$$
1-u_{i}^{Z}=\Phi\left(\frac{c-u_{i}^{S}}{\sigma_{S}}\right)
$$

by plug the above formula into formula (9.5), we will have following:

$$
\begin{align*}
\widetilde{\rho}_{i} & =\rho_{i\left(Y_{i j}, Z_{i j}\right)} \\
& =\frac{\rho_{i} \phi\left(\frac{c-u_{i}^{S}}{\sigma_{S}}\right)}{\sqrt{\Phi\left(\frac{c-u_{i}^{S}}{\sigma_{S}}\right)\left(1-\Phi\left(\frac{c-u_{i}^{S}}{\sigma_{S}}\right)\right)}} \\
& =\rho_{i} \frac{\phi\left(\frac{c-u_{i}^{S}}{\sigma_{S}}\right)}{\sqrt{\Phi\left(\frac{c-u_{i}^{S}}{\sigma_{S}}\right)\left(1-\Phi\left(\frac{c-u_{i}^{S}}{\sigma_{S}}\right)\right)}} \tag{9.6}
\end{align*}
$$

The maximum value of right side of formula (9.6) can be solved by the optimize function in R. The maximum value is 0.798 . Thus the maximum value of $\widetilde{\rho}_{i}$ is $0.7978846 \rho_{i}$. The range for $\rho_{i}$ is $[-1,1]$ and $\widetilde{\rho}_{i}$ should be in the range of $[-0.798,0.798]$. i.e., $\left|\widetilde{\rho}_{i}\right| \leq$ $0.798\left|\rho_{i}\right|$. This confirmed the result from Shih and Huang (1992).

Hence we can compute the log-likelihood of observed data $\left(Y_{i j}, Z_{i j}\right)$ by plugging (9.4) and (9.5) into (9.3). For example, assume there are 7 dose levels, $Y_{i j}$ follows Emax model and $Z_{i j}$ follows probit model. $\rho_{i}$ represents the correlation between the continuous efficacy variable and latent continuous safety variable.

Therefore by using the formulas (9.4) and (9.5) we have the following:

$$
u_{i}^{\curlyvee}=e_{0}+\frac{e_{\max } \text { dose }_{i}}{e d_{50}+\text { dose }_{i}},
$$

and

$$
u_{i}^{Z}=\Phi\left(\alpha+\beta d_{i}\right) .
$$

This implies

$$
u_{i}^{S}=c+\sigma_{S}\left(\alpha+\beta d_{i}\right)
$$

and

$$
\rho_{i}=\frac{\sqrt{2 \pi} \widetilde{\rho}_{i} u_{i}^{Z}\left(1-u_{i}^{Z}\right)}{\exp \left(-\frac{\left(c-u_{i}^{S}\right)^{2}}{2 \sigma_{S}^{2}}\right)} .
$$

where $\mathrm{i}=1, \ldots, 7$.

Thus the log-likelihood can be written as follows:

$$
\begin{aligned}
& L=\sum_{i=1}^{k} \sum_{j=1}^{n_{i}} \log \left\{f\left(y_{i j}, z_{i j}\right)\right\} \\
& =\sum_{i=1}^{7} \sum_{j=1}^{n_{i}}\left\{\log \left(\frac{1}{\sqrt{2 \pi \sigma_{Y}^{2}}}\right)\left[-\frac{1}{2 \sigma_{Y}^{2}}\left(y_{i j}-u_{i}^{Y}\right)\right]\right\} \\
& +\sum_{i=1}^{7} \sum_{j=1}^{n_{i}}\left\{z_{i j} \log \left[1-\Phi\left(\frac{-\sigma_{S}\left(\alpha+\beta d_{i}\right)-\frac{\sqrt{2 \pi \tilde{p}_{i}} u_{i}^{Z}\left(1-u_{i}^{Z}\right) \sigma_{S}}{\sigma_{\curlyvee} \exp \left(-\frac{\left(c-u_{i}^{S}\right)^{2}}{2 \sigma_{S}^{2}}\right)}\left(y_{i j}-u_{i}^{\gamma}\right)}{\sigma_{S} \sqrt{1-\frac{2 \pi \widetilde{\rho}_{i}^{2}\left(u_{i}^{Z}\left(1-u_{i}^{Z}\right)\right)^{2}}{\exp ^{2}\left(-\frac{\left(c-u_{i}^{S}\right)^{2}}{2 \sigma_{S}^{2}}\right)}}}\right)\right]\right\} \\
& +\sum_{i=1}^{7} \sum_{j=1}^{n_{i}}\left\{( 1 - z _ { i j } ) \operatorname { l o g } \left[\Phi\left(\frac{-\sigma_{S}\left(\alpha+\beta d_{i}\right)-\frac{\sqrt{2 \pi} \tilde{r}_{i} u_{i}^{Z}\left(1-u_{i}^{Z}\right) \sigma_{S}}{\sigma_{\Upsilon} \exp \left(-\frac{\left(c-u_{i}^{S}\right)^{2}}{2 \sigma_{S}^{2}}\right)}\left(y_{i j}-u_{i}^{Y}\right)}{\left.\left.\sigma_{S} \sqrt{1-\frac{2 \pi \tilde{\rho}_{i}^{2}\left(u_{i}^{Z}\left(1-u_{i}^{Z}\right)\right)^{2}}{\exp ^{2}\left(-\frac{\left(c-u_{i}^{S}\right)^{2}}{2 \sigma_{S}^{2}}\right)}}\right)\right]}\right] .\right.\right.
\end{aligned}
$$

### 9.2.2 Method 2: The latent variable threshold c is unknown

Under some instances there is no information about the latent continuous safety variable, then the likelihood can not be calculated from the approach described above. The likelihood may be derived purely on the information about the observed continuous efficacy and discrete safety outcome instead. The joint density of $\left(Y_{i j}, Z_{i j}\right)$ can be written as

$$
f_{Y_{i j}, Z_{i j}}\left(y_{i j}, z_{i j}\right)=f_{Z_{i j}}\left(z_{i j}\right) f_{Y_{i j} \mid Z_{i j}}\left(y_{i j} \mid z_{i j}\right)
$$

We assume that

$$
f_{Z_{i j}}\left(z_{i j}\right)=\left(g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)\right)^{Z_{i j}}\left(1-g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)\right)^{1-Z_{i j}}
$$

and we assume that the conditional distribution of is normal,

$$
\begin{aligned}
f_{Y_{i j} \mid Z_{i j}}\left(y_{i j} \mid z_{i j}\right) & =\frac{1}{\sqrt{2 \pi \sigma_{Y \mid\left(Z=z_{i j}\right)}^{2}}} \exp \left\{-\frac{1}{2 \sigma_{Y \mid\left(Z=z_{i j}\right)}^{2}}\left[y_{i j}-u_{i}^{Y}-\gamma_{i}\left(z_{i j}-u_{i}^{Z}\right)\right]\right\} \\
& =\frac{\exp \left\{-\frac{1}{2 \sigma_{Y \mid\left(Z=z_{i j}\right)}^{2}}\left[y_{i j}-f\left(d_{i}, \boldsymbol{\theta}_{Y}\right)-\gamma_{i}\left(z_{i j}-g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)\right)\right]\right\}}{\sqrt{2 \pi \sigma_{Y \mid\left(Z=z_{i j}\right)}^{2}}}
\end{aligned}
$$

where $\gamma_{i}$ is a parameter for regression of $Y_{i j}$ on $Z_{i j}$ for dose $i$, i.e,

$$
E\left(Y_{i j} \mid Z_{i j}\right)=u_{i}^{Y}+\gamma_{i}\left(Z_{i j}-u_{i}^{Z}\right)
$$

and

$$
\begin{aligned}
E\left(Y_{i j}\right) & =E\left(E\left(Y_{i j} \mid Z_{i j}\right)\right) \\
& =E\left(u_{i}^{Y}+\gamma_{i}\left(Z_{i j}-u_{i}^{Z}\right)\right) \\
& =u_{i}^{Y}+\gamma_{i} E\left(Z_{i j}-u_{i}^{Z}\right) \\
& =u_{i}^{Y} .
\end{aligned}
$$

Since

$$
\begin{aligned}
\operatorname{var}\left(Y_{i j}\right) & =E\left(\operatorname{var}\left(Y_{i j} \mid Z_{i j}\right)\right)+\operatorname{var}\left(E\left(Y_{i j} \mid Z_{i j}\right)\right) \\
& =E\left(\sigma_{Y \mid\left(Z=Z_{i j}\right)}^{2}\right)+\operatorname{var}\left(u_{i}^{Y}+\gamma_{i}\left(Z_{i j}-u_{i}^{Z}\right)\right) \\
& =\sigma_{Y \mid(Z=0)}^{2} P\left(Z_{i j}=0\right)+\sigma_{Y \mid Z=1}^{2} P\left(Z_{i j}=1\right)+\gamma_{i}^{2} \operatorname{var}\left(Z_{i j}-u_{i}^{Z}\right) \\
& =\sigma_{Y \mid(Z=0)}^{2}\left(1-u_{i}^{Z}\right)+\sigma_{Y \mid(Z=1)}^{2} u_{i}^{Z}+\gamma_{i}^{2} u_{i}^{Z}\left(1-u_{i}^{Z}\right),
\end{aligned}
$$

and

$$
\begin{gathered}
E\left(Y_{i j} Z_{i j}\right)=E\left(E\left(Y_{i j} Z_{i j} \mid Z_{i j}\right)\right) \\
=E\left(Z_{i j} E\left(Y_{i j} \mid Z_{i j}\right)\right) \\
=E\left(Z_{i j}\left(u_{i}^{Y}+\gamma_{i}\left(Z_{i j}-u_{i}^{Z}\right)\right)\right) \\
\left.=E\left(Z_{i j} u_{i}^{Y}+\gamma_{i} Z_{i j}\left(Z_{i j}-u_{i}^{Z}\right)\right)\right) \\
=u_{i}^{Y} E Z_{i j}+\gamma_{i} E\left(Z_{i j}^{2}-Z_{i j} u_{i}^{Z}\right) \\
\\
=u_{i}^{Y} E Z_{i j}+\gamma_{i}\left(u_{i}^{Z}-u_{i}^{Z} E Z_{i j}\right) \\
\\
=u_{i}^{Y} u_{i}^{Z}+\gamma_{i} u_{i}^{Z}\left(1-u_{i}^{Z}\right), \\
=\frac{\widetilde{\rho}_{i}=}{\operatorname{corr}\left(Y_{i j,} Z_{i j}\right)} \frac{\operatorname{cov}\left(Y_{i j}, Z_{i j}\right)}{\sqrt{\operatorname{var}\left(Y_{i j}\right) \operatorname{var}\left(Z_{i j}\right)}} \\
=\frac{E\left(Y_{i j} Z_{i j}\right)-E Y_{i j} E Z_{i j}}{\sqrt{\left(\sigma_{Y \mid(Z=0)}^{2}\left(1-u_{i}^{Z}\right)+\sigma_{Y \mid(Z=1)}^{2} u_{i}^{Z}+\gamma_{i}^{2} u_{i}^{Z}\left(1-u_{i}^{Z}\right)\right) u_{i}^{Z}\left(1-u_{i}^{Z}\right)}} \\
=\frac{u_{i}^{Y} u_{i}^{Z}+\gamma_{i} u_{i}^{Z}\left(1-u_{i}^{Z}\right)-u_{i}^{Y} u_{i}^{Z}}{\sqrt{\left(\sigma_{Y \mid(Z=0)}^{2}\left(1-u_{i}^{Z}\right)+\sigma_{Y \mid(Z=1)}^{2} u_{i}^{Z}+\gamma_{i}^{2} u_{i}^{Z}\left(1-u_{i}^{Z}\right)\right) u_{i}^{Z}\left(1-u_{i}^{Z}\right)}} \\
=
\end{gathered}
$$

this implies

$$
\gamma_{i}^{2} u_{i}^{Z}\left(1-u_{i}^{Z}\right)=\widetilde{\rho}_{i}^{2}\left(\sigma_{Y \mid(Z=0)}^{2}\left(1-u_{i}^{Z}\right)+\sigma_{Y \mid(Z=1)}^{2} u_{i}^{Z}\right)+\gamma_{i}^{2} u_{i}^{Z}\left(1-u_{i}^{Z}\right) \widetilde{\rho}_{i}^{2}
$$

and we have

$$
\gamma_{i}=\widetilde{\rho}_{i} \sqrt{\frac{\left(\sigma_{Y \mid(Z=0)}^{2}\left(1-u_{i}^{Z}\right)+\sigma_{Y \mid(Z=1)}^{2} u_{i}^{Z}\right)}{u_{i}^{Z}\left(1-u_{i}^{Z}\right)\left(1-\widetilde{\rho}_{i}^{2}\right)}} .
$$

Hence we can write the log-likelihood as follows:

$$
\begin{aligned}
L= & \sum_{i=1}^{k} \sum_{j=1}^{n_{i}} \log \left\{f_{Y_{i j}, Z_{i j}}\left(y_{i j}, z_{i j}\right)\right\} \\
= & \sum_{i=1}^{k} \sum_{j=1}^{n_{i}} \log \left\{f_{Z_{i j}}\left(z_{i j}\right) f_{Y_{i j} \mid Z_{i j}}\left(y_{i j} \mid z_{i j}\right)\right\} \\
= & \sum_{i=1}^{k} \sum_{j=1}^{n_{i}} \log \left\{\frac{1}{\left.\sqrt{2 \pi \sigma_{Y \mid\left(Z=z_{i j}\right)}^{2}}\left(g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)\right)^{z_{i j}}\left(1-g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)\right)^{1-z_{i j}}\right\}}\right. \\
& +\sum_{i=1}^{k} \sum_{j=1}^{n_{i}}\left\{-\frac{1}{2 \sigma_{Y \mid\left(Z=z_{i j}\right)}^{2}}\left[y_{i j}-f\left(d_{i}, \boldsymbol{\theta}_{Y}\right)-\gamma_{i}\left(z_{i j}-g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)\right)\right]\right\} \\
= & \sum_{i=1}^{k} \sum_{j=1}^{n_{i}} \log \left\{\frac{\left(u_{i}^{Z}\right)^{z_{i j}}\left(1-u_{i}^{Z}\right)^{1-z_{i j}}}{\left.\sqrt{2 \pi \sigma_{Y \mid\left(Z=z_{i j}\right)}^{2}}\right\}+}\right. \\
& \sum_{i=1 j=1}^{k} \sum^{n_{i}}\left\{-\frac{\left[y_{i j}-u_{i}^{Y}-\widetilde{\rho}_{i} \sqrt{\frac{\left(\sigma_{Y \mid(Z=0)}^{2}\left(1-u_{i}^{Z}\right)+\sigma_{Y \mid(Z=1)}^{2} u_{i}^{Z}\right)}{u_{i}^{Z}\left(1-u_{i}^{Z}\right)\left(1-\widetilde{\rho}_{i}^{2}\right)}}\left(z_{i j}-u_{i}^{Z}\right)\right]}{2 \sigma_{Y \mid\left(Z=z_{i j}\right)}^{2}}\right\} .
\end{aligned}
$$

If we assume $\sigma_{Y \mid(Z=1)}^{2}=\sigma_{Y \mid(Z=0)}^{2}=\sigma_{c}^{2}\left(\right.$ Note: $\left.\sigma_{c}^{2} \neq \sigma_{Y}^{2}\right)$, the log-likelihood can be further simplified as follow:

$$
\begin{align*}
L= & \sum_{i=1}^{k} \sum_{j=1}^{n_{i}} \log \left\{f_{Z_{i j}}\left(z_{i j}\right) f_{Y_{i j} \mid Z_{i j}}\left(y_{i j} \mid z_{i j}\right)\right\} \\
= & \sum_{i=1}^{k} \sum_{j=1}^{n_{i}} \log \left\{\frac{1}{\sqrt{2 \pi \sigma_{c}^{2}}}\left(u_{i}^{Z}\right)^{z_{i j}}\left(1-u_{i}^{Z}\right)^{1-z_{i j}}\right\} \\
& -\sum_{i=1}^{k} \sum_{j=1}^{n_{i}}\left\{\frac{1}{2 \sigma_{c}^{2}}\left[y_{i j}-u_{i}^{Y}\right]\right\} \\
& +\sum_{i=1 j=1}^{k} \sum^{n_{i}}\left\{\frac{\left[\widetilde{\rho}_{i} \sqrt{\frac{1}{u_{i}^{Z}\left(1-u_{i}^{Z}\right)\left(1-\widetilde{\rho}_{i}^{2}\right)}}\left(z_{i j}-u_{i}^{Z}\right)\right]}{2 \sigma_{c}}\right\} \tag{9.7}
\end{align*}
$$

For example, assume there are 7 dose levels and if $Y_{i j}$ follows Emax model and $Z_{i j}$ follows probit model. $\widetilde{\rho}_{i}$ represents the correlation between the continuous efficacy variable and discrete safety variable for each dose level $i$. The parameters to be estimated are: $\mathrm{e}_{0}, \mathrm{e}_{\max }$, $\operatorname{ed}_{50}, \alpha, \beta, \sigma_{y}^{2}, \widetilde{\rho}_{i}$. The log-likelihood can be computed by plugging the estimated parameters for the observed data into the formula (9.7).

Since

$$
u_{i}^{Y}=e_{0}+\frac{e_{\max } \text { dose }_{i}}{e d_{50}+\text { dose }_{i}},
$$

and

$$
u_{i}^{Z}=\Phi\left(\alpha+\beta d_{i}\right),
$$

we have

$$
\begin{aligned}
L= & \sum_{i=1}^{7} \sum_{j=1}^{n_{i}} \log \left\{\frac{1}{\sqrt{2 \pi \sigma_{c}^{2}}}\left(\Phi\left(\alpha+\beta d_{i}\right)\right)^{z_{i j}}\left(1-\Phi\left(\alpha+\beta d_{i}\right)\right)^{1-z_{i j}}\right\} \\
& -\frac{1}{2 \sigma_{Y}^{2}}\left[y_{i j}-\left(e_{0}+\frac{e_{\max } d o s e_{i}}{e d_{50}+d o s e_{i}}\right)\right] \\
& +\left\{\frac{\left[\widetilde{\rho}_{i} \sqrt{\frac{1}{\Phi\left(\alpha+\beta d_{i}\right)\left(1-\Phi\left(\alpha+\beta d_{i}\right)\right)\left(1-\tilde{\rho}_{i}^{2}\right)}}\left(z_{i j}-\Phi\left(\alpha+\beta d_{i}\right)\right)\right]}{2 \sigma_{c}}\right\} .
\end{aligned}
$$

### 9.3 Estimation of joint nonlinear bivariate continuous and dis-

 crete model
### 9.3.1 Estimation by IRGLS, ELS and QELS methods

Given a bivariate variable $\left[\begin{array}{l}Y_{i j} \\ Z_{i j}\end{array}\right]$ which represents the efficacy response $Y_{i j}$ and the safety response $Z_{i j}$. Here $Y_{i j}$ and $Z_{i j}$ can be continuous or discrete responses. The
derivation below apply to different types of responses which include bivariate continuous outcomes, bivariate discrete outcomes and bivariate continuous and discrete outcomes. The generalized nonlinear joint model can be written as in (9.1).

$$
\begin{aligned}
& \text { Let } \\
& \qquad u_{i}(\boldsymbol{\theta})=\left[\begin{array}{c}
u_{i}^{Y} \\
u_{i}^{Z}
\end{array}\right]=\left[\begin{array}{c}
E\left(Y_{i j}\right) \\
E\left(Z_{i j}\right)
\end{array}\right]=\left[\begin{array}{l}
f\left(d_{i}, \boldsymbol{\theta}_{Y}\right) \\
g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)
\end{array}\right], i=1, \ldots k, j=1, \ldots . n_{i}
\end{aligned}
$$

where $\boldsymbol{\theta}^{T}=\left(\boldsymbol{\theta}_{Y}^{T}, \boldsymbol{\theta}_{Z}^{T}\right)$, which refers to the vector of model parameters, $i$ to the dose group ( $i=1$ corresponds to placebo), and $j$ to the patient within dose group $i$, and

$$
\boldsymbol{\Sigma}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})=\operatorname{Var}\left(\left[\begin{array}{l}
Y_{i j} \\
Z_{i j}
\end{array}\right]\right) .
$$

The parameters to be estimated are $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$. The following derivations will illustrate how to apply the IRGLS, ELS and QELS to the above data. Here we need to mention that we do not use the full likelihood we derive above for the estimation. We will discuss this later in the discussion and future direction section for the full likelihood estimation approach.

Furthermore, let $\omega_{i j}=\left(Y_{i j}, Z_{i j}\right)^{T}, \gamma_{i j}=\left(Y_{i j}{ }^{2}, Z_{i j}{ }^{2}, Y_{i j} Z_{i j}\right)^{T}$, and

$$
\boldsymbol{\varphi}_{i j}=\binom{\boldsymbol{\omega}_{i j}}{\gamma_{i j}}=\left(\begin{array}{c}
Y_{i j} \\
Z_{i j} \\
Y_{i j}^{2} \\
Z_{i j}^{2} \\
Y_{i j} Z_{i j}
\end{array}\right)
$$

Assume that the first four moments of $\boldsymbol{\omega}_{i j}$ exist and depend on the parameter $(\boldsymbol{\theta}, \boldsymbol{\alpha})$, let

$$
\boldsymbol{f}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})=E\left(\boldsymbol{\varphi}_{i j}\right)=\binom{\mathbf{u}_{i}(\boldsymbol{\theta})}{\boldsymbol{v}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}=\left(\begin{array}{c}
u_{i}^{Y}\left(\boldsymbol{\theta}_{Y}\right) \\
u_{i}^{Z}\left(\boldsymbol{\theta}_{Z}\right) \\
E\left(Y_{i j}{ }^{2}\right) \\
E\left(Z_{i j}{ }^{2}\right) \\
E\left(Y_{i j} Z_{i j}\right)
\end{array}\right),
$$

and

$$
\operatorname{var}\left(\boldsymbol{\varphi}_{i j}\right)=\boldsymbol{\Psi}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})=\left(\begin{array}{cc}
\boldsymbol{\Sigma}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha}) & \boldsymbol{\Theta}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha}) \\
\boldsymbol{\Theta}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})^{T} & \boldsymbol{\Gamma}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})
\end{array}\right)
$$

where $\boldsymbol{u}_{i}(\boldsymbol{\theta})=E\left(\boldsymbol{\omega}_{i j}\right)=\left[\begin{array}{c}E\left(Y_{i j}\right) \\ E\left(Z_{i j}\right)\end{array}\right]=\left[\begin{array}{c}u_{i}^{Y} \\ u_{i}^{Z}\end{array}\right]$,

$$
\boldsymbol{v}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})=E\left(\boldsymbol{\gamma}_{i j}\right)=\left[\begin{array}{c}
E\left(Y_{i j}^{2}\right) \\
E\left(Z_{i j}^{2}\right) \\
E\left(Y_{i j} Z_{i j}\right)
\end{array}\right]
$$

$\boldsymbol{\Sigma}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})=\operatorname{var}\left(\boldsymbol{\omega}_{i j}\right), \boldsymbol{\Theta}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})=\operatorname{cov}\left(\boldsymbol{\omega}_{i j}, \boldsymbol{\gamma}_{i j}\right)$ and $\boldsymbol{\Gamma}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})=\operatorname{var}\left(\boldsymbol{\gamma}_{i j}\right)$. We can jointly estimate the parameter $(\boldsymbol{\theta}, \boldsymbol{\alpha})$ based on the following nonlinear model

$$
\boldsymbol{\varphi}_{i j}=\boldsymbol{f}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})+\boldsymbol{\epsilon}_{i j}
$$

where $E\left(\boldsymbol{\epsilon}_{i j}\right)=\mathbf{0}$ and $\operatorname{var}\left(\boldsymbol{\epsilon}_{i j}\right)=\boldsymbol{\Psi}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})$. The linearization method of Vonesh et al. (2001) can be applied. To be specific and complete, the following derives the details though the general formulas that have been introduced in the previous literature review section.

$$
\text { Let } \boldsymbol{\tau}=(\boldsymbol{\theta}, \boldsymbol{\alpha}) \text { and } \boldsymbol{t}^{T}=\widehat{\boldsymbol{\tau}}^{0^{T}}=\left(\widehat{\boldsymbol{\theta}}^{0 T}, \widehat{\boldsymbol{\alpha}}^{0 T}\right) \text { be an initial estimate of } \boldsymbol{\tau}^{T}=\left(\boldsymbol{\theta}^{T}, \boldsymbol{\alpha}^{T}\right)
$$

and assume $t$ is in the interior of parameter space of $\boldsymbol{\tau}$. By applying the usual Gauss-

Newton algorithm for the nonlinear regression, estimation may be carried out by taking a first-order Taylor series expansion of $f_{i}(\boldsymbol{\tau})$ about $\boldsymbol{\tau}=\boldsymbol{t}$ yielding the approximation $\boldsymbol{\varphi}_{i j}=$ $\boldsymbol{f}_{i}(\boldsymbol{\tau})+\boldsymbol{X}_{i t}(\boldsymbol{\tau}-\boldsymbol{t})+\boldsymbol{\xi}_{i j}$, we can rewrite in terms of the linear model

$$
\begin{equation*}
\boldsymbol{\varphi}_{i j}^{*}=\boldsymbol{X}_{i t} \boldsymbol{\tau}+\boldsymbol{\xi}_{i j,} \tag{9.8}
\end{equation*}
$$

where $\boldsymbol{X}_{i t}=\partial \boldsymbol{f}_{i}(\boldsymbol{\tau}) /\left.\partial \boldsymbol{\tau}^{T}\right|_{\tau=t}$ and $\boldsymbol{\varphi}_{i j}^{*}=\boldsymbol{\varphi}_{i j}-f_{i}(\boldsymbol{t})+\boldsymbol{X}_{i t} \boldsymbol{t}$, based on this linear model, an estimated generalized least squares estimate (EGLS) of $\tau$ is given by

$$
\begin{equation*}
\widehat{\boldsymbol{\tau}}=\left(\sum_{i} \sum_{j} \boldsymbol{X}_{i t}^{T} \boldsymbol{\Psi}_{i}^{-1}(\boldsymbol{t}) \boldsymbol{X}_{i t}\right)^{-1} \sum_{i} \sum_{j} \boldsymbol{X}_{i t}^{T} \boldsymbol{\Psi}_{i}^{-1}(\boldsymbol{t}) \boldsymbol{\varphi}_{i j}^{*}, \tag{9.9}
\end{equation*}
$$

where $\boldsymbol{\Psi}_{i j}(\boldsymbol{t})$ is the assumed covariance matrix of $\boldsymbol{\xi}_{i j}$ evaluated at $\boldsymbol{\tau}=\boldsymbol{t}$.
Iteratively reweighted generalized least squares (IRGLS) entails iterating between (9.8) and (9.9) by setting $\tau=t$ in (9.8) and then using (9.9) to obtain an updated estimate of $\boldsymbol{\tau}$. By repeating this process, obtain a sequence of one-step Gauss-Newton estimators, $\left\{\widehat{\tau}^{k} \mid \mathrm{k}=1,2, \ldots\right\}$, which, as $\mathrm{k} \rightarrow \infty$,yields a solution to the set on nonlinear "normal" estimating equations,

$$
\begin{equation*}
\boldsymbol{U}(\boldsymbol{\tau})=\sum_{i} \sum_{j}\left\{\boldsymbol{X}_{i}(\boldsymbol{\tau})^{T} \boldsymbol{\Psi}_{i}(\boldsymbol{\tau})^{-1}\left(\boldsymbol{\varphi}_{i j}-\boldsymbol{f}_{i}(\boldsymbol{\tau})\right)\right\}=0, \tag{9.10}
\end{equation*}
$$

where $X_{i}(\tau)=\partial f_{i}(\tau) / \partial \tau^{T}$. When fully iterated, the model-based asympotic variancecovariance matrix of the IRGLS estimate, $\widehat{\tau}$ is estimated by the inverse of the expected information matrix evaluated at the final estimate, i.e,

$$
\begin{equation*}
\widehat{\Omega}(\widehat{\boldsymbol{\tau}})=\left(\sum_{i} \sum_{j} \boldsymbol{X}_{i}(\boldsymbol{\tau})^{T} \boldsymbol{\Psi}_{i}(\boldsymbol{\tau})^{-1} \boldsymbol{X}_{i}(\boldsymbol{\tau})\right)^{-1} \tag{9.11}
\end{equation*}
$$

To safeguard against mis-specification of $\boldsymbol{\Psi}_{i}(\boldsymbol{\tau})$, one can use a robust estimator of the variance of $\widehat{\tau}$, which is given by the empirical "sandwich" estimator as follows:

$$
\begin{equation*}
\widehat{\boldsymbol{\Omega}}_{R}(\widehat{\boldsymbol{\tau}})=\widehat{\Omega}(\widehat{\boldsymbol{\tau}})\left(\sum_{i} \boldsymbol{X}_{i}(\boldsymbol{\tau})^{T} \boldsymbol{\Psi}_{i}(\boldsymbol{\tau})^{-1} \boldsymbol{e}_{i j} \boldsymbol{e}_{i j}^{T} \boldsymbol{\Psi}_{i}(\boldsymbol{\tau})^{-1} \boldsymbol{X}_{i}(\boldsymbol{\tau})\right) \widehat{\Omega}(\widehat{\boldsymbol{\tau}}), \tag{9.12}
\end{equation*}
$$

where $\boldsymbol{e}_{i j}=\boldsymbol{\varphi}_{i j}-\boldsymbol{f}_{i}(\widehat{\boldsymbol{\tau}})$.
In order to compute (9.11), we need to calculate the third- and fourth-order moments in $\Psi_{i}$.

If we let

$$
\begin{gathered}
\boldsymbol{s}_{i j}(\boldsymbol{\theta})=\operatorname{Vec}\left(\left(\boldsymbol{\omega}_{i j}-\mathbf{u}_{i}(\boldsymbol{\theta})\right)\left(\boldsymbol{\omega}_{i j}-\mathbf{u}_{i}(\boldsymbol{\theta})^{T}\right)=\left[\begin{array}{c}
\left(Y_{i j}-u_{i}^{\Upsilon}\right)^{2} \\
\left(Y_{i j}-u_{i}^{Y}\right)\left(Z_{i j}-u_{i}^{Z}\right) \\
\left(Y_{i j}-u_{i}^{Y}\right)\left(Z_{i j}-u_{i}^{Z}\right) \\
\left(Z_{i j}-u_{i}^{Z}\right)^{2}
\end{array}\right],\right. \\
\sigma_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})=\operatorname{Vec}\left(\boldsymbol{\Sigma}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})\right)=E\left[\boldsymbol{S}_{i j}(\boldsymbol{\theta})\right]=\left[\begin{array}{c}
\sigma_{Y, i}^{2} \\
\sigma_{Y Z, i} \\
\sigma_{Y Z, i} \\
\sigma_{Z, i}^{2}
\end{array}\right],
\end{gathered}
$$

and

$$
\boldsymbol{B}=\left[\begin{array}{llll}
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0
\end{array}\right]
$$

where $\operatorname{Vec}($.$) is the matrix operator that creates a column vector from a matrix \mathrm{A}$ by simply stacking the column vectors of A below one another, then under normality (Vonesh and Chinchilli, 1997), we have

$$
\boldsymbol{\Theta}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})=\operatorname{cov}\left(\boldsymbol{\omega}_{i j}, \gamma_{i j}\right)=\boldsymbol{\Sigma}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha}) \boldsymbol{T}_{i}(\boldsymbol{\theta})^{T},
$$

and

$$
\boldsymbol{\Gamma}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})=\operatorname{var}\left(\gamma_{i j}\right)=\boldsymbol{T}_{i}(\boldsymbol{\theta}) \boldsymbol{\Sigma}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha}) \boldsymbol{T}_{i}(\boldsymbol{\theta})^{T}+\boldsymbol{B} \operatorname{Var}\left[\boldsymbol{S}_{i j}(\boldsymbol{\theta})\right] \boldsymbol{B}^{T},
$$

where

$$
\begin{aligned}
\boldsymbol{T}_{i}(\boldsymbol{\theta}) & =\boldsymbol{B}\left\{\left[\boldsymbol{u}_{i}(\boldsymbol{\theta}) \otimes \boldsymbol{I}_{2}\right]+\left[\boldsymbol{I}_{2} \otimes u_{i}(\boldsymbol{\theta})\right]\right\} \\
& =\left[\begin{array}{llll}
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0
\end{array}\right]\left\{\left[\begin{array}{ll}
u_{i}^{Y} & 0 \\
u_{i}^{Z} & 0 \\
0 & u_{i}^{Y} \\
0 & u_{i}^{Z}
\end{array}\right]+\left[\begin{array}{ll}
u_{i}^{Y} & 0 \\
u_{i}^{Z} & 0 \\
0 & u_{i}^{Y} \\
0 & u_{i}^{Z}
\end{array}\right]\right\} \\
& =2\left[\begin{array}{lll}
u_{i}^{Y} & 0 \\
0 & u_{i}^{Z} \\
u_{i}^{Z} & 0
\end{array}\right]
\end{aligned}
$$

Hence we have

$$
\begin{aligned}
\boldsymbol{\Theta}_{i}(\boldsymbol{\theta}, \alpha) & =\left[\begin{array}{cc}
\sigma_{Y, i}^{2} & \sigma_{Y Z, i} \\
\sigma_{Y Z, i} & \sigma_{Z, i}^{2}
\end{array}\right] 2\left[\begin{array}{ccc}
u_{i}^{Y} & 0 & u_{i}^{Z} \\
0 & u_{i}^{Z} & 0
\end{array}\right] \\
& =2\left[\begin{array}{ccc}
\sigma_{Y, i}^{2} u_{i}^{Y} & \sigma_{Y Z, i} u_{i}^{Z} & \sigma_{Y, i}^{2} u_{i}^{Z} \\
\sigma_{Y Z, i} u_{i}^{Y} & \sigma_{Z, i}^{2} u_{i}^{Z} & \sigma_{Y Z, i} u_{i}^{Z}
\end{array}\right]
\end{aligned}
$$

Since

$$
\begin{aligned}
\operatorname{Var}\left[\boldsymbol{S}_{i j}(\boldsymbol{\theta})\right]= & {\left[\boldsymbol{I}_{4}+\boldsymbol{I}_{(2,2)}\right]\left[\boldsymbol{\Sigma}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha}) \otimes \boldsymbol{\Sigma}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})\right] } \\
= & 2\left[\begin{array}{llll}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{array}\right]\left[\begin{array}{cc}
\sigma_{Y, i}^{2} & \sigma_{Y Z, i} \\
\sigma_{Y Z, i} & \sigma_{Z, i}^{2}
\end{array}\right] \otimes\left[\begin{array}{cc}
\sigma_{Y, i}^{2} & \sigma_{Y Z, i} \\
\sigma_{Y Z, i} & \sigma_{Z, i}^{2}
\end{array}\right] \\
= & 2\left[\begin{array}{llll}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{array}\right]\left[\begin{array}{cccc}
\sigma_{Y, i}^{4} & \sigma_{Y, i}^{2} \sigma_{Y Z, i} & \sigma_{Y, i}^{2} \sigma_{Y Z, i} & \sigma_{Y Z, i}^{2} \\
\sigma_{Y, i}^{2} \sigma_{Y Z, i} & \sigma_{Y, i}^{2} \sigma_{Z, i}^{2} & \sigma_{Y Z, i}^{2} & \sigma_{Z, i}^{2} \sigma_{Y Z, i} \\
\sigma_{Y, i}^{2} \sigma_{Y Z, i} & \sigma_{Y Z, i}^{2} & \sigma_{Y, i}^{2} \sigma_{Z, i}^{2} & \sigma_{Z, i j}^{2} \sigma_{Y Z, i} \\
\sigma_{Y Z, i}^{2} & \sigma_{Z, i}^{2} \sigma_{Y Z, i} & \sigma_{Z, i}^{2} \sigma_{Y Z, i} & \sigma_{Z, i}^{4}
\end{array}\right] \\
= & 2\left[\begin{array}{ccc}
\sigma_{Y, i}^{4} & \sigma_{Y, i}^{2} \sigma_{Y Z, i} & \sigma_{Y, i}^{2} \sigma_{Y Z, i} \\
\sigma_{Y, i}^{2} \sigma_{Y Z, i}^{2} & \sigma_{Y Z, i}^{2} \\
\sigma_{Y, i}^{2} \sigma_{Y Z, i}^{2} & \sigma_{Y Z, i}^{2} & \sigma_{Z, i}^{2} \sigma_{Y Z, i} \\
\sigma_{Y Z, i}^{2} & \sigma_{Y, i}^{2} \sigma_{Z, i}^{2} & \sigma_{Z, i j}^{2} \sigma_{Y Z, i} \\
\sigma_{Y Z, i}^{2} & \sigma_{Z, i}^{2} \sigma_{Y Z, i} & \sigma_{Z, i}^{2} \sigma_{Y Z, i} \\
\sigma_{Z, i}^{4}
\end{array}\right],
\end{aligned}
$$

we have

$$
\begin{aligned}
& \boldsymbol{\Gamma}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})=\operatorname{var}\left(\boldsymbol{\gamma}_{i j}\right)=\boldsymbol{T}_{i}(\boldsymbol{\theta}) \boldsymbol{\Sigma}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha}) \boldsymbol{T}_{i}(\boldsymbol{\theta})^{T}+\boldsymbol{B V a r}\left[\boldsymbol{S}_{i j}(\boldsymbol{\theta})\right] \boldsymbol{B}^{T} \\
& =2\left[\begin{array}{cc}
u_{i}^{Y} & 0 \\
0 & u_{i}^{Z} \\
u_{i}^{Z} & 0
\end{array}\right] 2\left[\begin{array}{ccc}
\sigma_{Y, i}^{2} u_{i}^{Y} & \sigma_{Y Z, i} u_{i}^{Z} & \sigma_{Y, i}^{2} u_{i}^{Z} \\
\sigma_{Y Z, i} u_{i}^{Y} & \sigma_{Z, i}^{2} u_{i}^{Z} & \sigma_{Y Z, i} u_{i}^{Z}
\end{array}\right] \\
& =4\left[\begin{array}{ccc}
\sigma_{Y, i}^{2} u_{i}^{Y} & \sigma_{Y Z, i} u_{i}^{Y} u_{i}^{Z} & \sigma_{Y, i}^{2} u_{i}^{Y} u_{i}^{Z} \\
\sigma_{Y Z, i} u_{i}^{Y} & \sigma_{Z, i}^{2} u_{i}^{2 Z} & \sigma_{Y Z, i} u_{i}^{2 Z} \\
\sigma_{Y Z, i} u_{i}^{Y} u_{i}^{Z} & \sigma_{Y Z, i} u_{i}^{2 Z} & \sigma_{Y, i}^{2} u_{i}^{2 Z}
\end{array}\right] \\
& +2\left[\begin{array}{cccc}
\sigma_{Y, i}^{4} & \sigma_{Y, i}^{2} \sigma_{Y Z, i} & \sigma_{Y, i}^{2} \sigma_{Y Z, i} & \sigma_{Y Z, i}^{2} \\
\sigma_{Y Z, i}^{2} & \sigma_{Z, i}^{2} \sigma_{Y Z, i} & \sigma_{Z, i}^{2} \sigma_{Y Z, i} & \sigma_{Z, i}^{4} \\
\sigma_{Y, i}^{2} \sigma_{Y Z, i} & \sigma_{Y, i}^{2} \sigma_{Z, i}^{2} & \sigma_{Y Z, i}^{2} & \sigma_{Z, i}^{2} \sigma_{Y Z, i}
\end{array}\right]\left[\begin{array}{ccc}
1 & 0 & 0 \\
0 & 0 & 1 \\
0 & 0 & 0 \\
0 & 1 & 0
\end{array}\right] \\
& =4\left[\begin{array}{ccc}
\sigma_{Y, i}^{2} u_{i}^{Y} & \sigma_{Y Z, i} u_{i}^{Y} u_{i}^{Z} & \sigma_{Y, i}^{2} u_{i}^{Y} u_{i}^{Z} \\
\sigma_{Y Z, i} u_{i}^{Y} & \sigma_{Z, i}^{2} u_{i}^{2 Z} & \sigma_{Y Z, i} u_{i}^{2 Z} \\
\sigma_{Y Z, i} u_{i}^{Y} u_{i}^{Z} & \sigma_{Y Z, i} u_{i}^{2 Z} & \sigma_{Y, i}^{2} u_{i}^{2 Z}
\end{array}\right]+2\left[\begin{array}{ccc}
\sigma_{Y, i}^{4} & \sigma_{Y Z, i}^{2} & \sigma_{Y, i}^{2} \sigma_{Y Z, i} \\
\sigma_{Y Z, i}^{2} & \sigma_{Z, i}^{4} & \sigma_{Z, i}^{2} \sigma_{Y Z, i} \\
\sigma_{Y, i}^{2} \sigma_{Y Z, i} & \sigma_{Z, i}^{2} \sigma_{Y Z, i} & \sigma_{Y, i}^{2} \sigma_{Z, i}^{2}
\end{array}\right] \\
& =\left[\begin{array}{ccc}
4 \sigma_{Y, i}^{2} u_{i}^{Y}+2 \sigma_{Y, i}^{4} & 4 \sigma_{Y Z, i} u_{i}^{Y} u_{i}^{Z}+2 \sigma_{Y Z, i}^{2} & 4 \sigma_{Y, i}^{2} u_{i}^{Y} u_{i}^{Z}+2 \sigma_{Y, i}^{2} \sigma_{Y Z, i} \\
4 \sigma_{Y Z, i} u_{i}^{Y}+2 \sigma_{Y Z, i}^{2} & 4 \sigma_{Z, i}^{2} u_{i}^{2 Z}+2 \sigma_{Z, i}^{4} & 4 \sigma_{Y Z, i} u_{i}^{2 Z}+2 \sigma_{Z, i}^{2} \sigma_{Y Z, i} \\
4 \sigma_{Y Z, i} u_{i}^{Y} u_{i}^{Z}+2 \sigma_{Y, i}^{2} \sigma_{Y Z, i} & 4 \sigma_{Y Z, i} u_{i}^{2 Z}+2 \sigma_{Z, i}^{2} \sigma_{Y Z, i} & 4 \sigma_{Y, i}^{2} u_{i}^{2 Z}+2 \sigma_{Y, i}^{2} \sigma_{Z, i}^{2}
\end{array}\right] .
\end{aligned}
$$

The formula for the third and fourth moments in the above derivation can be used as "working" covariance matrix in the estimating equation in (9.10). Equation (9.10) is an unbiased estimating equation regardless of normality assumption and the M-estimation theory applies. Under fairly mild regularity conditions, the resulting estimates are consis-
tent, asymptotically normal.
Similar to the proof in Vonesh et al. (2001), alternatively $\boldsymbol{\theta}$ and $\alpha$ can be jointly estimated by maximizing the following objective function:

$$
\begin{equation*}
\boldsymbol{Q}_{E L S}(\boldsymbol{\theta}, \boldsymbol{\alpha})=\sum_{i} \sum_{j}\left\{\left(\boldsymbol{\omega}_{i j}-\mathbf{u}_{i}(\boldsymbol{\theta})\right)^{T} \boldsymbol{\Sigma}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})^{-1}\left(\boldsymbol{\omega}_{i j}-\mathbf{u}_{i}(\boldsymbol{\theta})\right)^{T}+\log \left|\boldsymbol{\Sigma}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})\right|\right\} . \tag{9.13}
\end{equation*}
$$

The function in (9.13) is the extended least squares objective function associated with the joint estimation of both $\theta$ and $\alpha$. The corresponding joint estimating equations for maximizing $Q_{E L S}(\theta, \alpha)$ are given as follows:

$$
\begin{align*}
\boldsymbol{U}_{E L S}(\boldsymbol{\theta}, \boldsymbol{\alpha}) & =\sum_{i=1}^{k} \sum_{j=1}^{n_{i}}\left[\begin{array}{cc}
\boldsymbol{D}_{i}(\boldsymbol{\theta}) & \mathbf{0} \\
\boldsymbol{E}_{i}(\boldsymbol{\theta}) & \boldsymbol{E}_{i}(\boldsymbol{\alpha})
\end{array}\right]^{T}\left[\begin{array}{cc}
\boldsymbol{\Sigma}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha}) & \mathbf{0} \\
\mathbf{0} & \boldsymbol{V}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})
\end{array}\right]^{-1}\left[\begin{array}{c}
\boldsymbol{\omega}_{i j}-\boldsymbol{u}_{i}(\boldsymbol{\theta}) \\
\boldsymbol{S}_{i j}(\boldsymbol{\theta})-\boldsymbol{\sigma}_{i}(\boldsymbol{\theta}, \boldsymbol{\alpha})
\end{array}\right] \\
& =\mathbf{0} \tag{9.14}
\end{align*}
$$

where

$$
\begin{aligned}
D_{i}(\theta) & =\partial u_{i}(\theta) / \partial \theta^{T} \\
E_{i}(\theta) & =\partial \sigma_{i}(\theta, \alpha) / \partial \theta^{T} \\
E_{i}(\alpha) & =\partial \sigma_{i}(\theta, \alpha) / \partial \alpha^{T} \\
\sigma_{i}(\theta, \alpha) & =E\left[S_{i j}(\theta)\right] \\
S_{i j}(\theta) & =\operatorname{Vec}\left(\left(\omega_{i j}-u_{i}(\theta)\right)\left(\omega_{i j}-u_{i}(\theta)\right)^{T}\right), \\
V_{i}(\theta, \alpha) & =2 \Sigma_{i}(\theta, \alpha) \otimes \Sigma_{i}(\theta, \alpha) .
\end{aligned}
$$

It is shown that under normality, the estimation in (9.10) and the ELS method are equivalent to the GEE2 method. The empirical sandwich estimator can be used to achieve asymptotically valid inference provided that mean and variance structure have been correctly specified.

If we set $E_{i}(\boldsymbol{\theta})=0$ in (9.14), one gets the set of quasi-extended least squares (QELS) estimating equations described in Vonesh et al. (2001). The advantage of using QELS estimating equations is that under certain conditions, one can still achieve a consistent estimate of $\boldsymbol{\theta}$ even if the variance-covariance structure is mis-specified.

As mentioned previously since IRGLS procedure for estimating $\tau$, which consists simply of repeated application of Taylor series linearization and estimated generalized least squares, is equivalent to maximum likelihood estimation and it is equivalent to the ELS estimating approach assuming normality. Hence only ELS and QELS will be used in the following fitting procedure because the focus of this chapter is only for estimation of the parameters in order to find the mean model for efficacy and safety which allow us to select the best dose for Phase III. The comparison with IRGLS and using the maximum likelihood will be a future direction for different estimating approaches regarding the joint modeling.

### 9.3.2 Fitting the joint nonlinear bivariate continuous and discrete model with ELS or QELS approach

The advantage of using the ELS approach is that the 3rd and 4th order moments are automatically generated as though from a multivariate normal distribution even though data may not have multivariate normal. In this way, we can essentially treat data (i.e., continuous and discrete response variables) as being multivariate normal even though it is not and still achieve consistent parameter estimates of the mean and variancecovariance parameters when mean and variance-covariance are not mis-specified. Alternatively we can use the quasi-ELS approach (QELS) which uses a slightly different working
covariance structure. In addition if the underlying third- and fourth-order moments do not deviate too greatly from those of a Gaussian distribution, then the ELS estimators will be more efficient than GEE-based estimators.

Here we assume a bivariate continuous and binary variable. The ELS and QELS macros were kindly provided by Dr. Vonesh, these macros need to be adapted to the current nonlinear bivariate continuous and binary outcome setting. The key for the estimating joint nonlinear model parameters is how to specify the mean model for continuous and binary variable and the variance-covariance structure for the bivariate variable. The mean model for continuous variable can be nonlinear model such as Emax model, and mean model for binary variable can be logistic or other link functions. In addition, since the correlation between bivariate outcome depends on the individual dose level, the correlation need to be estimated in each dose level and specified in the variance-covariance structure. The initial values for all mean model parameters, variance and correlation at each dose level need to be correctly specified. These initial values for all mean model parameters, variance are initially obtained by the separate model fittings for the continuous and binary variable while the initial value for the correlation is obtained by direct pearson correlation coefficients between the bivariate outcomes at each dose level.

### 9.4 Evaluation of the nonlinear joint model fitting

### 9.4.1 Parameter estimate bias and relative efficiency

The evaluations of parameter estimate bias and relative efficiency are similar to what are described in Section 4.2.2 of Part I.

### 9.4.2 Evaluate the model fitting

There are many different criteria to use to evaluate the model fitting. Many of criteria are slightly variation of another and difference maybe subtle. For the nonlinear joint bivariate mixed type model, since our estimation method based on QELS is equivalent to the maximum likelihood under the normality, thus using Alaike's information criterion may be a fairly good choice. This criteria uses likelihood as the basis but other criteria may not take into account of the nonlinear response which may not be an appropriate choice.

## AIC criteria:

Alaike's information criterion (AIC) is a measure of the goodness of fit of an estimated statistical model. It is a test between models and is a tool for model selection. The model having the lowest AIC is the best model. In general case, the AIC is:

$$
A I C=2 k-2 \ln (L)
$$

where $k$ is the number of parameters in the statistical model, and $L$ is the maximized value of the likelihood for the estimated model.

As illustrated in Section 9.2, assume there are 7 dose levels, $Y_{i j}$ follows Emax model and $Z_{i j}$ follows probit model. $\widetilde{\rho}_{i}$ represents the correlation between the continuous efficacy variable and discrete safety variable for dose level $i$. The parameters to be estimated are: $e_{0}, e_{\max }, e d_{50}, \alpha, \beta, \sigma_{Y, i}^{2}, \widetilde{\rho}_{i}$.

$$
\begin{aligned}
A I C & =2 k-2 \ln (L) \\
& =38-2 \ln (L)
\end{aligned}
$$

As we noted, the AIC is based on maximum likelihood estimation while ELS and QELS approach do not directly use maximum likelihood, though ELS or QELS approach is equivalent to the maximum likelihood approach under the normality. The modification to AIC where the likelihood is replaced by the quasi-likelihood (QIC) (Pan, 2001) may be investigated in the future as the criteria for the model selection.

### 9.4.3 Model mis-specification

To test robustness of ELS or QELS method, the data simulated from certain model will be fitted for different model by the ELS or QELS methods. Here the focus is on the discrete variable model mis-specification. Chi-square goodness-of-fit test can be applied to the discrete distribution. It allows us to evaluate how "close" the observed values are to those would be expected given in the model. Chi-square test uses a measure of goodness of fit which is defined as the sum of squares of differences of observed and expected outcome frequencies divided by the expected value, i.e.,

$$
\chi^{2}=\sum_{i=1}^{r} \sum_{j=0}^{1} \frac{\left(\mathrm{O}_{i, j}-\mathrm{E}_{i, j}\right)^{2}}{\mathrm{E}_{i, j}}
$$

where $i$ is the dose level and $j$ is binary outcome. Degree of freedom=r-1. The parameter estimates from the QELS method are obtained and the mean model for the binary outcome with the particular distribution are used in the model fitting.

$$
P\left(Z_{i j}=1 \mid d_{i}\right)=g\left(d_{i}, \boldsymbol{\theta}_{Z}\right),
$$

where $g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)$ can be inverse of logistic, probit or complementary log-log links.
There are different expected mean value at each dose level. $E_{i, j}$ is computed as
follows:

$$
E_{i, 1}=g\left(d_{i}, \boldsymbol{\theta}_{Z}\right) * n_{i}, \text { and } E_{i, 0}=n_{i}-E_{i, 1}
$$

where $n_{i}$ is number of patients at each dose level.

### 9.5 Simulation of correlated bivariate continuous and discrete responses

The simulation will be based on Method 1 described in Section 9.2.1, i.e., we will first simulate a bivariate correlated continuous random variable and then use a threshold to obtain correlated continuous and discrete variables.

### 9.5.1 Simulation Procedures

1. Compute mean model for the latent continuous variable.

- Assume threshold value $c$ which is used to as cut off value to obtain the binary safety variable from the latent continuous safety variable.
- Assume that the latent continuous safety variable is standardized with variance

1. 

- Assume correlation $\rho$ between latent continuous safety variable and continuous efficacy variable.
- Assume the mean model and variance for continuous efficacy variable.
- Compute that the latent continuous safety mean model based on the assumed discrete binary variable mean model using formula (9.4).
- Compute the correlation between the continuous efficacy variable and discrete safety variable at each dose level using formula (9.5).

Note: For simplicity and consistency with Part I, here we assume $\rho_{i,}, \sigma_{Y, i}^{2}$ and $\sigma_{S, i}^{2}$ are equal for all dose levels.
2. Simulate bivariate continuous normal efficacy and latent safety responses.
3. Use cut off value $c$ on the latent continuous variable to obtain discrete safety random variable.

## Example:

For the simulation purpose, we assume that we are interested in a continuous efficacy outcome $Y_{i j}$ measured by decreased diastolic blood pressure (DBP) from baseline and a discrete safety outcome $Z_{i j}$ measured by renal dysfunction (yes or no). The continuous efficacy variable decreased diastolic blood pressure (DBP) from baseline follows the Emax model with mean model $2.5+14.5 *$ dose $/(.2+$ dose $)$ and the standard deviation for decreased diastolic blood pressure is 7 mmHg ; The binary renal dysfunction outcome is an indicator of the decreased GFR from baseline $\left(S_{i j}\right)$. The cut off is $3.5 \mathrm{ml} / \mathrm{min} / 1.73$ $\mathrm{m}^{2}$ which means that if the decreased GFR from baseline is $>3.5 \mathrm{ml} / \mathrm{min} / 1.73 \mathrm{~m}^{2}$ then binary renal dysfunction is coded as 1 (yes) otherwise 0 (no). Next we consider logistic, probit or complementary log-log model for the binary safety outcome. For simplicity the linear regression is used in the simulation for the links we selected, to be specific, we assume that the mean model of the binary outcome is $P\left(Z_{i j}=1 \mid d_{i}\right)=\frac{\exp \left(-2.123+3.728 d_{i}\right)}{1+\exp \left(-2.123+3.728 d_{i}\right)}$ for $\operatorname{logit}, P\left(Z_{i j}=1 \mid d_{i}\right)=\Phi\left(-1.166+1.853 d_{i}\right)$ for probit, and $P\left(Z_{i j}=1 \mid d_{i}\right)=1-$
$\exp \left(-\exp \left(-2.123+2.828 d_{i}\right)\right)$ for complementary log-log respectively. The mean model of the latent continuous GFR outcome can be derived based on formula (9.4) for different assumed link functions for the safety binary outcome. Let $\operatorname{var}\left(S_{i j}\right)=1, n=50 /$ dose group and dose values $0,0.05,0.2,0.4,0.6,0.8$ and 1 . The data are simulated 500 times for each $\rho=0,0.4,0.8$, respectively from the above continuous efficacy and safety models. The data are generated using the rmvnorm function in R . Then the latent continuous response $S_{i j}$ is dichotomized to $Z_{i j}$.

### 9.5.2 Evaluation of Simulation Results

1. Fit the discrete safety variable using the correct link function used in the simulation.
2. Fit the correct nonlinear regression model for the continuous variable using gnls (Pinheiro and Bates, 2000) function in $R$ to get the parameter estimates for the continuous model.
3. Estimate the correlations between bivariate continuous and discrete outcomes for each dose level using Pearson correlation coefficients for the simulated data.
4. Compare all the estimates of the mean model parameters and correlations with true parameter values.

The following Tables 9.1-9.3 are the simulated outputs with different regression models for the discrete safety outcomes. As shown in the table, the proposed method provides very consistent parameter estimates including correlations in each dose level for different models for the discrete variable. Furthermore, the simulation results show the correctness of theoritical result we derived in formula (9.6).

Table 9.1: Simulated bivariate continuous and discrete outcomes (assuming discrete outcome follows probit regression)

|  |  |  |  | $\widetilde{\rho}_{i}$ |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\rho$ | Parameter | $\alpha$ | $\beta$ | $\widetilde{\rho}_{0}$ | $\widetilde{\rho}_{0.05}$ | $\widetilde{\rho}_{0.2}$ | $\widetilde{\rho}_{0.4}$ | $\widetilde{\rho}_{0.6}$ | $\widetilde{\rho}_{0.8}$ | $\widetilde{\rho}_{1}$ |  |
| 0 | True value | -1.166 | 1.853 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |  |
|  | Simu. value | -1.176 | 1.873 | -.008 | .005 | -.000 | .001 | .003 | -.009 | -.006 |  |
| 0.4 | True value | -1.166 | 1.853 | .247 | .257 | .284 | .310 | .320 | .313 | .292 |  |
|  | Simu. value | -1.170 | 1.862 | .244 | .257 | .281 | .313 | .320 | .310 | .286 |  |
| 0.8 | True value | -1.166 | 1.853 | .495 | .515 | .568 | .618 | .638 | .627 | .585 |  |
|  | Simu. value | -1.174 | 1.864 | .489 | .510 | .569 | .627 | .637 | .629 | .583 |  |

Notes: $\widetilde{\rho}_{i}$ is the correlation between bivariate continuous and discrete outcomes at dose level $i$.
$\rho$ is the correlation used for simulating the bivariate continuous outcomes.

Table 9.2: Simulated bivariate continuous and discrete outcomes (assuming discrete outcome follows logistic regression)

|  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\rho$ | Parameter | $\alpha$ | $\beta$ | $\widetilde{\rho}_{0}$ | $\widetilde{\rho}_{0.05}$ | $\widetilde{\rho}_{0.2}$ | $\widetilde{\rho}_{0.4}$ | $\widetilde{\rho}_{0.6}$ | $\widetilde{\rho}_{0.8}$ | $\widetilde{\rho}_{1}$ |
| 0 | True value | -2.123 | 3.728 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
|  | Simu. value | -2.129 | 3.743 | .004 | -.000 | -.002 | -.002 | .002 | .007 | -.007 |
| 0.4 | True value | -2.123 | 3.728 | .239 | .250 | .280 | .310 | .319 | .303 | .269 |
|  | Simu. value | -2.122 | 3.720 | 241 | .250 | .272 | .313 | .317 | .299 | .269 |
| 0.8 | True value | -2.123 | 3.728 | .477 | .499 | .561 | .621 | .638 | .606 | .537 |
|  | Simu. value | -2.131 | 3.732 | .474 | .497 | .556 | .621 | .639 | .608 | .531 |

Notes: $\widetilde{\rho}_{i}$ is the correlation between bivariate continuous and discrete outcomes at dose level $i$. $\rho$ is the correlation used for simulating the bivariate continuous outcomes.

Table 9.3: Simulated bivariate continuous and discrete outcomes (assuming discrete outcome follows complementary log-log regression)

|  |  |  | $\widetilde{\rho}_{i}$ |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\rho$ | Parameter | $\alpha$ | $\beta$ | $\widetilde{\rho}_{0}$ | $\widetilde{\rho}_{0.05}$ | $\widetilde{\rho}_{0.2}$ | $\widetilde{\rho}_{0.4}$ | $\widetilde{\rho}_{0.6}$ | $\widetilde{\rho}_{0.8}$ | $\widetilde{\rho}_{1}$ |
| 0 | True value | -2.123 | 2.828 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
|  | Simu. value | -2.134 | 2.843 | -.011 | -.002 | -.012 | .007 | .004 | -.005 | -.001 |
| 0.4 | True value | -2.123 | 2.828 | .242 | .252 | .277 | .305 | .319 | .306 | .253 |
|  | Simu. value | -2.157 | 2.877 | .232 | .239 | .266 | .312 | .313 | .297 | .253 |
| 0.8 | True value | -2.123 | 2.828 | .484 | .502 | .553 | .610 | .638 | .612 | .506 |
|  | Simu. value | -2.153 | 2.870 | .473 | .486 | .544 | .613 | .638 | .615 | .506 |

Notes: $\widetilde{\rho}_{i}$ is the correlation between bivariate continuous and discrete outcomes at dose level $i$. $\rho$ is the correlation used for simulating the bivariate continuous outcomes.

### 9.6 Fitting the joint nonlinear continuous and discrete responses

 modelAll the parameter estimates for the joint nonlinear continuous and discrete responses model are obtained using SAS IML. The ELS and QELS methods discussed in Section 9.3.1 are used for the estimation. Assume the 1st and 2nd order moments are from the multivariate nonlinear model, the first two moments are used to set up the mean model and covariance structure. Furthermore we assume that the 3rd and 4th order moments of a general multivariate response vector coincide with the multivariate normal response. The ELS or QELS algorithms are applied under this set up. The efficiency of ELS depends on how far the third and fourth order moments deviate from those calculated using normal distributions.

### 9.6.1 Comparison of parameter estimates and their standard errors by the ELS and QELS approaches

The results of parameter estimates of ELS and QELS approaches are displayed in Tables 9.4 to 9.6. As summarized in Table 9.7, QELS yields better estimates on average. Especially with stronger correlations QELS approach has smaller standard deviations for both continuous and discrete model parameter estimates overall. Model-based and robust standard errors are similar for discrete variable parameter estimates. This is not surprising since the information about the discrete variable is not as informative as the latent continuous variable even though the correlation between the bivariate continuous and discrete variable increases. Overall, it seems that QELS method is more robust under various strength of bivariate correlation or the different models selected for the discrete safety variable.

| Method | Bivariate Cont. Corr. ( $\rho$ ) | Summary | Parameters |  |  |  |  |  |  |  |  |  |  |  | Efficacy Variance $\sigma_{\gamma}^{2}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Efficacy |  |  | Safety |  | Cont. and discrete bivariate corr by dose |  |  |  |  |  |  |  |
|  |  |  | e0 | ed50 | emax | $\alpha$ | $\beta$ | $\widetilde{\rho}_{0}$ | $\widetilde{\rho}_{0.05}$ | $\widetilde{\rho}_{0.2}$ | $\widetilde{\rho}_{0.4}$ | $\widetilde{\rho}_{0.6}$ | $\widetilde{\rho}_{0.8}$ | $\widetilde{\rho}_{1}$ |  |
| ELS | 0.0 | Mean | 2.39 | 0.23 | 15.16 | -1.19 | 1.89 | -0.01 | -0.01 | -0.00 | 0.00 | -0.01 | 0.01 | -0.00 | 48.84 |
|  |  | $\mathrm{SE}_{S}$ | 0.95 | 0.12 | 1.76 | 0.14 | 0.25 | 0.14 | 0.15 | 0.16 | 0.16 | 0.15 | 0.15 | 0.15 | 3.54 |
|  |  | $\mathrm{SE}_{M}$ | 0.90 | 0.10 | 1.76 | 0.08 | 0.15 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 3.69 |
|  |  | RSE | 0.90 | 0.10 | 1.77 | 0.14 | 0.23 | 0.12 | 0.12 | 0.13 | 0.14 | 0.14 | 0.13 | 0.14 | 3.67 |
|  | 0.4 | Mean | 2.55 | 0.23 | 14.93 | -1.18 | 1.88 | 0.24 | 0.26 | 0.27 | 0.31 | 0.31 | 0.29 | 0.29 | 48.88 |
|  |  | $\mathrm{SE}_{S}$ | 0.91 | 0.11 | 1.80 | 0.13 | 0.23 | 0.13 | 0.14 | 0.13 | 0.12 | 0.13 | 0.13 | 0.13 | 3.38 |
|  |  | $\mathrm{SE}_{M}$ | 0.88 | 0.10 | 1.72 | 0.08 | 0.15 | 0.13 | 0.13 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 3.67 |
|  |  | RSE | 0.88 | 0.10 | 1.76 | 0.14 | 0.23 | 0.11 | 0.11 | 0.12 | 0.12 | 0.12 | 0.12 | 0.11 | 3.66 |
|  | 0.8 | Mean | 2.51 | 0.22 | 14.82 | -1.17 | 1.88 | 0.49 | 0.52 | 0.58 | 0.62 | 0.64 | 0.62 | 0.58 | 48.78 |
|  |  | $\mathrm{SE}_{S}$ | 0.87 | 0.10 | 1.70 | 0.13 | 0.22 | 0.08 | 0.10 | 0.08 | 0.07 | 0.07 | 0.07 | 0.08 | 3.45 |
|  |  | $\mathrm{SE}_{M}$ | 0.80 | 0.08 | 1.53 | 0.08 | 0.15 | 0.10 | 0.09 | 0.08 | 0.07 | 0.07 | 0.07 | 0.08 | 3.49 |
|  |  | RSE | 0.82 | 0.09 | 1.67 | 0.13 | 0.22 | 0.08 | 0.08 | 0.07 | 0.07 | 0.06 | 0.07 | 0.07 | 3.66 |
| QELS | 0.0 | Mean | 2.39 | 0.23 | 15.18 | -1.19 | 1.88 | -0.01 | -0.01 | 0.00 | 0.00 | -0.01 | 0.01 | -0.00 | 48.83 |
|  |  | $\mathrm{SE}_{S}$ | 0.95 | 0.12 | 1.78 | 0.14 | 0.25 | 0.15 | 0.15 | 0.16 | 0.16 | 0.15 | 0.15 | 0.15 | 3.55 |
|  |  | $\mathrm{SE}_{M}$ | 0.91 | 0.10 | 1.77 | 0.13 | 0.22 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 3.69 |
|  |  | RSE | 0.90 | 0.10 | 1.77 | 0.13 | 0.22 | 0.12 | 0.12 | 0.13 | 0.14 | 0.14 | 0.13 | 0.13 | 3.67 |
|  | 0.4 | Mean | 2.55 | 0.23 | 14.92 | -1.18 | 1.88 | 0.24 | 0.26 | 0.27 | 0.31 | 0.32 | 0.29 | 0.30 | 48.88 |
|  |  | $\mathrm{SE}_{S}$ | 0.91 | 0.11 | 1.80 | 0.13 | 0.22 | 0.13 | 0.14 | 0.13 | 0.12 | 0.13 | 0.13 | 0.13 | 3.39 |
|  |  | $\mathrm{SE}_{M}$ | 0.89 | 0.10 | 1.76 | 0.13 | 0.22 | 0.13 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 3.67 |
|  |  | RSE | 0.88 | 0.10 | 1.76 | 0.13 | 0.22 | 0.11 | 0.11 | 0.12 | 0.12 | 0.12 | 0.12 | 0.11 | 3.65 |
|  | 0.8 | Mean | 2.51 | 0.22 | 14.83 | -1.17 | 1.87 | 0.49 | 0.51 | 0.58 | 0.62 | 0.64 | 0.62 | 0.58 | 48.82 |
|  |  | $\mathrm{SE}_{S}$ | 0.88 | 0.10 | 1.67 | 0.12 | 0.21 | 0.08 | 0.10 | 0.08 | 0.07 | 0.07 | 0.07 | 0.08 | 3.45 |
|  |  | $\mathrm{SE}_{M}$ | 0.82 | 0.09 | 1.65 | 0.13 | 0.22 | 0.10 | 0.09 | 0.08 | 0.07 | 0.07 | 0.07 | 0.08 | 3.43 |
|  |  | RSE | 0.82 | 0.09 | 1.66 | 0.13 | 0.22 | 0.08 | 0.08 | 0.07 | 0.07 | 0.06 | 0.07 | 0.07 | 3.54 |
| Notes: SE $_{S}$ - Standard deviation; $\mathrm{SE}_{M}$ - Average model-based standard errors; RSE - Average robust standard errors. $\rho$ is the correlation used for simulating the bivariate continuous outcomes. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Table 9.5: Comparison of parameter estimates and their standard errors between the ELS and QELS approaches-Emax for efficacy outcome and Logit for safety outcome

| Method | Bivariate Cont. Corr. ( $\rho$ ) | Summary | Parameters |  |  |  |  |  |  |  |  |  |  |  | Efficacy Variance $\sigma_{\gamma}^{2}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Efficacy |  |  | Safety |  | Cont. and discrete bivariate corr by dose |  |  |  |  |  |  |  |
|  |  |  | e0 | ed50 | emax | $\alpha$ | $\beta$ | $\widetilde{\rho}_{0}$ | $\widetilde{\rho}_{0.05}$ | $\widetilde{\rho}_{0.2}$ | $\widetilde{\rho}_{0.4}$ | $\widetilde{\rho}_{0.6}$ | $\widetilde{\rho}_{0.8}$ | $\widetilde{\rho}_{1}$ |  |
| ELS | 0.0 | Mean | 2.55 | 0.24 | 15.06 | -2.17 | 3.81 | -0.01 | -0.01 | -0.02 | 0.00 | 0.01 | -0.01 | 0.00 | 48.88 |
|  |  | $\mathrm{SE}_{S}$ | 0.87 | 0.12 | 1.69 | 0.26 | 0.45 | 0.15 | 0.15 | 0.14 | 0.15 | 0.15 | 0.15 | 0.15 | 3.79 |
|  |  | $\mathrm{SE}_{M}$ | 0.90 | 0.11 | 1.80 | 0.14 | 0.25 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 3.69 |
|  |  | RSE | 0.89 | 0.11 | 1.81 | 0.26 | 0.43 | 0.12 | 0.12 | 0.13 | 0.14 | 0.14 | 0.14 | 0.13 | 3.68 |
|  | 0.4 | Mean | 2.41 | 0.23 | 15.13 | -2.15 | 3.78 | 0.24 | 0.24 | 0.27 | 0.32 | 0.31 | 0.30 | 0.26 | 48.66 |
|  |  | $\mathrm{SE}_{S}$ | 0.91 | 0.12 | 2.03 | 0.27 | 0.47 | 0.13 | 0.13 | 0.13 | 0.13 | 0.12 | 0.13 | 0.13 | 3.88 |
|  |  | $\mathrm{SE}_{M}$ | 0.88 | 0.10 | 1.71 | 0.14 | 0.25 | 0.13 | 0.13 | 0.12 | 0.12 | 0.12 | 0.12 | 0.13 | 3.66 |
|  |  | RSE | 0.87 | 0.10 | 1.75 | 0.26 | 0.43 | 0.11 | 0.11 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 3.65 |
|  | 0.8 | Mean | 2.51 | 0.22 | 14.83 | -2.15 | 3.80 | 0.47 | 0.50 | 0.56 | 0.62 | 0.64 | 0.60 | 0.53 | 48.78 |
|  |  | $\mathrm{SE}_{S}$ | 0.87 | 0.09 | 1.69 | 0.27 | 0.45 | 0.09 | 0.11 | 0.08 | 0.07 | 0.07 | 0.07 | 0.09 | 3.43 |
|  |  | $\mathrm{SE}_{M}$ | 0.81 | 0.08 | 1.52 | 0.14 | 0.25 | 0.10 | 0.10 | 0.08 | 0.07 | 0.07 | 0.08 | 0.09 | 3.51 |
|  |  | RSE | 0.82 | 0.09 | 1.68 | 0.25 | 0.43 | 0.09 | 0.08 | 0.08 | 0.07 | 0.06 | 0.07 | 0.08 | 3.66 |
| QELS | 0.0 | Mean | 2.55 | 0.24 | 15.05 | -2.16 | 3.79 | -0.01 | -0.01 | -0.02 | 0.00 | 0.01 | -0.01 | 0.00 | 48.86 |
|  |  | $\mathrm{SE}_{S}$ | 0.87 | 0.12 | 1.69 | 0.24 | 0.43 | 0.16 | 0.16 | 0.14 | 0.15 | 0.15 | 0.15 | 0.15 | 3.78 |
|  |  | $\mathrm{SE}_{M}$ | 0.90 | 0.11 | 1.81 | 0.24 | 0.41 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 3.69 |
|  |  | RSE | 0.89 | 0.11 | 1.81 | 0.24 | 0.41 | 0.12 | 0.12 | 0.13 | 0.14 | 0.14 | 0.13 | 0.13 | 3.68 |
|  | 0.4 | Mean | 2.41 | 0.23 | 15.11 | -2.15 | 3.76 | 0.24 | 0.24 | 0.27 | 0.32 | 0.31 | 0.30 | 0.26 | 48.67 |
|  |  | $\mathrm{SE}_{S}$ | 0.91 | 0.11 | 1.96 | 0.25 | 0.45 | 0.13 | 0.13 | 0.13 | 0.13 | 0.12 | 0.13 | 0.13 | 3.89 |
|  |  | $\mathrm{SE}_{M}$ | 0.88 | 0.10 | 1.73 | 0.24 | 0.41 | 0.13 | 0.13 | 0.12 | 0.12 | 0.12 | 0.12 | 0.13 | 3.65 |
|  |  | RSE | 0.87 | 0.10 | 1.74 | 0.24 | 0.41 | 0.11 | 0.11 | 0.12 | 0.12 | 0.12 | 0.12 | 0.11 | 3.64 |
|  | 0.8 | Mean | 2.51 | 0.22 | 14.83 | -2.15 | 3.79 | 0.47 | 0.50 | 0.56 | 0.62 | 0.64 | 0.60 | 0.53 | 48.82 |
|  |  | $\mathrm{SE}_{S}$ | 0.87 | 0.09 | 1.66 | 0.24 | 0.41 | 0.09 | 0.11 | 0.08 | 0.07 | 0.07 | 0.07 | 0.08 | 3.45 |
|  |  | $\mathrm{SE}_{M}$ | 0.82 | 0.09 | 1.66 | 0.24 | 0.41 | 0.10 | 0.09 | 0.08 | 0.07 | 0.07 | 0.08 | 0.09 | 3.45 |
|  |  | RSE | 0.82 | 0.09 | 1.67 | 0.24 | 0.41 | 0.08 | 0.08 | 0.07 | 0.07 | 0.06 | 0.07 | 0.08 | 3.53 |
| Notes: $\mathrm{SE}_{S}$ - Standard deviation; $\mathrm{SE}_{M}$ - Average model-based standard errors; RSE - Average robust standard errors. $\rho$ is the correlation used for simulating the bivariate continuous outcomes. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Table 9.6: Comparison of parameter estimates and their standard errors between the ELS and QELS approaches-Emax for efficacy and Cloglog for safety

| Method | Bivariate Cont. Corr. ( $\rho$ ) | Summary | Parameters |  |  |  |  |  |  |  |  |  |  |  | Efficacy <br> Variance <br> $\sigma_{Y}^{2}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Efficacy |  |  | Safety |  | Cont. and discrete bivariate corr by dose |  |  |  |  |  |  |  |
|  |  |  | e0 | ed50 | emax | $\alpha$ | $\beta$ | $\widetilde{\rho}_{0}$ | $\widetilde{\rho}_{0.05}$ | $\widetilde{\rho}_{0.2}$ | $\widetilde{\rho}_{0.4}$ | $\widetilde{\rho}_{0.6}$ | $\widetilde{\rho}_{0.8}$ | $\widetilde{\rho}_{1}$ |  |
| ELS | 0.0 | Mean | 2.38 | 0.23 | 15.15 | -2.17 | 2.89 | -0.01 | -0.01 | -0.01 | 0.00 | -0.01 | 0.01 | -0.00 | 48.79 |
|  |  | $\mathrm{SE}_{S}$ | 0.95 | 0.12 | 1.77 | 0.23 | 0.34 | 0.14 | 0.15 | 0.16 | 0.15 | 0.15 | 0.16 | 0.13 | 3.48 |
|  |  | $\mathrm{SE}_{M}$ | 0.90 | 0.10 | 1.76 | 0.12 | 0.16 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 3.69 |
|  |  | RSE | 0.90 | 0.10 | 1.76 | 0.22 | 0.30 | 0.12 | 0.12 | 0.13 | 0.14 | 0.14 | 0.13 | 0.13 | 3.66 |
|  | 0.4 | Mean | 2.56 | 0.23 | 14.95 | -2.16 | 2.88 | 0.23 | 0.25 | 0.27 | 0.31 | 0.31 | 0.28 | 0.26 | 48.87 |
|  |  | $\mathrm{SE}_{S}$ | 0.90 | 0.11 | 1.83 | 0.23 | 0.33 | 0.13 | 0.14 | 0.13 | 0.12 | 0.13 | 0.13 | 0.13 | 3.38 |
|  |  | $\mathrm{SE}_{M}$ | 0.88 | 0.10 | 1.72 | 0.12 | 0.16 | 0.13 | 0.13 | 0.12 | 0.12 | 0.12 | 0.12 | 0.13 | 3.67 |
|  |  | RSE | 0.88 | 0.10 | 1.77 | 0.22 | 0.30 | 0.11 | 0.11 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 3.66 |
|  | 0.8 | Mean | 2.51 | 0.23 | 14.91 | -2.14 | 2.88 | 0.48 | 0.50 | 0.56 | 0.61 | 0.64 | 0.61 | 0.50 | 48.74 |
|  |  | $\mathrm{SE}_{S}$ | 0.86 | 0.10 | 1.79 | 0.23 | 0.32 | 0.08 | 0.10 | 0.08 | 0.07 | 0.07 | 0.07 | 0.10 | 3.41 |
|  |  | $\mathrm{SE}_{M}$ | 0.80 | 0.09 | 1.52 | 0.12 | 0.16 | 0.10 | 0.09 | 0.09 | 0.08 | 0.07 | 0.08 | 0.10 | 3.51 |
|  |  | RSE | 0.81 | 0.09 | 1.69 | 0.21 | 0.29 | 0.08 | 0.08 | 0.08 | 0.07 | 0.06 | 0.07 | 0.09 | 3.66 |
| QELS | 0.0 | Mean | 2.39 | 0.23 | 15.18 | -1.19 | 1.88 | -0.01 | -0.01 | 0.00 | 0.00 | -0.01 | 0.01 | -0.00 | 48.83 |
|  |  | $\mathrm{SE}_{S}$ | 0.95 | 0.12 | 1.78 | 0.14 | 0.25 | 0.15 | 0.15 | 0.16 | 0.16 | 0.15 | 0.15 | 0.15 | 3.55 |
|  |  | $\mathrm{SE}_{M}$ | 0.91 | 0.10 | 1.77 | 0.13 | 0.22 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 3.69 |
|  |  | RSE | 0.90 | 0.10 | 1.77 | 0.13 | 0.22 | 0.12 | 0.12 | 0.13 | 0.14 | 0.14 | 0.13 | 0.13 | 3.67 |
|  | 0.4 | Mean | 2.55 | 0.23 | 14.92 | -1.18 | 1.88 | 0.24 | 0.26 | 0.27 | 0.31 | 0.32 | 0.29 | 0.30 | 48.88 |
|  |  | $\mathrm{SE}_{S}$ | 0.91 | 0.11 | 1.80 | 0.13 | 0.22 | 0.13 | 0.14 | 0.13 | 0.12 | 0.13 | 0.13 | 0.13 | 3.39 |
|  |  | $\mathrm{SE}_{M}$ | 0.89 | 0.10 | 1.76 | 0.13 | 0.22 | 0.13 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 3.67 |
|  |  | RSE | 0.88 | 0.10 | 1.76 | 0.13 | 0.22 | 0.11 | 0.11 | 0.12 | 0.12 | 0.12 | 0.12 | 0.11 | 3.65 |
|  | 0.8 | Mean | 2.51 | 0.22 | 14.83 | -1.17 | 1.87 | 0.49 | 0.51 | 0.58 | 0.62 | 0.64 | 0.62 | 0.58 | 48.82 |
|  |  | $\mathrm{SE}_{S}$ | 0.88 | 0.10 | 1.67 | 0.12 | 0.21 | 0.08 | 0.10 | 0.08 | 0.07 | 0.07 | 0.07 | 0.08 | 3.45 |
|  |  | $\mathrm{SE}_{M}$ | 0.82 | 0.09 | 1.65 | 0.13 | 0.22 | 0.10 | 0.09 | 0.08 | 0.07 | 0.07 | 0.07 | 0.08 | 3.43 |
|  |  | RSE | 0.82 | 0.09 | 1.66 | 0.13 | 0.22 | 0.08 | 0.08 | 0.07 | 0.07 | 0.06 | 0.07 | 0.07 | 3.54 |
| Notes: $\mathrm{SE}_{S}$ - Standard deviation; $\mathrm{SE}_{M}$ - Average model-based standard errors; RSE- Average robust standard errors. $\rho$ is the correlation used for simulating the bivariate continuous outcomes. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Table 9.7: Comparison of the ELS and QELS approaches for joint nonlinear continuous and discrete outcomes

| Correlation between continuous and latent coninuous outcomes | Continuous variable parameter estimates | Discrete variable parameter estimates |
| :---: | :---: | :---: |
|  | ELS Method vs. QELS method | ELS Method vs. QELS method |
| 0.4 | Similar $\mathrm{SE}_{S}$ | Similar $\mathrm{SE}_{S}$ for probit, ELS has lower $\mathrm{SE}_{S}$ for logit, Cloglog |
|  | Similar difference between $\mathrm{SE}_{M}$ and $\mathrm{SE}_{S}$ | QELS has $\mathrm{SE}_{M}$ closer to observed $\mathrm{SE}_{S}$ than ELS |
|  | ELS: $\mathrm{SE}_{S} \geq \mathrm{SE}_{M}$ for probit, Cloglog | ELS: $\mathrm{SE}_{S} \geq \mathrm{SE}_{M}$ |
|  | QELS: $\mathrm{SE}_{S} \geq \mathrm{SE}_{M}$ for probit, Cloglog | QELS: $\mathrm{SE}_{S} \geq \mathrm{SE}_{M}$ |
| 0.8 | QELS has lower $\mathrm{SE}_{S}$ $\mathrm{SE}_{S}, \mathrm{SE}_{M}$, RSE decrease for both QELS and ELS method when corr. increase | QELS has lower SE $_{S}$ <br> $\mathrm{SE}_{S}$ decrease and $\mathrm{SE}_{M}$, RSE remain same for both QELS and ELS for probit, Cloglog when corr. increase $\mathrm{SE}_{S}, \mathrm{SE}_{M}$, RSE remain same for both QELS and ELS method for logit. when corr. increase |
|  | ELS: $\mathrm{SE}_{S} \geq \mathrm{SE}_{M}$ | ELS: $\mathrm{SE}_{S} \geq \mathrm{SE}_{M}$ |
|  | QELS: $\mathrm{SE}_{S} \geq \mathrm{SE}_{M}$ | QELS: $\mathrm{SE}_{S} \leq \mathrm{SE}_{M}$ |
|  | QELS has $\mathrm{SE}_{M}$ closer to observed $\mathrm{SE}_{S}$ than ELS | QELS has $\mathrm{SE}_{M}$ closer to observed $\mathrm{SE}_{S}$ than ELS |
|  | QELS has similar RSE values with ELS | QELS has lower RSE values than ELS method |

### 9.6.2 The parameter estimates results from joint fitting and separate fitting

## QELS parameter estimates:

As shown in Tables 9.8 to 9.10 and Figures 9.1 to 9.7, the joint model parameter estimates for continuous efficacy Emax model and discrete safety probit model with the stronger bivariate correlation have lower MSE and more efficiency than the parameter estimates from the separate fitting. The relative efficiency of joint fitting vs. separate fitting ranged from 1.06 to 1.527 for bivariate correlation of 0.495 to 0.638 for different dose level. The efficiency of joint fitting vs. separate fitting ranged from 0.976-1.020 when the bivariate correlation is 0 . The joint parameter estimates tend to have less percent bias with stronger correlation ( $0.3 \%$ to $11.6 \%$ ) than those with no correlation ( $-4.3 \%$ to $16.3 \%$ ). With stronger correlation the parameter estimates from joint model fitting have lower MSE than the parameter estimates from the separate model fitting. The correlation estimates from the joint model at different dose levels have increased efficiency (1.076 to 1.508) with respect to the simple Pearson correlation coefficient estimates when the true correlations between the safety and efficacy variables increases. The percent bias for the joint model estimates becomes smaller as the bivariate correlation becomes stronger. The variance estimates for the continuous efficacy variable remain similar for different bivariate correlation.

Tables 9.11 to 9.16 show the joint model fitting with logit or complementary log$\log$ model for the discrete safety variable, the results are similar to the probit model fitting regarding to the pattern of bias, MSE and efficiency by comparing the joint model fitting with separate model fitting under different correlation strength.
Table 9.8: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and probit for safety outcome (QELS approach): $\rho=0$

| Parameter | True value | Separate Fitting* |  |  |  | Joint Fitting |  |  |  | $\begin{gathered} \text { Efficiency } \\ \hline \text { Joint } \\ \text { vs. sep. } \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  | Median | Mean (Std) | Bias(\%) | MSE | Median | Mean (Std) | Bias(\%) | MSE |  |
| e0 | 2.5 | 2.434 | 2.402(0.9481) | -0.098(-3.9\%) | 0.904 | 2.431 | 2.393(0.9513) | -0.107(-4.3\%) | 0.912 | 0.993 |
| ed50 | 0.2 | 0.208 | 0.232(0.1169) | 0.032(16.0\%) | 0.015 | 0.209 | 0.233(0.1157) | 0.033(16.3\%) | 0.014 | 1.020 |
| emax | 14.5 | 14.881 | 15.152(1.7845) | 0.652( $4.5 \%)$ | 3.594 | 14.875 | 15.183(1.7838) | 0.683( 4.7\%) | 3.632 | 1.001 |
| $\alpha$ | -1.166 | -1.178 | -1.187(0.1341) | -0.021( $1.8 \%$ ) | 0.018 | -1.173 | $1.186(0.1351)$ | -0.020(1.7\%) | 0.019 | 0.984 |
| $\beta$ | 1.853 | 1.869 | 1.885(0.2429) | 0.032(1.7\%) | 0.060 | 1.869 | 1.882(0.2458) | 0.029(1.5\%) | 0.061 | 0.976 |
| $\sigma_{Y}$ | 7 | 6.998 | 7.009(0.2561) | 0.009( (0.1\%) | 0.065 | 6.973 | $6.984(0.2542)$ | -0.016(-0.2\%) | 0.065 | 1.015 |
| $\widetilde{\rho}_{0}$ | 0 | -0.006 | -0.005(0.1359) | -0.005 | 0.018 | -0.006 | -0.008(0.1450) | -0.008 | 0.021 | 0.878 |
| $\widetilde{\rho}_{0.05}$ | 0 | -0.012 | -0.005(0.1438) | -0.005 | 0.021 | -0.010 | -0.010(0.1542) | -0.010 | 0.024 | 0.870 |
| $\widetilde{\rho}_{0.2}$ | 0 | -0.001 | -0.002(0.1485) | -0.002 | 0.022 | 0.005 | 0.001(0.1582) | 0.001 | 0.025 | 0.881 |
| $\widetilde{\rho}_{0.4}$ | 0 | -0.005 | 0.007(0.1609) | 0.007 | 0.026 | 0.001 | 0.005(0.1596) | 0.005 | 0.025 | 1.016 |
| $\widetilde{\rho}_{0.6}$ | 0 | -0.007 | -0.010(0.1466) | -0.010 | 0.021 | -0.007 | -0.010(0.1496) | -0.010 | 0.022 | 0.960 |
| $\widetilde{\rho}_{0.8}$ | 0 | 0.017 | 0.013(0.1511) | 0.013 | 0.023 | 0.017 | 0.014(0.1540) | 0.014 | 0.024 | 0.963 |
| $\widetilde{\rho}_{1}$ | 0 | -0.002 | -0.003(0.1488) | -0.003 | 0.022 | -0.005 | -0.002(0.1491) | -0.002 | 0.022 | 0.995 |

Table 9.9: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and probit for safety outcome (QELS approach): $\rho=0.4$

Table 9.10: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and probit for safety outcome (QELS approach): $\rho=0.8$



Figure 9.1: Efficiency of parameter estimates based on separate and joint modeling - Emax for efficacy outcome and probit for safety outcome (QELS approach)


Figure 9.2: Percent bias of efficacy parameter estimates based on separate and joint modeling - Emax for efficacy outcome and probit for safety outcome (QELS approach)


Figure 9.3: Percent bias of safety parameter estimates based on separate and joint modeling - Emax for efficacy outcome and probit for safety outcome (QELS approach)


Figure 9.4: MSE of parameter estimates based on separate and joint modeling - Emax for efficacy outcome and probit for safety outcome (QELS approach)


Figure 9.5: Efficiency of parameter estimates for correlations based on simple Pearson correlation coefficient estimate and joint modeling - Emax for efficacy outcome and probit for safety outcome (QELS approach)


Figure 9.6: Percent bias of parameter estimates for correlations based on simple Pearson correlation coefficient estimate and joint modeling - Emax for efficacy outcome and probit for safety outcome (QELS approach)


Figure 9.7: MSE of parameter estimates for correlations based on simple Pearson correlation coefficient estimate and joint modeling - Emax for efficacy outcome and probit for safety outcome (QELS approach)
Table 9.11: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and logit for safety outcome (QELS approach): $\rho=0$

Table 9.12: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and logit for safety outcome (QELS approach): $\rho=0.4$

| Para- <br> meter | True value | Separate Fitting* |  |  |  | Joint Fitting |  |  |  | Efficiency |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  | Jointvs. sep. |
|  |  | Median | Mean (Std) | Bias(\%) | MSE |  | Median | Mean (Std) | Bias(\%) | MSE |
| e0 | 2.5 | 2.398 | 2.406(0.9307) | -0.094(-3.7\%) | 0.871 | 2.363 | 2.410(0.9066) | -0.090(-3.6\%) | 0.826 | 1.054 |
| ed50 | 0.2 | 0.201 | 0.229(0.1078) | 0.029(14.5\%) | 0.012 | 0.205 | 0.230(0.1172) | 0.030(14.9\%) | 0.015 | 0.845 |
| emax | 14.5 | 14.968 | 15.109(1.8707) | 0.609( 4.2\%) | 3.853 | 14.973 | 15.129(2.0318) | 0.629(4.3\%) | 4.504 | 0.848 |
| $\alpha$ | -2.123 | -2.123 | -2.145(0.2517) | -0.022( 1.0\%) | 0.064 | -2.114 | -2.151(0.2714) | -0.028( 1.3\%) | 0.074 | 0.860 |
| $\beta$ | 3.728 | 3.723 | 3.761(0.4479) | 0.033(0.9\%) | 0.201 | 3.713 | 3.775(0.4737) | 0.047( 1.3\%) | 0.226 | 0.894 |
| $\sigma_{Y}$ | 7 | 7.017 | 6.999(0.2792) | -0.001(-0.0\%) | 0.078 | 6.984 | 6.970(0.2774) | -0.030(-0.4\%) | 0.077 | 1.013 |
| $\widetilde{\rho}_{0}$ | . 239 | 0.237 | 0.238(0.1347) | -0.001(-0.4\%) | 0.018 | 0.250 | 0.243(0.1319) | 0.004( 1.8\%) | 0.017 | 1.043 |
| $\widetilde{\rho}_{0.05}$ | . 250 | 0.251 | 0.235(0.1277) | -0.015(-5.9\%) | 0.016 | 0.259 | 0.242(0.1297) | -0.008(-3.2\%) | 0.017 | 0.968 |
| $\widetilde{\rho}_{0.2}$ | . 280 | 0.281 | 0.276(0.1275) | -0.004(-1.5\%) | 0.016 | 0.273 | 0.274(0.1253) | -0.006(-2.2\%) | 0.016 | 1.036 |
| $\widetilde{\rho}_{0.4}$ | . 310 | 0.317 | 0.315(0.1258) | 0.004( 1.3\%) | 0.016 | 0.317 | 0.315(0.1256) | 0.005( 1.6\%) | 0.016 | 1.004 |
| $\widetilde{\rho}_{0.6}$ | . 319 | 0.309 | 0.312(0.1216) | -0.007(-2.1\%) | 0.015 | 0.303 | 0.311(0.1214) | -0.008(-2.5\%) | 0.015 | 1.003 |
| $\widetilde{\rho}_{0.8}$ | . 303 | 0.295 | 0.299(0.1310) | -0.004(-1.2\%) | 0.017 | 0.301 | 0.301(0.1292) | -0.003(-0.8\%) | 0.017 | 1.028 |
| $\widetilde{\rho}_{1}$ | . 269 | 0.262 | 0.258(0.1346) | -0.011(-4.0\%) | 0.018 | 0.268 | 0.259(0.1334) | -0.010(-3.6\%) | 0.018 | 1.017 |
| Notes: outputs are from 500 simulations and $\mathrm{n}=50$ per dose group. $\rho$ is the correlation used for simulating the bivariate continuous outcomes. <br> *: All $\widetilde{\rho}$ s for separate fitting are obtained from Pearson correlation coefficient between continuous and discrete variable for each dose level. |  |  |  |  |  |  |  |  |  |  |

Table 9.13: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and logit for safety outcome (QELS approach): $\rho=0.8$

| Parameter | True value | Separate Fitting* |  |  |  | Joint Fitting |  |  |  | EfficiencyJointvs. sep. |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  | Median | Mean (Std) | Bias(\%) | MSE | Median | Mean (Std) | Bias(\%) | MSE |  |
| e0 | 2.5 | 2.558 | 2.562(0.9297) | 0.062( $2.5 \%$ ) | 0.864 | 2.538 | $2.510(0.8697)$ | 0.010( 0.4\%) | 0.753 | 1.143 |
| ed50 | 0.2 | 0.206 | 0.238(0.1182) | 0.038(19.1\%) | 0.015 | 0.204 | 0.223(0.0919) | 0.023(11.5\%) | 0.009 | 1.653 |
| emax | 14.5 | 14.819 | 4.981(1.8597) | 0.481(3.3\%) | 3.673 | 14.722 | 14.830(1.6593) | 0.330( $2.3 \%$ ) | 2.849 | 1.256 |
| $\alpha$ | -2.123 | -2.138 | $2.146(0.2513)$ | -0.023( 1.1\%) | 0.063 | -2.131 | -2.147(0.2437) | -0.024(1.2\%) | 0.060 | 1.064 |
| $\beta$ | 3.728 | 3.781 | 3.788(0.4254) | 0.060(1.6\%) | 0.184 | 3.761 | 3.789(0.4139) | 0.061( 1.6\%) | 0.174 | 1.056 |
| $\sigma_{y}$ | 7 | 7.010 | 7.017(0.2457) | 0.017(0.2\%) | 0.060 | 6.980 | 6.983(0.2464) | -0.017(-0.2\%) | 0.061 | 0.994 |
| $\widetilde{\rho}_{0}$ | . 477 | 0.469 | 0.467(0.1053) | -0.010(-2.0\%) | 0.011 | 0.473 | 0.471(0.0863) | -0.006(-1.2\%) | 0.007 | 1.489 |
| $\widetilde{\rho}_{0.05}$ | . 499 | 0.509 | 0.495(0.1200) | -0.004(-0.8\%) | 0.014 | 0.512 | 0.501(0.1063) | 0.002 (0.3\%) | 0.011 | 1.274 |
| $\widetilde{\rho}_{0.2}$ | . 561 | 0.555 | 0.560(0.0912) | -0.001(-0.2\%) | 0.008 | 0.563 | 0.564(0.0780) | 0.004(0.6\%) | 0.006 | 1.370 |
| $\widetilde{\rho}_{0.4}$ | . 621 | 0.626 | 0.621(0.0807) | 0.001( 0.1\%) | 0.006 | 0.624 | $0.622(0.0746)$ | 0.001( 0.2\%) | 0.006 | 1.171 |
| $\widetilde{\rho}_{0.6}$ | . 638 | 0.644 | 0.639(0.0773) | 0.001( 0.1\%) | 0.006 | 0.644 | $0.639(0.0716)$ | 0.001( 0.2\%) | 0.005 | 1.166 |
| $\widetilde{\rho}_{0.8}$ | . 606 | 0.601 | 0.599(0.0808) | -0.007(-1.2\%) | 0.007 | 0.603 | $0.600(0.0715)$ | -0.006(-1.0\%) | 0.005 | 1.278 |
| $\widetilde{\rho}_{1}$ | . 537 | 0.530 | 0.528(0.1041) | -0.009(-1.8\%) | 0.011 | 0.531 | 0.535(0.0843) | -0.002(-0.4\%) | 0.007 | 1.525 |

Table 9.14: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and Cloglog for safety outcome (QELS approach): $\rho=0$

Table 9.15: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and Cloglog for safety outcome (QELS approach): $\rho=0.4$

| Parameter | True value | Separate Fitting* |  |  |  | Joint Fitting |  |  |  | $\begin{gathered} \hline \text { Efficiency } \\ \hline \text { Joint } \\ \text { vs. sep. } \\ \hline \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  | Median | Mean (Std) | Bias(\%) | MSE | Median | Mean (Std) | Bias(\%) | MSE |  |
| e0 | 2.5 | 2.584 | 2.562(0.9269) | 0.062( $2.5 \%$ ) | 0.859 | 2.547 | 2.548(0.9074) | 0.048( $1.9 \%$ ) | 0.822 | 1.044 |
| ed50 | 0.2 | 0.205 | 0.237(0.1163) | 0.037(18.6\%) | 0.015 | 0.211 | 0.233(0.1100) | 0.033(16.7\%) | 0.013 | 1.117 |
| emax | 14.5 | 14.738 | 14.966(1.8308) | 0.466(3.2\%) | 3.552 | 14.875 | 14.933(1.8040) | 0.433(3.0\%) | 3.425 | 1.030 |
| $\alpha$ | -2.123 | -1.167 | -1.180(0.1296) | -0.014( 1.2\%) | 0.017 | -1.164 | -1.179(0.1333) | -0.013(1.1\%) | 0.018 | 0.945 |
| $\beta$ | 2.828 | 2.875 | 2.880(0.2224) | 0.027( $1.4 \%$ ) | 0.050 | 2.870 | 2.882(0.2270) | 0.029(1.6\%) | 0.052 | 0.959 |
| $\sigma_{Y}$ | 7 | 7.006 | 7.017(0.2437) | 0.017( (0.2\%) | 0.059 | 6.979 | 6.987(0.2417) | -0.013(-0.2\%) | 0.058 | 1.017 |
| $\widetilde{\rho}_{0}$ | . 242 | 0.237 | 0.241(0.1302) | -0.006(-2.5\%) | 0.017 | 0.242 | 0.241(0.1272) | -0.007(-2.7\%) | 0.016 | 1.048 |
| $\widetilde{\rho}_{0.05}$ | . 252 | 0.267 | 0.254(0.1361) | -0.004(-1.4\%) | 0.018 | 0.263 | 0.257(0.1386) | -0.001(-0.3\%) | 0.019 | 0.964 |
| $\widetilde{\rho}_{0.2}$ | . 277 | 0.276 | 0.272(0.1300) | -0.012(-4.2\%) | 0.017 | 0.279 | 0.273(0.1272) | -0.011(-4.0\%) | 0.016 | 1.045 |
| $\widetilde{\rho}_{0.4}$ | . 305 | 0.312 | 0.313(0.1275) | 0.004( 1.2\%) | 0.016 | 0.316 | 0.310(0.1228) | 0.002( 0.5\%) | 0.015 | 1.079 |
| $\widetilde{\rho}_{0.6}$ | . 319 | 0.326 | 0.315(0.1351) | -0.004(-1.2\%) | 0.018 | 0.319 | 0.315(0.1306) | -0.005(-1.4\%) | 0.017 | 1.071 |
| $\widetilde{\rho}_{0.8}$ | . 306 | 0.285 | 0.288(0.1327) | -0.026(-8.2\%) | 0.018 | 0.283 | 0.288(0.1313) | -0.026(-8.2\%) | 0.018 | 1.022 |
| $\widetilde{\rho}_{1}$ | . 253 | 0.307 | 0.291(0.1339) | -0.001(-0.5\%) | 0.018 | 0.305 | 0.294(0.1285) | 0.002( 0.5\%) | 0.016 | 1.086 |
| Notes: outputs are from 500 simulations and $n=50$ per dose group. $\rho$ is the correlation used for simulating the bivariate continuous outcomes. <br> *: All $\widetilde{\rho}_{\text {s for separate fitting are obtained from Pearson correlation coefficient between continuous and discrete variable for each dose lever }}$ |  |  |  |  |  |  |  |  |  |  |

Table 9.16: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and Cloglog for safety outcome (QELS approach): $\rho=0.8$

| Parameter | True value | Separate Fitting* |  |  |  | Joint Fitting |  |  |  | Efficiency <br> Joint <br> vs. sep. |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  | Median | Mean (Std) | Bias(\%) | MSE | Median | Mean (Std) | Bias(\%) | MSE |  |
| e0 | 2.5 | 2.558 | 2.562(0.9297) | 0.062( $2.5 \%$ ) | 0.864 | 2.539 | 2.508(0.8651) | 0.008( 0.3\%) | 0.745 | 1.155 |
| ed50 | 0.2 | 0.206 | 0.238(0.1182) | 0.038(19.1\%) | 0.015 | 0.207 | 0.225(0.0976) | 0.025(12.7\%) | 0.010 | 1.465 |
| emax | 14.5 | 14.819 | 4.981(1.8597) | 0.481( $3.3 \%)$ | 3.673 | 14.726 | 14.880(1.7212) | 0.380( $2.6 \%$ ) | 3.092 | 1.167 |
| $\alpha$ | -2.123 | -2.128 | 2.135(0.2157) | 0.012( $0.6 \%)$ | 0.046 | -2.113 | -2.139(0.2100) | -0.016(0.8\%) | 0.044 | 1.054 |
| $\beta$ | 3.728 | 3.845 | 3.859(0.3050) | 0.031( $1.1 \%$ ) | 0.094 | 3.859 | 3.865(0.2956) | 0.037( $1.3 \%$ ) | 0.088 | 1.065 |
| $\sigma_{Y}$ | 7 | 7.010 | 7.017(0.2457) | 0.017( $0.2 \%$ ) | 0.060 | 6.984 | 6.982(0.2451) | -0.018(-0.3\%) | 0.060 | 1.005 |
| $\widetilde{\rho}_{0}$ | . 484 | 0.474 | 0.476(0.1012) | -0.008(-1.7\%) | 0.010 | 0.486 | 0.480(0.0810) | -0.004(-0.9\%) | 0.007 | 1.561 |
| $\widetilde{\rho}_{0.05}$ | . 502 | 0.510 | 0.498(0.1160) | -0.004(-0.7\%) | 0.013 | 0.509 | 0.504(0.1026) | 0.002( $0.4 \%)$ | 0.010 | 1.279 |
| $\widetilde{\rho}_{0.2}$ | . 553 | 0.551 | 0.553(0.0968) | -0.000(-0.1\%) | 0.009 | 0.562 | 0.559(0.0833) | 0.005(0.9\%) | 0.007 | 1.352 |
| $\widetilde{\rho}_{0.4}$ | . 610 | 0.620 | 0.615(0.0795) | 0.005(0.8\%) | 0.006 | 0.618 | 0.615(0.0732) | 0.005(0.7\%) | 0.005 | 1.178 |
| $\widetilde{\rho}_{0.6}$ | . 638 | 0.650 | 0.642(0.0739) | 0.004(0.6\%) | 0.005 | 0.646 | 0.643(0.0696) | 0.004( 0.7\%) | 0.005 | 1.129 |
| $\widetilde{\rho}_{0.8}$ | . 612 | 0.617 | 0.609(0.0792) | -0.004(-0.6\%) | 0.006 | 0.614 | 0.608(0.0716) | -0.004(-0.7\%) | 0.005 | 1.223 |
| $\widetilde{\rho}_{1}$ | . 506 | 0.505 | 0.495(0.1096) | -0.011(-2.2\%) | 0.012 | 0.500 | 0.499(0.0909) | -0.007(-1.4\%) | 0.008 | 1.454 |

## ELS parameter estimates:

The joint model estimates from the ELS approach (Emax model for continuous efficacy outcome and probit model for discrete safety outcome) are shown in Tables 9.17 to 9.19 , the joint model parameter estimates with the stronger bivariate correlation have lower MSE and more efficiency than the parameter estimates from the separate fitting. The efficiency of joint fitting vs. separate fitting ranged from 0.939-1.494 for bivariate correlation of 0.495 to 0.638 at different dose level while the efficiency of joint fitting vs. separate fitting ranged from 0.949-1.033 for no bivariate correlation. The joint parameter estimates tend to have less percent bias ( $0.4 \%$ to $11.8 \%$ ) with stronger correlation compared with no correlation ( $-4.4 \%$ to $15.8 \%$ ). The joint model estimates for correlation at different dose levels have increased efficiency ( 1.092 to 1.486 ) with the bivariate correlation of ( 0.495 to 0.638 ) by comparing to the pearson correlation coefficient estimates for the bivariate outcomes. The percent bias is relative smaller for the joint estimates than for separate estimates when the bivariate correlation become stronger. The variance estimate for the continuous efficacy variable are similar for different bivariate correlation.

Tables 9.20 to 9.25 show the joint model fitting with logit or complementary log$\log$ model for the discrete safety variable, the results are similar to the probit model fitting regarding to the pattern of percent bias, MSE and efficiency by comparing the joint model fitting with separate model fitting under the different correlation strength.

In general, the discrete variable regression parameter estimates does not improve much with joint modeling when correlation becomes stronger. This is not surprising since the discrete variable comprises much less information than latent continuous variable af-
ter dichotomization. Model-based and robust standard error remain similar for discrete variable parameter estimates regardless of correlation for ELS method. But overall the percent bias are still relatively small ( $\leq 2.3 \%$ ). By comparing QELS and ELS approach for the nonlinear joint modeling of mixed type outcomes, QELS method yields more efficient parameter estimates and is more robust under the various strength of bivariate correlation and the different models selected for the discrete safety variable.
Table 9.17: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and probit for safety outcome (ELS approach): $\rho=0$

| Parameter | True value | Separate Fitting* |  |  |  | Joint Fitting |  |  |  | $\begin{gathered} \text { Efficiency } \\ \hline \text { Joint } \\ \text { vs. sep. } \\ \hline \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  | Median | Mean (Std) | Bias(\%) | MSE | Median | Mean (Std) | Bias(\%) | MSE |  |
| e0 | 2.5 | 2.434 | 2.402(0.9481) | -0.098(-3.9\%) | 0.904 | 2.387 | 2.389(0.9543) | -0.111(-4.4\%) | 0.918 | 0.987 |
| ed50 | 0.2 | 0.208 | 0.232(0.1169) | 0.032(16.0\%) | 0.015 | 0.208 | 0.232(0.1156) | 0.032(15.8\%) | 0.014 | 1.022 |
| emax | 14.5 | 14.881 | 15.152(1.7845) | 0.652( $4.5 \%)$ | 3.594 | 14.837 | 15.163(1.7557) | 0.663(4.6\%) | 3.507 | 1.033 |
| $\alpha$ | -1.166 | -1.178 | -1.187(0.1341) | -0.021( 1.8\%) | 0.018 | -1.182 | -1.189(0.1376) | -0.023( 1.9\%) | 0.019 | 0.949 |
| $\beta$ | 1.853 | 1.869 | 1.885(0.2429) | 0.032(1.7\%) | 0.060 | 1.880 | $1.888(0.2492)$ | 0.035(1.9\%) | 0.063 | 0.950 |
| $\sigma_{Y}$ | 7 | 6.998 | 7.009(0.2561) | 0.009( $0.1 \%$ ) | 0.065 | 6.974 | 6.984(0.2533) | -0.016(-0.2\%) | 0.064 | 1.023 |
| $\widetilde{\rho}_{0}$ | 0 | -0.006 | -0.005(0.1359) | -0.005( \%) | 0.018 | -0.005 | -0.008(0.1439) | -0.008( \%) | 0.021 | 0.892 |
| $\widetilde{\rho}_{0.05}$ | 0 | -0.012 | -0.005(0.1438) | -0.005(\%) | 0.021 | -0.014 | -0.010(0.1528) | -0.010(\%) | 0.023 | 0.886 |
| $\widetilde{\rho}_{0.2}$ | 0 | -0.001 | -0.002(0.1485) | -0.002(\%) | 0.022 | 0.002 | -0.001(0.1604) | -0.001(\%) | 0.026 | 0.857 |
| $\widetilde{\rho}_{0.4}$ | 0 | -0.005 | 0.007(0.1609) | 0.007( \%) | 0.026 | -0.001 | 0.002(0.1593) | 0.002( \%) | 0.025 | 1.019 |
| $\widetilde{\rho}_{0.6}$ | 0 | -0.007 | -0.010(0.1466) | -0.010( \%) | 0.021 | -0.007 | -0.010(0.1493) | -0.010(\%) | 0.022 | 0.964 |
| $\widetilde{\rho}_{0.8}$ | 0 | 0.017 | 0.013(0.1511) | 0.013(\%) | 0.023 | 0.018 | 0.013(0.1541) | 0.013(\%) | 0.024 | 0.961 |
| $\widetilde{\rho}_{1}$ | 0 | -0.002 | -0.003(0.1488) | -0.003(\%) | 0.022 | -0.007 | -0.004(0.1489) | -0.004(\%) | 0.022 | 0.998 |

Table 9.18: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and probit for safety outcome (ELS approach): $\rho=0.4$

| Parameter | True value | Separate Fitting* |  |  |  | Joint Fitting |  |  |  | Efficiency <br> Joint <br> vs. sep. |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  | Median | Mean (Std) | Bias(\%) | MSE | Median | Mean (Std) | Bias(\%) | MSE |  |
| e0 | 2.5 | 2.584 | 2.562(0.9269) | 0.062( $2.5 \%$ ) | 0.859 | 2.547 | 2.548(0.9074) | 0.048( $1.9 \%$ ) | 0.822 | 1.044 |
| ed50 | 0.2 | 0.205 | 0.237(0.1163) | 0.037(18.6\%) | 0.015 | 0.211 | 0.233(0.1100) | 0.033(16.7\%) | 0.013 | 1.117 |
| emax | 14.5 | 14.738 | 14.966(1.8308) | 0.466(3.2\%) | 3.552 | 14.875 | 14.933(1.8040) | 0.433(3.0\%) | 3.425 | 1.030 |
| $\alpha$ | -1.166 | -1.167 | -1.180(0.1296) | -0.014(1.2\%) | 0.017 | -1.164 | -1.179(0.1333) | -0.013( (1.1\%) | 0.018 | 0.945 |
| $\beta$ | 1.853 | 1.875 | 1.880(0.2224) | 0.027( $1.4 \%$ ) | 0.050 | 1.870 | 1.882(0.2270) | 0.029(1.6\%) | 0.052 | 0.959 |
| $\sigma_{Y}$ | 7.0 | 7.006 | 7.017(0.2437) | 0.017( $0.2 \%$ ) | 0.059 | 6.979 | 6.987(0.2417) | -0.013(-0.2\%) | 0.058 | 1.017 |
| $\widetilde{\rho}_{0}$ | . 247 | 0.237 | 0.241(0.1302) | -0.006(-2.5\%) | 0.017 | 0.242 | 0.241(0.1272) | -0.007(-2.7\%) | 0.016 | 1.048 |
| $\widetilde{\rho}_{0.05}$ | . 257 | 0.267 | 0.254(0.1361) | -0.004(-1.4\%) | 0.018 | 0.263 | 0.257(0.1386) | -0.001(-0.3\%) | 0.019 | 0.964 |
| $\widetilde{\rho}_{0.2}$ | . 284 | 0.276 | 0.272(0.1300) | -0.012(-4.2\%) | 0.017 | 0.279 | 0.273(0.1272) | -0.011(-4.0\%) | 0.016 | 1.045 |
| $\widetilde{\rho}_{0.4}$ | . 310 | 0.312 | 0.313(0.1275) | 0.004( $1.2 \%$ ) | 0.016 | 0.316 | 0.310(0.1228) | 0.002( 0.5\%) | 0.015 | 1.079 |
| $\widetilde{\rho}_{0.6}$ | . 320 | 0.326 | 0.315(0.1351) | -0.004(-1.2\%) | 0.018 | 0.319 | 0.315(0.1306) | -0.005(-1.4\%) | 0.017 | 1.071 |
| $\widetilde{\rho}_{0.8}$ | . 313 | 0.285 | 0.288(0.1327) | -0.026(-8.2\%) | 0.018 | 0.283 | 0.288(0.1313) | -0.026(-8.2\%) | 0.018 | 1.022 |
| $\widetilde{\rho}_{1}$ | . 247 | 0.307 | 0.291(0.1339) | -0.001(-0.5\%) | 0.018 | 0.305 | 0.294(0.1285) | 0.002( 0.5\%) | 0.016 | 1.086 |
| Notes: outputs are from 500 simulations and $n=50$ per dose group. $\rho$ is the correlation used for simulating the bivariate continuous outcomes. <br> *: All $\widetilde{\rho}$ s for separate fitting are obtained from Pearson correlation coefficient between continuous and discrete variable for each dose level. |  |  |  |  |  |  |  |  |  |  |

Table 9.19: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and probit for safety outcome (ELS approach): $\rho=0.8$

| Parameter | True value | Separate Fitting* |  |  |  | Joint Fitting |  |  |  | Efficiency <br> Joint <br> vs. sep. |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  | Median | Mean (Std) | Bias(\%) | MSE | Median | Mean (Std) | Bias(\%) | MSE |  |
| e0 | 2.5 | 2.558 | 2.562(0.9297) | 0.062( $2.5 \%$ ) | 0.864 | 2.483 | 2.511(0.8748) | 0.011( 0.4\%) | 0.762 | 1.129 |
| ed50 | 0.2 | 0.206 | 0.238(0.1182) | 0.038(19.1\%) | 0.015 | 0.210 | 0.224(0.0967) | 0.024(11.8\%) | 0.010 | 1.494 |
| emax | 14.5 | 14.819 | 14.981(1.8597) | 0.481( $3.3 \%)$ | 3.673 | 14.601 | 14.822(1.6994) | 0.322( $2.2 \%$ ) | 2.977 | 1.197 |
| $\alpha$ | -1.166 | -1.168 | -1.172(0.1283) | -0.006(0.5\%) | 0.016 | -1.172 | -1.175(0.1324) | -0.009( $0.7 \%$ ) | 0.018 | 0.939 |
| $\beta$ | 1.853 | 1.883 | 1.873(0.2238) | 0.020( $1.1 \%$ ) | 0.050 | 1.853 | 1.875(0.2237) | 0.022( $1.2 \%$ ) | 0.050 | 1.001 |
| $\sigma_{Y}$ | 7.0 | 7.010 | 7.017(0.2457) | 0.017( $0.2 \%$ ) | 0.060 | 6.978 | 6.980(0.2470) | -0.020(-0.3\%) | 0.061 | 0.989 |
| $\widetilde{\rho}_{0}$ | . 495 | 0.494 | 0.488(0.0996) | -0.006(-1.3\%) | 0.010 | 0.498 | 0.491(0.0817) | -0.004(-0.8\%) | 0.007 | 1.486 |
| $\widetilde{\rho}_{0.05}$ | . 515 | 0.523 | 0.509(0.1144) | -0.006(-1.2\%) | 0.013 | 0.522 | 0.516(0.1002) | 0.001( $0.2 \%)$ | 0.010 | 1.305 |
| $\widetilde{\rho}_{0.2}$ | . 568 | 0.581 | 0.572(0.0896) | 0.004( (0.7\%) | 0.008 | 0.571 | 0.576(0.0769) | 0.008( $1.3 \%$ ) | 0.006 | 1.357 |
| $\widetilde{\rho}_{0.4}$ | . 618 | 0.623 | 0.620(0.0794) | 0.002( $0.3 \%)$ | 0.006 | 0.623 | 0.619(0.0735) | 0.002( 0.3\%) | 0.005 | 1.167 |
| $\widetilde{\rho}_{0.6}$ | . 638 | 0.648 | 0.642(0.0740) | 0.004(0.6\%) | 0.005 | 0.646 | 0.643(0.0700) | 0.004( 0.7\%) | 0.005 | 1.118 |
| $\widetilde{\rho}_{0.8}$ | . 627 | 0.628 | 0.623(0.0702) | -0.004(-0.6\%) | 0.005 | 0.632 | 0.621(0.0672) | -0.006(-0.9\%) | 0.005 | 1.092 |
| $\widetilde{\rho}_{1}$ | . 585 | 0.592 | 0.578(0.0982) | -0.007(-1.2\%) | 0.010 | 0.584 | 0.582(0.0824) | -0.004(-0.6\%) | 0.007 | 1.421 |
| Notes: outputs are from 500 simulations and $n=50$ per dose group. $\rho$ is the correlation used for simulating the bivariate continuous outcomes. <br> *: All $\widetilde{\rho}_{\text {s }}$ for separate fitting are obtained from Pearson correlation coefficient between continuous and discrete variable for each dose level. |  |  |  |  |  |  |  |  |  |  |

Table 9.20: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and logit for safety outcome (ELS approach): $\rho=0$

| Parameter | True value | Separate Fitting* |  |  |  | Joint Fitting |  |  |  | $\begin{gathered} \text { Efficiency } \\ \hline \text { Joint } \\ \text { vs. sep. } \\ \hline \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  | Median | Mean (Std) | Bias(\%) | MSE | Median | Mean (Std) | Bias(\%) | MSE |  |
| e0 | 2.5 | 2.601 | 2.550(0.8825) | 0.050( $2.0 \%$ ) | 0.777 | 2.599 | 2.548(0.8686) | 0.048( $1.9 \%$ ) | 0.753 | 1.032 |
| ed50 | 0.2 | 0.215 | 0.241(0.1127) | 0.041(20.3\%) | 0.014 | 0.214 | 0.240(0.1167) | 0.040(20.1\%) | 0.015 | 0.932 |
| emax | 14.5 | 14.915 | 15.051(1.6381) | 0.551( $3.8 \%$ ) | 2.973 | 14.829 | 15.057(1.6907) | 0.557(3.8\%) | 3.154 | 0.939 |
| $\alpha$ | -2.123 | -2.147 | -2.168(0.2414) | -0.045( $2.1 \%$ ) | 0.060 | -2.149 | -2.167(0.2572) | -0.044( $2.1 \%$ ) | 0.068 | 0.881 |
| $\beta$ | 3.728 | 3.767 | 3.805(0.4291) | 0.077( $2.1 \%$ ) | 0.189 | 3.801 | 3.814(0.4537) | 0.086( $2.3 \%$ ) | 0.212 | 0.894 |
| $\sigma_{Y}$ | 7 | 7.011 | 7.011(0.2692) | 0.011( $0.2 \%$ ) | 0.072 | 6.987 | 6.986(0.2707) | -0.014(-0.2\%) | 0.073 | 0.990 |
| $\widetilde{\rho}_{0}$ | 0 | -0.017 | -0.009(0.1453) | -0.009(\%) | 0.021 | -0.021 | -0.015(0.1547) | -0.015(\%) | 0.024 | 0.882 |
| $\widetilde{\rho}_{0.05}$ | 0 | 0.003 | -0.009(0.1427) | -0.009(\%) | 0.020 | 0.001 | -0.008(0.1534) | -0.008( \%) | 0.023 | 0.866 |
| $\widetilde{\rho}_{0.2}$ | 0 | -0.031 | -0.024(0.1379) | -0.024( \%) | 0.020 | -0.028 | -0.023(0.1391) | -0.023(\%) | 0.020 | 0.984 |
| $\widetilde{\rho}_{0.4}$ | 0 | -0.002 | -0.000(0.1441) | -0.000(\%) | 0.021 | -0.002 | 0.002(0.1468) | 0.002(\%) | 0.021 | 0.963 |
| $\widetilde{\rho}_{0.6}$ | 0 | 0.007 | 0.006(0.1537) | 0.006( \%) | 0.024 | 0.007 | 0.006(0.1532) | 0.006( \%) | 0.023 | 1.007 |
| $\widetilde{\rho}_{0.8}$ | 0 | -0.005 | -0.009(0.1471) | -0.009(\%) | 0.022 | -0.007 | -0.009(0.1515) | -0.009(\%) | 0.023 | 0.943 |
| $\widetilde{\rho}_{1}$ | 0 | -0.006 | 0.002(0.1486) | 0.002( \%) | 0.022 | -0.002 | 0.004(0.1512) | 0.004(\%) | 0.023 | 0.967 |

Table 9.21: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and logit for safety outcome (ELS approach): $\rho=0.4$

| Parameter | True value | Separate Fitting* |  |  |  | Joint Fitting |  |  |  | EfficiencyJointvs. sep. |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  | Median | Mean (Std) | Bias(\%) | MSE | Median | Mean (Std) | Bias(\%) | MSE |  |
| e0 | 2.5 | 2.398 | 2.406(0.9307) | -0.094(-3.7\%) | 0.871 | 2.363 | 2.410(0.9066) | -0.090(-3.6\%) | 0.826 | 1.054 |
| ed50 | 0.2 | 0.201 | 0.229(0.1078) | 0.029(14.5\%) | 0.012 | 0.205 | 0.230(0.1172) | 0.030(14.9\%) | 0.015 | 0.845 |
| emax | 14.5 | 14.968 | 15.109(1.8707) | 0.609(4.2\%) | 3.853 | 14.973 | 15.129(2.0318) | 0.629(4.3\%) | 4.504 | 0.848 |
| $\alpha$ | -2.123 | -2.123 | -2.145(0.2517) | -0.022 ( 1.0\%) | 0.064 | -2.114 | -2.151(0.2714) | -0.028( $1.3 \%$ ) | 0.074 | 0.860 |
| $\beta$ | 3.728 | 3.723 | 3.761(0.4479) | 0.033( $0.9 \%$ ) | 0.201 | 3.713 | 3.775(0.4737) | 0.047( $1.3 \%$ ) | 0.226 | 0.894 |
| $\sigma_{Y}$ | 7 | 7.017 | 6.999(0.2792) | -0.001(-0.0\%) | 0.078 | 6.984 | 6.970(0.2774) | -0.030(-0.4\%) | 0.077 | 1.013 |
| $\widetilde{\rho}_{0}$ | . 239 | 0.237 | 0.238(0.1347) | -0.001(-0.4\%) | 0.018 | 0.250 | 0.243(0.1319) | 0.004( $1.8 \%$ ) | 0.017 | 1.043 |
| $\widetilde{\rho}_{0.05}$ | . 250 | 0.251 | 0.235(0.1277) | -0.015(-5.9\%) | 0.016 | 0.259 | 0.242(0.1297) | -0.008(-3.2\%) | 0.017 | 0.968 |
| $\widetilde{\rho}_{0.2}$ | . 280 | 0.281 | 0.276(0.1275) | -0.004(-1.5\%) | 0.016 | 0.273 | 0.274(0.1253) | -0.006(-2.2\%) | 0.016 | 1.036 |
| $\widetilde{\rho}_{0.4}$ | . 310 | 0.317 | 0.315(0.1258) | 0.004( $1.3 \%$ ) | 0.016 | 0.317 | 0.315(0.1256) | 0.005( 1.6\%) | 0.016 | 1.004 |
| $\widetilde{\rho}_{0.6}$ | . 319 | 0.309 | 0.312(0.1216) | -0.007(-2.1\%) | 0.015 | 0.303 | 0.311(0.1214) | -0.008(-2.5\%) | 0.015 | 1.003 |
| $\widetilde{\rho}_{0.8}$ | . 303 | 0.295 | 0.299(0.1310) | -0.004(-1.2\%) | 0.017 | 0.301 | 0.301(0.1292) | -0.003(-0.8\%) | 0.017 | 1.028 |
| $\widetilde{\rho}_{1}$ | . 269 | 0.262 | 0.258(0.1346) | -0.011(-4.0\%) | 0.018 | 0.268 | 0.259(0.1334) | -0.010(-3.6\%) | 0.018 | 1.017 |
| Notes: outputs are from 500 simulations and $n=50$ per dose group. $\rho$ is the correlation used for simulating the bivariate continuous outcomes. <br> *: All $\widetilde{\rho}$ s for separate fitting are obtained from Pearson correlation coefficient between continuous and discrete variable for each dose level. |  |  |  |  |  |  |  |  |  |  |

Table 9.22: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and logit for safety outcome (ELS approach): $\rho=0.8$

| Parameter | True value | Separate Fitting* |  |  |  | Joint Fitting |  |  |  | Efficiency |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  | Joint vs. sep. |
|  |  | Median | Mean (Std) | Bias(\%) | MSE |  | Median | Mean (Std) | Bias(\%) | MSE |
| e0 | 2.5 | 2.558 | 2.562(0.9297) | 0.062( 2.5\%) | 0.864 | 2.527 | 2.513(0.8701) | 0.013( 0.5\%) | 0.753 | 1.142 |
| ed50 | 0.2 | 0.206 | 0.238(0.1182) | 0.038(19.1\%) | 0.015 | 0.208 | 0.223(0.0919) | 0.023(11.6\%) | 0.009 | 1.654 |
| emax | 14.5 | 14.819 | 14.981(1.8597) | 0.481( 3.3\%) | 3.673 | 14.754 | 14.834(1.6903) | 0.334( $2.3 \%)$ | 2.955 | 1.210 |
| $\alpha$ | -2.123 | -2.138 | -2.146(0.2513) | -0.023( 1.1\%) | 0.063 | -2.108 | -2.149(0.2665) | -0.026( 1.2\%) | 0.071 | 0.889 |
| $\beta$ | 3.728 | 3.781 | 3.788(0.4254) | 0.060( 1.6\%) | 0.184 | 3.763 | 3.796(0.4499) | 0.068( 1.8\%) | 0.206 | 0.894 |
| $\sigma_{e}$ | 7 | 7.010 | 7.017(0.2457) | 0.017( 0.2\%) | 0.060 | 6.976 | 6.980(0.2454) | -0.020(-0.3\%) | 0.060 | 1.002 |
| $\widetilde{\rho}_{0}$ | . 477 | 0.469 | 0.467(0.1053) | -0.010(-2.0\%) | 0.011 | 0.480 | 0.472(0.0871) | -0.005(-1.1\%) | 0.008 | 1.462 |
| $\widetilde{\rho}_{0.05}$ | . 499 | 0.509 | 0.495(0.1200) | -0.004(-0.8\%) | 0.014 | 0.512 | 0.502(0.1065) | 0.003( 0.6\%) | 0.011 | 1.270 |
| $\widetilde{\rho}_{0.2}$ | . 561 | 0.555 | 0.560(0.0912) | -0.001(-0.2\%) | 0.008 | 0.564 | 0.565(0.0792) | 0.004( 0.7\%) | 0.006 | 1.327 |
| $\widetilde{\rho}_{0.4}$ | . 621 | 0.626 | 0.621(0.0807) | 0.001( 0.1\%) | 0.006 | 0.623 | 0.621(0.0742) | 0.000( 0.1\%) | 0.005 | 1.182 |
| $\widetilde{\rho}_{0.6}$ | . 638 | 0.644 | 0.639(0.0773) | 0.001( 0.1\%) | 0.006 | 0.645 | 0.639(0.0717) | 0.001( 0.1\%) | 0.005 | 1.162 |
| $\widetilde{\rho}_{0.8}$ | . 606 | 0.601 | 0.599(0.0808) | -0.007(-1.2\%) | 0.007 | 0.602 | 0.599(0.0718) | -0.007(-1.2\%) | 0.005 | 1.267 |
| $\widetilde{\rho}_{1}$ | . 537 | 0.530 | 0.528(0.1041) | -0.009(-1.8\%) | 0.011 | 0.534 | 0.534(0.0865) | -0.003(-0.5\%) | 0.007 | 1.448 |
| Notes: outputs are from 500 simulations and $n=50$ per dose group. $\rho$ is the correlation used for simulating the bivariate continuous outcomes. <br> *: All $\widetilde{\rho}$ s for separate fitting are obtained from Pearson correlation coefficient between continuous and discrete variable for each dose level. |  |  |  |  |  |  |  |  |  |  |

Table 9.23: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and Cloglog for safety outcome (ELS approach): $\rho=0$

| Parameter | True value | Separate Fitting* |  |  |  | Joint Fitting |  |  |  | Efficiency |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  | $\begin{gathered} \text { Joint } \\ \text { vs. sep. } \\ \hline \end{gathered}$ |
|  |  | Median | Mean (Std) | Bias(\%) | MSE |  | Median | Mean (Std) | Bias(\%) | MSE |
| e0 | 2.5 | 2.434 | 2.402(0.9481) | -0.098(-3.9\%) | 0.904 | 2.352 | 2.376(0.9464) | -0.124(-5.0\%) | 0.907 | 1.004 |
| ed50 | 0.2 | 0.208 | 0.232(0.1169) | 0.032(16.0\%) | 0.015 | 0.205 | 0.230(0.1150) | 0.030(15.2\%) | 0.014 | 1.032 |
| emax | 14.5 | 14.881 | 15.152(1.7845) | 0.652( $4.5 \%)$ | 3.594 | 14.865 | 15.154(1.7719) | 0.654( 4.5\%) | 3.552 | 1.014 |
| $\alpha$ | -2.123 | -2.151 | -2.162(0.2184) | -0.039( $1.8 \%$ ) | 0.049 | -2.142 | -2.166(0.2308) | -0.043( $2.0 \%$ ) | 0.055 | 0.896 |
| $\beta$ | 2.828 | 2.860 | 2.883(0.3305) | 0.055( $2.0 \%$ ) | 0.112 | 2.854 | 2.892(0.3389) | 0.064( $2.2 \%$ ) | 0.118 | 0.951 |
| $\sigma_{Y}$ | 7 | 6.998 | 7.009(0.2561) | 0.009( 0.1\%) | 0.065 | 6.971 | 6.981(0.2498) | -0.019(-0.3\%) | 0.062 | 1.052 |
| $\widetilde{\rho}_{0}$ | 0 | 0.000 | -0.003(0.1343) | -0.003(\%) | 0.018 | -0.001 | -0.007(0.1449) | -0.007(\%) | 0.021 | 0.859 |
| $\widetilde{\rho}_{0.05}$ | 0 | -0.014 | -0.005(0.1445) | -0.005( \%) | 0.021 | -0.015 | -0.008(0.1532) | -0.008( \%) | 0.023 | 0.889 |
| $\widetilde{\rho}_{0.2}$ | 0 | -0.005 | -0.009(0.1462) | -0.009(\%) | 0.021 | -0.012 | -0.010(0.1562) | -0.010( \%) | 0.024 | 0.876 |
| $\widetilde{\rho}_{0.4}$ | 0 | -0.008 | 0.006(0.1533) | 0.006( \%) | 0.023 | -0.013 | 0.002(0.1521) | 0.002( \%) | 0.023 | 1.016 |
| $\widetilde{\rho}_{0.6}$ | 0 | -0.007 | -0.011(0.1458) | -0.011(\%) | 0.021 | -0.012 | -0.011(0.1482) | -0.011(\%) | 0.022 | 0.968 |
| $\widetilde{\rho}_{0.8}$ | 0 | 0.018 | 0.010(0.1586) | 0.010( \%) | 0.025 | 0.007 | 0.011(0.1629) | 0.011( \%) | 0.027 | 0.948 |
| $\widetilde{\rho}_{1}$ | 0 | 0.005 | -0.000(0.1338) | -0.000(\%) | 0.018 | 0.006 | -0.001(0.1350) | -0.001( \%) | 0.018 | 0.983 |
| Notes: outputs are from 500 simulations and $n=50$ per dose group. $\rho$ is the correlation used for simulating the bivariate continuous outcomes. <br> *: All $\widetilde{\rho}$ s for separate fitting are obtained from Pearson correlation coefficient between continuous and discrete variable for each dose level. |  |  |  |  |  |  |  |  |  |  |

Table 9.24: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and Cloglog for safety outcome (ELS approach): $\rho=0.4$

| Parameter | True value | Separate Fitting* |  |  |  | Joint Fitting |  |  |  | Efficiency <br> Joint <br> vs. sep. |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  | Median | Mean (Std) | Bias(\%) | MSE | Median | Mean (Std) | Bias(\%) | MSE |  |
| e0 | 2.5 | 2.584 | 2.562(0.9269) | 0.062( $2.5 \%$ ) | 0.859 | 2.581 | 2.555(0.9016) | 0.055( $2.2 \%$ ) | 0.812 | 1.057 |
| ed50 | 0.2 | 0.205 | 0.237(0.1163) | 0.037(18.6\%) | 0.015 | 0.206 | 0.235(0.1097) | 0.035(17.3\%) | 0.013 | 1.123 |
| emax | 14.5 | 14.738 | 14.966(1.8308) | 0.466( (3.2\%) | 3.552 | 14.927 | 14.952(1.8328) | 0.452(3.1\%) | 3.546 | 0.998 |
| $\alpha$ | -2.123 | -2.142 | -2.152(0.2153) | -0.029(1.4\%) | 0.047 | -2.123 | -2.158(0.2262) | -0.035 ( 1.6\%) | 0.052 | 0.906 |
| $\beta$ | 2.828 | 2.851 | 2.871(0.3115) | 0.043( $1.5 \%$ ) | 0.098 | 2.855 | 2.883(0.3261) | 0.055(1.9\%) | 0.109 | 0.913 |
| $\sigma_{Y}$ | 7 | 7.006 | 7.017(0.2437) | 0.017( $0.2 \%$ ) | 0.059 | 6.974 | 6.986(0.2411) | -0.014(-0.2\%) | 0.058 | 1.022 |
| $\widetilde{\rho}_{0}$ | . 242 | 0.232 | $0.234(0.1324)$ | -0.008(-3.3\%) | 0.017 | 0.241 | $0.234(0.1316)$ | -0.008(-3.4\%) | 0.017 | 1.011 |
| $\widetilde{\rho}_{0.05}$ | . 252 | 0.259 | $0.245(0.1353)$ | -0.006(-2.3\%) | 0.018 | 0.262 | 0.248(0.1392) | -0.003(-1.1\%) | 0.019 | 0.945 |
| $\widetilde{\rho}_{0.2}$ | . 277 | 0.266 | 0.264(0.1342) | -0.012(-4.5\%) | 0.018 | 0.268 | $0.265(0.1314)$ | -0.012(-4.2\%) | 0.017 | 1.044 |
| $\widetilde{\rho}_{0.4}$ | . 305 | 0.310 | 0.309(0.1280) | 0.004( $1.3 \%$ ) | 0.016 | 0.306 | 0.306(0.1217) | 0.001( $0.3 \%)$ | 0.015 | 1.107 |
| $\widetilde{\rho}_{0.6}$ | . 319 | 0.325 | 0.315(0.1353) | -0.004(-1.3\%) | 0.018 | 0.316 | 0.315(0.1308) | -0.005(-1.4\%) | 0.017 | 1.070 |
| $\widetilde{\rho}_{0.8}$ | . 306 | 0.278 | 0.280(0.1349) | -0.026(-8.5\%) | 0.019 | 0.279 | 0.279(0.1311) | -0.027(-8.7\%) | 0.018 | 1.059 |
| $\widetilde{\rho}_{1}$ | . 253 | 0.258 | 0.256(0.1312) | 0.003( $1.2 \%$ ) | 0.017 | 0.259 | 0.256(0.1295) | 0.003( $1.4 \%$ ) | 0.017 | 1.026 |

Table 9.25: Parameter estimates based on separate and joint modeling - Emax for efficacy outcome and Cloglog for safety outcome (ELS approach): $\rho=0.8$

| Parameter | True value | Separate Fitting* |  |  |  | Joint Fitting |  |  |  | Efficiency <br> Joint <br> vs. sep. <br> 1.156 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  | Median | Mean (Std) | Bias(\%) | MSE | Median | Mean (Std) | Bias(\%) | MSE |  |
| e0 | 2.5 | 2.558 | 2.562(0.9297) | 0.062( $2.5 \%$ ) | 0.864 | 2.508 | 2.510(0.8648) | 0.010(0.4\%) | 0.744 | 1.156 |
| ed50 | 0.2 | 0.206 | 0.238(0.1182) | 0.038(19.1\%) | 0.015 | 0.210 | $0.227(0.1001)$ | 0.027(13.5\%) | 0.011 | 1.393 |
| emax | 14.5 | 14.819 | 14.981(1.8597) | 0.481( $3.3 \%$ ) | 3.673 | 14.736 | 14.914(1.7853) | 0.414( $2.9 \%$ ) | 3.343 | 1.085 |
| $\alpha$ | -2.123 | -2.128 | -2.135(0.2157) | -0.012( 0.6\%) | 0.046 | -2.132 | -2.145(0.2268) | -0.022 ( 1.0\%) | 0.052 | 0.905 |
| $\beta$ | 3.728 | 3.845 | 3.859(0.3050) | 0.031( $1.1 \%$ ) | 0.094 | 3.859 | 3.876(0.3199) | 0.048( $1.7 \%$ ) | 0.104 | 0.909 |
| $\sigma_{Y}$ | 7 | 7.010 | 7.017(0.2457) | 0.017(0.2\%) | 0.060 | 6.976 | 6.977(0.2440) | -0.023(-0.3\%) | 0.060 | 1.014 |
| $\widetilde{\rho}_{0}$ | . 484 | 0.474 | 0.476(0.1012) | -0.008(-1.7\%) | 0.010 | 0.485 | 0.480(0.0808) | -0.004(-0.8\%) | 0.007 | 1.567 |
| $\widetilde{\rho}_{0.05}$ | . 502 | 0.510 | $0.498(0.1160)$ | -0.004(-0.7\%) | 0.013 | 0.513 | $0.504(0.1028)$ | 0.002( 0.5\%) | 0.011 | 1.274 |
| $\widetilde{\rho}_{0.2}$ | . 553 | 0.551 | 0.553(0.0968) | -0.000(-0.1\%) | 0.009 | 0.559 | 0.559(0.0841) | 0.005(1.0\%) | 0.007 | 1.325 |
| $\widetilde{\rho}_{0.4}$ | . 610 | 0.620 | 0.615(0.0795) | 0.005(0.8\%) | 0.006 | 0.618 | 0.614(0.0730) | 0.003( 0.6\%) | 0.005 | 1.185 |
| $\widetilde{\rho}_{0.6}$ | . 638 | 0.650 | 0.642(0.0739) | 0.004( (0.6\%) | 0.005 | 0.646 | 0.642(0.0698) | 0.004( 0.6\%) | 0.005 | 1.123 |
| $\widetilde{\rho}_{0.8}$ | . 612 | 0.617 | 0.609(0.0792) | -0.004(-0.6\%) | 0.006 | 0.611 | 0.607(0.0712) | -0.006(-1.0\%) | 0.005 | 1.240 |
| $\widetilde{\rho}_{1}$ | . 506 | 0.505 | 0.495(0.1096) | -0.011(-2.2\%) | 0.012 | 0.502 | 0.497(0.0953) | -0.009(-1.7\%) | 0.009 | 1.324 |
| Notes: outputs are from 500 simulations and $n=50$ per dose group. $\rho$ is the correlation used for simulating the bivariate continuous outcomes. <br> *: All $\widetilde{\rho}_{\text {s }}$ for separate fitting are obtained from Pearson correlation coefficient between continuous and discrete variable for each dose level. |  |  |  |  |  |  |  |  |  |  |

### 9.6.3 The model fitting from the model mis-specification

The model mis-specifications are performed to test the robustness of the QELS method. Here we only consider QELS approach in this section since QELS provides better overall parameter estimates for joint nonlinear modeling of mixed type outcomes in the previous simulation example. The simulations are performed using correct and misspecified model assumption. For example, if the discrete response is simulated from the probit regression, the model for the discrete variable will be fitted with correct probit model, mis-specified logit or Cloglog respectively. The results were compared using Chisquare test to study the impact of model mid-specification.

As illustrated in the following Tables 9.26 to 9.28 , there are 7 dose levels and $d f=(7-1)=6$. As expected chi-square for both correct model and mis-specified models are not significant with P-values $\geq 0.3$. The correct model fitting has the smallest $\chi^{2}$ value and highest p -value and mis-specifed model still have expected value relatively close to the observed value and $p$-values are not significant. These patterns remain true for either strong or weak correlations. The results shown in Tables 9.26 to 9.28 support that the QELS approach is reasonably robust to model mis-specification.

Table 9.26: Model mis-specification fitting when there is no correlation between bivariate responses

|  |  | Chi-square |  |
| :---: | :---: | :---: | :---: |
|  |  | $\sum_{i=1}^{k} \sum_{j=0}^{1} \frac{\left(\mathrm{O}_{i, j}-\mathrm{E}_{i, j}\right)^{2}}{\mathrm{E}_{i, j}}$ |  |
| Data-simulated | Model fitting | Mean(std) | p-value |
| Probit $(\rho=0)$ | Probit | $5.19(3.080)$ | .520 |
|  | Logit | $5.21(3.087)$ | .517 |
|  | Cloglog | $5.86(3.457)$ | .439 |
|  |  |  |  |
| $\operatorname{logit}(\rho=0)$ | Logit | $5.37(3.320)$ | .497 |
|  | Probit | $5.44(3.385)$ | .487 |
|  | Cloglog | $6.27(3.820)$ | .394 |
|  |  |  |  |
| Cloglog $(\rho=0)$ | Cloglog | $5.23(3.328)$ | .514 |
|  | Probit | $6.60(4.403)$ | .359 |
|  | Logit | $6.21(4.011)$ | .400 |

Note: $\rho$ is the correlation used for simulating the bivariate continuous outcomes.

Table 9.27: Model mis-specification fitting when there is weak correlation between bivariate responses

|  |  | Chi-square |  |
| :---: | :---: | :---: | :---: |
|  |  | $\sum_{i=1}^{k} \sum_{j=0}^{1} \frac{\left(\mathrm{O}_{i, j}-\mathrm{E}_{i, j}\right)^{2}}{\mathrm{E}_{i, j}}$ |  |
| Data-simulated | Model fitting | Mean(std) | p-value |
| Probit $(\rho=0.4)$ | Probit | $4.81(2.943)$ | .568 |
|  | Logit | $4.87(3.035)$ | .561 |
|  | Cloglog | $5.36(3.125)$ | .500 |
|  |  |  |  |
| $\operatorname{logit}(\rho=0.4)$ | Logit | $5.14(3.318)$ | .526 |
|  | Probit | $5.18(3.341)$ | .521 |
|  | Cloglog | $6.02(3.568)$ | .421 |
|  |  |  |  |
| $\operatorname{Cloglog}(\rho=0.4)$ | Cloglog | $4.97(2.930)$ | .548 |
|  | Probit | $6.42(4.123)$ | .378 |
|  | Logit | $6.14(4.015)$ | .408 |

Note: $\rho$ is the correlation used for simulating the bivariate continuous outcomes.

Table 9.28: Model mis-specification fitting when there is strong correlation between bivariate responses

|  |  | Chi-square |  |
| :---: | :---: | :---: | :---: |
|  |  | $\sum_{i=1}^{k} \sum_{j=0}^{1} \frac{\left(\mathrm{O}_{i, j}-\mathrm{E}_{i, j}\right)^{2}}{\mathrm{E}_{i, j}}$ |  |
| Data-simulated | Model fitting | Mean(std) | p-value |
| Probit $(\rho=0.8)$ | Probit | $5.23(3.263)$ | .515 |
|  | Logit | $5.28(3.219)$ | .510 |
|  | Cloglog | $5.78(3.017)$ | .448 |
|  |  |  |  |
| $\operatorname{logit}(\rho=0.8)$ | Logit | $5.20(3.17)$ | .518 |
|  | Probit | $5.29(3.40)$ | .507 |
|  | Cloglog | $6.11(3.45)$ | .411 |
|  |  |  |  |
| Cloglog $(\rho=0.8)$ | Cloglog | $5.48(3.214)$ | .484 |
|  | Probit | $7.23(6.919)$ | .300 |
|  | Logit | $6.6(4.152)$ |  |
| Note: $\rho$ is the correlation used for simulating the bivariate continuous outcomes. |  |  |  |

### 9.7 How to determine $M E D$ and $M S D$

### 9.7.1 $M E D$ definition

MED definition will be the same as the definition in Part I Section 2.2.1.

### 9.7.2 MSD definition

Following the rationale of MSD definition for continuous outcome in Section 2.2.1, for discrete outcome, the model-based approaches allow $M S D \in\left(d_{1}, d_{k}\right]$. Given a model $g(., \boldsymbol{\theta})$,

$$
M S D=\underset{d \in\left(d_{1}, d_{k}\right]}{\operatorname{argmax}}\left\{g(d, \theta) \leq g\left(d_{1}, \boldsymbol{\theta}\right)+\triangle\right\},
$$

$g(., \boldsymbol{\theta})$ can be the inverse of logit, probit or complementary log-log link if the discrete response follows logistic, probit or complementary log-log regression respectively. Here, MSD is defined as the maximum dose which shows clinically acceptable toxicity in terms of measurable probability of safety events. $\triangle$ will be the clinical acceptable difference of probability for the interest of safety outcome, that is, the largest relevant difference of probability, by which we expect a dose to be not too worse than placebo.

Following the above definition, two different rules are proposed to estimate true MSD. Denote $U_{d}$ the upper $1-2 \gamma$ confidence limit of predicted mean value $p_{d}$ at dose $d$ based on the model $g(., \boldsymbol{\theta})$.

$$
\begin{aligned}
& \widehat{M S D_{1}}=\operatorname{argmax}_{d \in\left(d_{1}, d_{k}\right]}\left\{U_{d} \leq g\left(d_{1}, \widehat{\boldsymbol{\theta}}\right)+\Delta\right\}, \\
& \widehat{M S D_{2}}=\operatorname{argmax}_{d \in\left(d_{1}, d_{k}\right]}\left\{p_{d} \leq g\left(d_{1}, \widehat{\boldsymbol{\theta}}\right)+\Delta\right\}
\end{aligned}
$$

### 9.8 Procedures/Strategies of finding $M E D$ and $M S D$ for bivariate continuous efficacy and discrete safety data

The procedures and strategies are similar to what we described in the Section 4.2.5. As shown in Figure 4.1, the first step is to confirm if there exists PoC for efficacy (alpha=0.05), there will be no dose-finding continuation for this drug which means the drug will not be carried to the Phase III development. When the PoC for the efficacy response is established, the second step is to show the PoC for the safety response (alpha=0.2). Only efficacy response is studied further to identify MED for the Phase III program when there is no established PoC for safety response. Joint modeling for efficacy and safety responses is performed under the scenario there are both PoC for efficacy and safety responses.

The details are referred in Section 4.2.5 and are not repeated here.

### 9.9 Design for separate and joint model fitting

### 9.9.1 Separate model fitting

To get the MED and MSD from separate model fitting by ignoring the correlation:

1. The following set of candidate efficacy models such as Linlog, Emax, exponential, quadratic are chosen for the continuous response; the logistic, probit, complementary $\log -\log$ models are selected for fitting discrete safety data. Nonlinear the regression model may be used for discrete safety data.
2. Fit the efficacy data separately and choose best model with alpha $=0.05$ and the lowest AIC; If PoC for efficacy is established, fit discrete safety data separately and choose best model with alpha= 0.2 and the lowest AIC.
3. Obtain the $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ based on the models from step 2.

### 9.9.2 Joint model fitting

When there is no PoC for efficacy established as shown in Figure 4.4, there are no continued model fittings and Phase III development will not be proceeded. On the other hand when there is PoC for efficacy but with no PoC for safety, then we only focus on determining the $M E D$. In the following Sections we will describe the strategy of estimating the MED and MSD from the joint model fittings after PoC for both efficacy and safety are established. Two stratigies will be used to find the final joint model as follow:

## Strategy I - Keep most significant model from separate efficacy and safety

 model fitting1. First get all the significant efficacy models, of which $t$-statistics are greater than critical value $q^{\Upsilon}$ and all the significant safety models which z-statistics are greater than critical value $q^{Z}$ for controlling FWER from separate model fitting. Alpha $=0.05$ is pre-specified for PoC for efficacy and alpha $=0.2$ is pre-specified for PoC for safety respectively. The PoC for the safety discrete response implies that there is significant differences in probabilities of safety events among doses. The wald statistic under null hypothesis is asymptotic multivaraite normally distributed. The detailed proofs are derived in Appendix Section 13.2. The computation of the critical value $q^{Z}$ should account for the multiplicity to control the FWER at a pre-specified alpha level (Hochberg and Tamhane, 1987). After the PoC for both the efficacy and response are confirmed, the next step is to determine the most significant efficacy and safety models based on the lowest AIC criteria, and keep the parameter estimates from the separate model fitting, these values will be used as starting values for joint model fitting.
2. Joint model fitting by QELS approach. This is done by specifying the correct mean model and variance-covariance structure for the bivariate outcomes.
3. Obtain $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ based on updated efficacy and safety models from joint models fitting. Base on the definition of $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$, fit the updated marginal of joint model to obtain the $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$.
4. Suggest an optimal dose or a range of doses for Phase III, and evaluate the Phase III

## Program Development.

## Strategy II - Keep all significant models from separate efficacy and safety mod-

 els1. Keep all the significant efficacy models, of which $t$-statistics are bigger than critical value $q^{Y}$, and all the significant safety models, of which z-statistics are greater than critical value $q^{Z}$, for controlling FWER from separate fitting. Alpha $=0.05$ is prespecified for PoC for efficacy and alpha $=0.2$ is pre-specified for PoC for safety respectively. The parameter estimates are kept and will be used as starting values for joint model fitting.
2. Joint model fitting by QELS approach for all the combinations of efficacy and safety models selected from separate model fitting in step 1, the next step is to determine the best combination based on lowest AIC from all the joint models fitting. The method for computing AIC is based on observed likelihood described in Section 9.2.
3. Obtain the $\widehat{M E D}_{2}$ and $\widehat{M S D}_{1}$ based on the estimated marginal efficacy and safety models of the best fitted joint model.
4. Suggest an optimal dose or a range of doses for Phase III, and evaluate the Phase III Program Development..

## Chapter 10

## Suggest Dose(s) for the Phase III

## Program Development Based Joint

## Continuous and Discrete Responses

In this chapter, we will propose several methods to select an optimal dose or a range of doses for Phase III and evaluate the possibility of success in Phase III program. Two of the proposed methods are based on probability functions, another two are based on utility function. If $\widehat{M S D}_{1}$ is smaller than $\widehat{M E D}_{2}$, there will be no appropriate dose for the Phase III program.

### 10.1 Identify dose(s) for Phase III program through joint criteria of continuous efficacy and discrete safety responses

There are two proposed methods described in this section.

### 10.1.1 Methods

Method 1: The recommended dose(s) will be determined by

$$
\underset{d \in[M E D, M S D]}{\operatorname{argmax}}\{P(Y>a, Z=0 \mid d) \geq \xi\}, \text { or }
$$

a dose range in $[M E D, M S D]$ such that $P(Y>a, Z=0 \mid d) \geq \xi$,
where Y is continuous efficacy variable and Z is discrete safety variable, $\mathrm{Z}=0$ means no toxicity while $a$ is the criteria for Phase III success and $\xi$ is the success probability. The parameters for the joint bivariate density are estimated from the joint fitted model using the QELS approach in Sections 9.3.1 and 9.3.2. These estimates will be used to get the estimated joint bivariate density and hence calculate the probabilities we are interested in.

When the latent variable threshold $c$ and standard deviation $\sigma_{S}^{2}$ of latent variable are known, the joint density is,

$$
\begin{aligned}
f(y, z \mid d)= & f_{Y}(y \mid d) f_{Z \mid Y}(z \mid y, d) \\
= & \frac{\exp \left[-\frac{\left(y-u_{i}^{Y}\right.}{2 \sigma_{Y}^{2}}\right]}{\sqrt{2 \pi \sigma_{Y}^{2}}}\left(\Phi\left(\frac{c-\left(u_{i}^{Z}+\rho \frac{\sigma_{S}}{\sigma_{Y}}\left(y-u_{i}^{Y}\right)\right)}{\sigma_{S} \sqrt{\left(1-\rho^{2}\right)}}\right)\right)^{1-z} \\
& \times\left(1-\Phi\left(\frac{c-\left(u_{i}^{Z}+\rho \frac{\sigma_{S}}{\sigma_{Y}}\left(y-u_{i}^{Y}\right)\right)}{\sigma_{S} \sqrt{\left(1-\rho^{2}\right)}}\right)\right)^{z}
\end{aligned}
$$

(See Section 9.2.1 for details).

If we do not have the information for the latent variable and assume $\sigma_{Y \mid(Z=1)}^{2}=$ $\sigma_{Y \mid(Z=0)}^{2}=\sigma_{c}^{2}$, then the following joint density will be used,

$$
\begin{aligned}
f(y, z \mid d)= & f_{Z}(z \mid d) f_{Y \mid Z}(y \mid z, d) \\
= & \left(\frac{1}{\sqrt{2 \pi \sigma_{c}^{2}}}\left(u_{i}^{Z}\right)^{Z}\left(1-u_{i}^{Z}\right)^{1-Z}\right) \\
& \times \exp \left(\frac{-\left(y-u_{i}^{Y}-\widetilde{\rho} \sqrt{\frac{1}{u_{i}^{Z}\left(1-u_{i}^{Z}\right)\left(1-\tilde{\rho}^{2}\right)}}\left(z-u_{i}^{Z}\right)\right)}{2 \sigma_{c}}\right) .
\end{aligned}
$$

Method 2: Instead of using the joint density, the marginal distribution approach can also be used, i.e., the recommended dose(s) will be,

$$
\underset{d \in[M E D, M S D]}{\operatorname{argmax}}\left\{P(Y>a \mid d) \geq \xi_{1} \text { and } P(Z=1 \mid d)<\xi_{2}\right\}, \text { or }
$$

a dose range in $[M E D, M S D]$ such that $\left\{P(Y>a \mid d) \geq \xi_{1}\right.$ and $\left.P(Z=1 \mid d)<\xi_{2}\right\}$,
where $Y$ is the continuous efficacy variable and $Z$ is the discrete safety variable, $Z=1$ means toxicity response, $a$ is the criteria for Phase III success, $\xi_{1}$ and $\xi_{2}$ are the success probabilities. The probability $P(Y>a \mid d)$ and $P(Z=1 \mid d)$ are calculated based on the estimated marginal density functions derived from either separate model fitting or joint model fitting using the QELS approach in Sections 9.3.1 and 9.3.2.

### 10.2 Identify dose(s) for Phase III through utility function based on trade off of efficacy and safety responses

### 10.2.1 Methods

The utility function method to determine the recommended dose is based on the following utility function,

$$
F(d)=e f f(d)-k * \operatorname{saf}(d)
$$

The dose is determined by maximizing the utility function $\mathrm{F}(\mathrm{d})$, i.e., $\underset{d \in[M E D, M S D]}{\operatorname{argmax}} F(d)$.
Here, $k$ is the weight for the discounted safety from the efficacy; $\operatorname{eff}(d)$ and $\operatorname{saf}(d)$ can represent

1) $P(Y>a \mid d)$ and $P(Z=1 \mid d)$ respectively, where the estimated marginal distributions are derived from either separate model fitting or joint model fitting using the QELS approach in Sections 9.3.1 and 9.3.2. $Y$ is the continuous efficacy variable and $Z$ is the discrete safety variable, $a$ is a criteria for Phase III success.
2) Standardized response $e f f(d)=\widehat{Y}(d)$ and $\operatorname{saf}(d)=\widehat{Z}(d)$, where $\widehat{Y}(d)=\frac{E(Y(d))}{\sqrt{\operatorname{Var}(Y(d))}}=$ $\frac{f\left(d, \theta_{Y}\right)}{\sigma_{Y}}$ and $\widehat{Z}(d)=E(Z(d))=g\left(d, \theta_{Z}\right) \cdot \theta_{Y}, \theta_{Z}$, and $\sigma_{Y}$ are estimated from the mean efficacy and safety models based on either the separate marginal model fitting or the joint model fitting using the QELS approach in Sections 9.3.1 and 9.3.2.

## Chapter 11

## Conclusions

In Part II of this dissertation, we developed the methodology for identifying the final dose for Phase III program based on bivariate nonlinear models for continuous and discrete endpoints. The methodology includes several procedures and is summarized as follows:

1. The strategy to simulate the bivariate continuous and discrete outcomes: The approach utilizes the association between the discrete variable and latent continuous variable in terms of the mean model and the correlations at each dose level.

The derivations are in Section 9.2.1. The parameter values from the simulation results are very close to the true parameter values from which the data are simulated. The results show that the methods are valid for the different link functions selected for the discrete variables and for various strength of correlations between the bivariate variables.
2. Estimation for joint continuous and discrete models: The ELS and QELS methods are both explored and applied to the mixed type data. In this dissertation, dif-
ferent link functions for the discrete variable and various bivariate correlations, from no bivariate correlations to strong bivariate correlations, are considered. As expected, both ELS and QELS methods produce relatively consistent estimates for all the parameters. The standard errors from the QELS method are relatively smaller than the standard errors from the ELS method. This confirmed that using the QELS estimating equations, we can still achieve a consistent estimate of mean model parameter estimates even if the variance-covariance structure is mis-specified. The robust standard errors of the parameter estimates for the continuous variable obtained from the ELS method and QELS methods are similar. And the robust errors for the discrete variable parameter estimates obtained from the QELS are little lower than the robust errors obtained from the ELS method. Overall the QELS method is more efficient and robust for joint model fitting of the continuous and discrete responses .

## 3. Comparison of efficiency of the parameter estimates from joint model fitting

and separate model fitting: The data with different correlations between the bivariate variables and using different link functions for the discrete variable are simulated. The relative efficiency of the parameter estimates for the continuous outcome from the joint model fitting with respect to the separate model fitting increases as the correlation between the bivariate increases. Efficiencies of the joint model fitting with respect to the separate model fitting are the similar for the parameter estimates for the discrete outcome model. This is not surprising since there are tremendous information loss when the observed variable is dichotomized obtained from the latent continuous outcome.
4. Joint model fitting evaluation: The full likelihood is derived in Section 9.2.

There are two approaches to obtain the likelihood depending on if the information of the latent continuous outcome is known. The AIC can be computed from the observed likelihood after we have the parameter estimates from the ELS or QELS methods.

In addition, the model mis-specification is used to test the robustness of the QELS approach. The Chi-square results from the model mis-specification do show that QELS approach is robust.
5. Define the MSD for the discrete safety variable, and develop the strategies of finding MED and MSD for bivariate continuous efficacy and discrete safety data: The strategies of finding MED and MSD for the joint mixed type efficacy and safety responses are similar to the bivariate continuous responses in Part I. The overall strategy is to prove the PoC for the safety response (alpha=0.2) only after the PoC for the efficacy response (alpha=0.05) is established. The two approaches developed in Part I are still valid here for the mixed type responses.
6. Identify an optimal dose or a range of doses for Phase III program: There are two concepts developed for identifying an optimal dose or a range of doses for the Phase III program. First approach is through the joint success criteria of the continuous efficacy and discrete safety responses for the Phase III program. The parameter estimates from the joint modeling are obtained from the QELS method. The joint success criteria can be based on either the joint distribution or the marginal distributions obtained from the joint model estimation. The second approach is the utility function based on the trade off of the continuous efficacy and discrete safety success criteria. The rationales are similar to the bivariate continuous responses in Part I.

## Chapter 12

## Discussion and Future Directions

### 12.1 Discussion

Part I develops the dose-finding methodology for continuous bivariate responses while Part II develops the dose finding methodology for mixed typed outcomes. There are some key differences for these two methodologies which are discussed below.

The motivation for Part I of the thesis is based on an example of the joint continuous bivariate efficacy and safety data from a clinical trial. The estimation of the joint nonlinear continuous model uses the full likelihood approach and there is normality assumption for the bivariate normal outcomes. The simulation results do show the joint modeling provide less biased and more efficient parameter estimates when the bivariate data have stronger correlation. The whole developed methodology starts from the extension of MSD, joint modeling, determination of the best final joint model and identifying the final optimal dose for maximizing success of the Phase III program. We illustrate an example through simulations of bivariate normal outcomes and show that the methodology
is valid and applicable.
After Part I is finished, we continue to explore different methodologies when the bivariate data are the mixed type of continuous and discrete. There are some challenges to use the full likelihood approach for joint model estimations not only the nonlinear modeling of mixed type outcomes, but also the modeling of second moment parameters since the correlations between the bivariate variables are dose dependent. In Part II of the dissertation we used the extended least squares approach for the estimation of the nonlinear joint model of mixed type outcomes. The simulation results show that the parameter estimates obtained from the joint modeling approach are more efficient than the parameter estimates obtained from the separate model especially for the estimates of model for continuous outcome. For the discrete variable, as expected there are some loss of information when the discrete variable is dichotomized from the latent continuous variable. It seems that efficiency remains similar comparing the joint modeling with the separate modeling for discrete portion of outcomes. The standard errors from the QELS method are relatively smaller than the standard errors from the ELS method. The rationale is that using the QELS estimating equations can still achieve a consistent estimate of mean model parameter estimates even if the variance-covariance structure is mis-specified. Furthermore the QELS approach is a robust approach even when the model is mis-specified.

It was shown by Vonesh et al. (2001) that the ELS and QELS approach are equivalent to the maximum likelihood estimation assuming normality. Furthermore the advantage of using the ELS approach is that the 3rd and 4th order moments are automatically generated as though from a multivariate normal distribution even though data may not
follow a multivariate normal. This does add the robustness of this approach since normality assumption is not a necessarily condition for applying this method to the mixed type data. Although the comparisons of the different estimation methods are not the focus of the Part II, it may be a future direction to explore how the ELS or QELS approach is different from the full likelihood approach when normality condition is not met in terms of the efficiency of the parameter estimates.

Other challenges we face when we develop the dose finding methodology for the mixed type bivariate data are the criteria for model selections when the data are not the bivariate normal, we compute the AIC based on the observed likelihood. We discuss the two approaches to compute the observed likelihood which depends on the information we have for the latent continuous variable. But as we know the ELS or QELS do not use the full likelihood, so in the future we may explore the alternatives for the model fitting criteria like QIC (Pan, 2001), instead of using the AIC.

In addition, how to simulate the bivariate mixed type data mimic the clinical setting for our research purpose is another challenge we face in Part II of the dissertation. We do derive the association between the mean model for the discrete variable and the mean model for the latent continuous variable, the relationship of the correlation between the bivariate mixed type data and the correlation between the bivariate continuous variable at each dose level. These relationships can be utilized for either observed likelihood computation or the joint density criteria for the dose optimization.

The two strategies for determining the final model and how to identify the final optimal dose(s) are similar in Part I and Part II. One point need to be addressed here is
that since Part II apply to the discrete data, the success criteria will be different from what is defined in Part I. In Part I the success criteria for the safety outcome criteria is that the probability of the continuous safety response less than the certain value, for the discrete data the success criteria is the probability of no toxicity for the safety outcome for the Phase III program.

Though we mentioned in Part II that bivariate mixed type data consists of the joint continuous efficacy and discrete safety outcomes, this is not always the case. In the clinical trials, the mixed type data can be discrete efficacy data and continuous safety data or the mixed type data can both be generated from the efficacy or safety outcomes depending on the individual clinical setting.

As we can see the dose finding methodology in Part II is developed based on nonlinear joint models for the mixed type bivariate data, but the estimation procedure can also be applied to bivariate continuous data. Certainly there may have more advantages to use the full likelihood estimation methodology in Part I when the data are bivariate normal. In cases that the joint continuous data are not bivariate normal we may expect the ELS or QELS method may provide better parameter estimates than the estimations in Part I.

### 12.2 Future Directions

Assessment of dose-response profiles for efficacy and safety outcomes are the key for reliable evaluations of the risk-benefit profile of a drug as well as the selection of final doses to be carried into Phase III program. Due to broad applications of the dose-finding
approach and various clinical trial designs, there are lots of interesting issues remain unresolved. The following directions are considered as important potential directions in the future.

### 12.2.1 Joint nonlinear continuous and discrete model estimation using full like- <br> lihood approach

In Part II, we use the extended least squares method to estimate the joint nonlinear continuous and discrete model. We do have the relative consistent and efficient parameter estimates using this approach. Although for Part II of the dissertation, the focus is not comparing the different estimating approaches, in the future, utilizing the full likelihood may be another direction for estimations. Fitmaurice and Laird (1995) developed the estimating method using likelihood equations. But the author did realize that the full likelihood-based approach is computationally intensive and the solution to likelihood equations becomes intractable when the number of nuisance parameters increased. Furthermore this paper only consider the linear model for continuous outcome. In addition, for our dose finding approach the correlations are different at each dose level.

There are other two key aspects make full likelihood not robust. If the distribution we use in the likelihood do not reflect the true distribution, i.e., the normality is not met, then the likelihood approach is not accurate. On the other hand we test robustness of the model mis-specification for the ELS or QELS. ELS and QELS can be applied to mixed type responses without the need to derive the likelihood. They can still get reasonable estimates even the normality is not met.

Thus the full likelihood approach can be a future direction but may not be attrac-
tive approach due to the above issues.

### 12.2.2 Extension of dose finding based on bivariate discrete variable

In Part I of this dissertation we developed the methodology to estimate the joint bivariate continuous outcomes based on full likelihood, how to find the MED and MSD and how to identify the final dose(s) to carry into Phase III. In Part II of this dissertation we also developed the methodology to estimate the joint bivariate mixed type outcomes based on extended least squares (ELS) or quasi-extended least squares method (QELS), and furthermore we discussed the approaches to find the MSD for discrete variable and identify the final dose for Phase III. Having done these work, we do realize that in reality it is also common that in clinical trials dose finding need to be based on the outcomes which are both discrete variables. Under this clinical setting, the extension of dose finding based on the bivariate discrete variable is an interesting topic in the future. It is worth to mention that ELS or QELS approach adapted into the estimation of bivariate mixed type outcomes in Part II can also be applied to the estimation of the joint nonlinear models of the bivariate discrete variables.

### 12.2.3 Comparison of estimating methods in Part I and Part II

In part I , the estimating method for the joint continuous bivariate variable is based on the full likelihood under the normality assumption. This approach only apply to bivariate continuous variable under normality. While the estimating method in Part II is based on the extended least squares and normality assumption is not required. The estimating method in Part II can be applied to different types of responses, there is no
restriction of the type of random variable under normality, it may still produce correct estimates under the non-normal situation. Thus we consider the approach in Part II is more robust and has broader application. While Part I does have the advantage since it may produce more efficient estimated by the full likelihood approach under normality, it may not produce correct estimates under the non-normal situation. Thus the comparison of estimations of bivariate continuous variables in Part I and Part II may be a potential topic. We do expect under normality the method in Part I may generate more efficient parameter estimates while the method in Part II may have some advantages under non-normal assumptions.

### 12.2.4 Quasi-likelihood criterion (QIC) for model selections

Model selection is an important mid-step in our procedures for finding the best joint model. In Part II we propose to use the observed likelihood to compute AIC. AIC is based on the likelihood and asymptotic properties of the maximum likelihood estimator. It is described in Vonesh et al. (2001) that minimizing the QELS objective function is equivalent to maximizing the joint log-likelihood function under normality assumption. Though Vonesh et al. (2001) did show that the assumption of normality is not required to achieve the consistent and reasonably efficient estimators, Pan (2001) proposed a modification to the penalty term in the usual AIC and used the quasi-likelihood for the mean and second moment parameters. Since QELS is the non-likelihood based, though we do derive the likelihood and use the observed likelihood for the AIC computation, Qualsi-likelihood criterion (QIC) may be an alternative to our observed likelihood based AIC method.

## Chapter 13

## Appendix

### 13.1 Comparison of Likelihoods Between Method 1 and Method

2

Sections 9.2.1 and 9.2.2 introduce two approaches to derive the likelihood depending on whether we have the information about the latent variable. Method 1 assumes that we do have the underlying distribution about the latent variable. In this case we assume the bivariate continuous variable follows a bivariate normal distribution and hence $Y_{i j} \mid S_{i j}$ follows a normal distribution. Method 2 derives the likelihood directly from the bivariate continuous and discrete variable when we do not have information about the underlying variable. In this case, we assume the conditional distribution $f_{Y_{i j} \mid Z_{i j}}\left(y_{i j} \mid z_{i j}\right)$ is a normal distribution. In this section, we want to explore if $Y_{i j} \mid\left(S_{i j}>c\right)$ is normally distributed given that $\left(Y_{i j}, S_{i j}\right)$ follows a bivariate normal distribution in Method 1.

$$
\begin{aligned}
& f_{Y_{i j} \mid Z_{i j}}(y \mid z=1)= f_{Y_{i j} \mid\left(s_{i j}>c\right)}(y)=\frac{\int_{c}^{\infty} f\left(Y_{i j,}, S_{i j}\right)}{P(y, s) d s} \\
& P\left(S_{i j}>c\right) \\
&= \int_{c}^{\infty} \frac{1}{2 \pi\left|\Sigma_{i}\right|^{1 / 2}} \exp \left(-\frac{1}{2}\left(\left[\begin{array}{c}
y-u_{i}^{Y} \\
s-u_{i}^{S}
\end{array}\right] \Sigma_{i}^{-1}\left[\begin{array}{c}
y-u_{i}^{Y} \\
s-u_{i}^{S}
\end{array}\right]\right)\right) d s \\
& P\left(S_{i j}>c\right) \\
&= \frac{\int_{c}^{\infty} \frac{1}{2 \pi\left|\Sigma_{i}\right|^{1 / 2}} \exp \left(-\frac{1}{2}\left(\frac{\left(y-u_{i}^{Y}\right)^{2}}{\sigma_{Y}^{2}}+2 \rho_{i} \frac{\left(y-u_{i}^{Y}\right)\left(s-u_{i}^{S}\right)}{\sigma_{Y} \sigma_{S}}+\frac{\left(s-u_{i}^{S}\right)^{2}}{\sigma_{S}^{2}}\right)\right) d s}{P\left(S_{i j}>c\right)} \\
&= k_{1} \exp \left(-\frac{\left(y-u_{i}^{Y}\right)^{2}}{2 \sigma_{Y}^{2}}\right) \\
& \times \int_{c}^{\infty} \exp \left(-\frac{1}{2 \sigma_{S}^{2}}\left(2 \rho_{i} \frac{\sigma_{S}}{\sigma_{Y}}\left(y-u_{i}^{Y}\right)\left(s-u_{i}^{S}\right)+\left(s-u_{i}^{S}\right)^{2}\right)\right) d s \\
&= k_{1} \exp \left(-\frac{\left(y-u_{i}^{Y}\right)^{2}}{2 \sigma_{Y}^{2}}+\frac{1}{2 \sigma_{S}^{2}}\left(\rho_{i} \frac{\sigma_{S}}{\sigma_{Y}}\left(y-u_{i}^{Y}\right)\right)^{2}\right) \\
& \times \int_{c}^{\infty} \exp \left(-\frac{\left(\left(\frac{\rho_{i} \sigma_{S}\left(y-u_{i}^{Y}\right)}{\sigma_{Y}}\right)^{2}+\frac{2 \rho_{i} \sigma_{S}\left(y-u_{i}^{Y}\right)\left(s-u_{i}^{S}\right)}{\sigma_{Y}}+\left(s-u_{i}^{S}\right)^{2}\right)}{2 \sigma_{S}^{2}}\right) d s \\
& \times \sum_{c} \exp \left(-\frac{\left(y-u_{i}^{Y}\right)^{2}}{2 \sigma_{Y}^{2}}+\frac{\rho_{i}^{2}}{2 \sigma_{Y}^{2}}\left(y-u_{i}^{Y}\right)^{2}\right) \\
& \exp \left(-\frac{1}{2 \sigma_{S}^{2}}\left(s-u_{i}^{S}+\rho_{i} \frac{\sigma_{S}}{\sigma_{Y}}\left(y-u_{i}^{Y}\right)\right)^{2}\right) d s
\end{aligned}
$$

let $z=\frac{s-u_{i}^{S}+\rho_{i} \frac{\sigma_{S}}{\sigma_{\gamma}}\left(y-u_{i}^{Y}\right)}{\sigma_{S}}$, then

$$
\begin{align*}
f_{Y_{i j} \mid Z_{i j}}(y \mid z= & 1)=k_{1} \exp \left(-\frac{\left(1-\rho_{i}^{2}\right)\left(y-u_{i}^{Y}\right)^{2}}{2 \sigma_{Y}^{2}}\right) \sqrt{2 \pi} \sigma_{S} \\
& \times \int_{\frac{c-u_{i}^{S}+\rho_{i} \frac{\sigma_{S}}{\sigma_{Y}}\left(y-u_{i}^{Y}\right)}{\sigma_{S}}}^{\infty} \frac{1}{\sqrt{2 \pi}} \exp \left(-\frac{z^{2}}{2}\right) d z \\
= & k_{1} \exp \left(-\frac{\left(1-\rho_{i}^{2}\right)\left(y-u_{i}^{Y}\right)^{2}}{2 \sigma_{Y}^{2}}\right) \\
& \times \sqrt{2 \pi} \sigma_{S}\left[1-\Phi\left(\frac{c-u_{i}^{S}+\rho_{i} \frac{\sigma_{S}}{\sigma_{Y}}\left(y-u_{i}^{Y}\right)}{\sigma_{S}}\right)\right] \tag{13.1}
\end{align*}
$$

where $k_{1}=\frac{1}{2 \pi\left|\Sigma_{i}\right|^{1 / 2} P\left(S_{i j}>c\right)}$. Apparently from formula (13.1), the $f_{Y_{i j} \mid Z_{i j}}(y \mid z=1)$ is not a pdf of a normal distribution. Thus the distribution from Method 1 is different from the distribution from Method 2. But for both likelihoods, they have the same first two moment parameters, and we do have all the parameter estimates using ELS or QELS methods.

### 13.2 Proof of Concept Study for Binary Outcomes

### 13.2.1 Contrast Tests with Multiple Dose Response Models

For the purpose of detecting an overall trend, we assume the following model:

$$
u_{i}^{Z}=p_{i}^{Z}=P\left(Z_{i j}=1 \mid d_{i}\right)=g\left(d_{i}, \boldsymbol{\theta}_{Z}\right) \text { for } i=1, \ldots, k, j=1, \ldots n_{i},
$$

where $p_{i}^{Z}=g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)$ are the probability of $\left(Z_{i j}=1\right)$ at dose $d_{i}$. Let $\mathbf{p}^{Z}=\left(p_{1}^{Z}, \ldots, p_{k}^{Z}\right)$. The MLEs of $p_{i}^{Z}$ and $\mathbf{p}^{Z}$ are

$$
\hat{p}_{i}^{Z}=\frac{\sum_{j=1}^{n_{i}} Z_{i j}}{n_{i}}
$$

and $\widehat{\mathbf{p}}^{Z}=\left(\hat{p}_{1}^{Z}, \ldots, \hat{p}_{k}^{Z}\right)$, respectively.
Assume that we are given a set of models $\mathcal{M}=\left\{M_{m}: m=1, \ldots, M\right\}$. For each model $M_{m}$, we want to determine if there are significant differences in probabilities of events among all doses by considering the following hypotheses:

$$
\begin{gathered}
H_{0}^{m}: \sum_{i=1}^{k} c_{m i} p_{i}^{Z}=\mathbf{c}_{m}^{\prime} \mathbf{p}^{Z}=0, \\
\text { vs. } H_{1}^{m}: \sum_{i=1}^{k} c_{m i} p_{i}^{Z}=\mathbf{c}_{m}^{\prime} \mathbf{p}^{Z}>0,
\end{gathered}
$$

where contrast vectors $\mathbf{c}_{m}^{\prime}=\left(c_{m 1}, \ldots, c_{m k}\right)$ are known constants subject to $\mathbf{c}_{m}^{\prime} \mathbf{1}=0, m=$ $1, \ldots, M$.

The wald test statistic for the hypotheses can be written as

$$
Z_{m}=\frac{\mathbf{c}_{m}^{\prime} \widehat{\mathbf{p}}^{Z}}{\sqrt{\mathbf{c}_{m}^{\prime} I_{N}^{-1}\left(\widehat{\mathbf{p}}^{Z}\right) \mathbf{c}_{m}}}, m=1, \ldots, M
$$

where $I_{N}\left(\mathbf{p}^{Z}\right)$ is the Fisher information matrix for $\mathbf{p}$. Since

$$
\begin{gathered}
I_{N}(\mathbf{p})=\left[\begin{array}{cccc}
\frac{n_{1}}{p_{1}^{Z}\left(1-p_{1}^{Z}\right)} & 0 & \cdots & 0 \\
0 & \frac{n_{2}}{p_{2}^{Z\left(1-p_{2}^{Z}\right)}} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \frac{n_{k}}{p_{k}^{Z}\left(1-p_{k}^{Z}\right)}
\end{array}\right], \\
\mathbf{c}_{m}^{\prime} I_{N}^{-1}\left(\widehat{\mathbf{p}}^{Z}\right) \mathbf{c}_{m}=\sum_{i=1}^{k} \hat{p}_{i}^{Z}\left(1-\widehat{p}_{i}^{Z}\right) c_{m i}^{2} / n_{i} \text { and } \\
Z_{m}=\frac{\sum_{i=1}^{k} c_{m i} \widehat{p}_{i}^{Z}}{\sqrt{\sum_{i=1}^{k} \hat{p}_{i}^{Z}\left(1-\widehat{p}_{i}^{Z}\right) c_{m i}^{2} / n_{i}}}, m=1, \ldots, M .
\end{gathered}
$$

It is shown that $Z_{m}$ is asymptotically normal with mean 0 and variance 1 under null hypotheses. In the proof of concept study, we want to test $H_{0}: \mathbf{c}_{m}^{\prime} \mathbf{p}^{Z}=0$ for all $m=1, \cdots, M$
and control the family wise error rate (FWER). Since $Z_{1}, \ldots, Z_{M}$ are correlated, we need to study the asymptotic joint distribution of $\mathbf{Z}=\left(Z_{1}, \ldots, Z_{M}\right)^{\prime}$ under $H_{0}$.

The Central Limit Theorem shows that $\sqrt{n_{i}}\left(\widehat{p}_{i}^{Z}-p_{i}^{Z}\right) \xrightarrow{d} N\left(0, p_{i}^{Z}\left(1-p_{i}^{Z}\right)\right)$ as $n_{i} \rightarrow \infty$ and

$$
\sqrt{N}\left(\widehat{\mathbf{p}}^{Z}-\mathbf{p}^{Z}\right) \xrightarrow{d} N\left(\mathbf{0}, V\left(\mathbf{p}^{Z}\right)\right) \text { as } N \rightarrow \infty,
$$

where $N=\sum_{i=1}^{k} n_{i}$,

$$
V\left(\mathbf{p}^{\mathbf{Z}}\right)=\left[\begin{array}{cccc}
r_{1} p_{1}^{Z}\left(1-p_{1}^{Z}\right) & 0 & \cdots & 0 \\
0 & r_{2} p_{2}^{Z}\left(1-p_{2}^{Z}\right) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & r_{k} p_{k}^{Z}\left(1-p_{k}^{Z}\right)
\end{array}\right]
$$

and $r_{i}=N / n_{i}$. Let $C=\left(\mathbf{c}_{1}, \cdots, \mathbf{c}_{M}\right)$, we have

$$
\sqrt{N}\left(C^{\prime} \hat{\mathbf{p}}^{Z}-C^{\prime} \mathbf{p}^{Z}\right) \xrightarrow{d} N\left(\mathbf{0}, C^{\prime} V\left(\mathbf{p}^{Z}\right) C\right) \text { as } N \rightarrow \infty,
$$

where

$$
\begin{aligned}
& C^{\prime} V\left(\mathbf{p}^{Z}\right) C \\
& =\left[\begin{array}{c}
\mathbf{c}_{1}^{\prime} \\
\mathbf{c}_{2}^{\prime} \\
\vdots \\
\mathbf{c}_{M}^{\prime}
\end{array}\right]\left[\begin{array}{cccc}
r_{1} p_{1}^{Z}\left(1-p_{1}^{Z}\right) & 0 & \cdots & 0 \\
0 & r_{2} p_{2}^{Z}\left(1-p_{2}^{Z}\right) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & r_{k} p_{k}^{Z}\left(1-p_{k}^{Z}\right)
\end{array}\right]\left(\mathbf{c}_{1}, \cdots, \mathbf{c}_{M}\right) \\
& =\left[\begin{array}{cccc}
c_{11} r_{1} p_{1}^{Z}\left(1-p_{1}^{Z}\right) & c_{12} r_{2} p_{2}^{Z}\left(1-p_{2}^{Z}\right) & \cdots & c_{1 k} r_{k} p_{k}^{Z}\left(1-p_{k}^{Z}\right) \\
c_{21} r_{1} p_{1}^{Z}\left(1-p_{1}^{Z}\right) & c_{22} r_{2} p_{2}^{Z}\left(1-p_{2}^{Z}\right) & \cdots & c_{2 k} r_{k} p_{k}^{Z}\left(1-p_{k}^{Z}\right) \\
\vdots & \vdots & \ddots & \vdots \\
c_{M 1} r_{1} p_{1}^{Z}\left(1-p_{1}^{Z}\right) & c_{M 2} r_{2} p_{2}^{Z}\left(1-p_{2}^{Z}\right) & \cdots & c_{M k} r_{k} p_{k}^{Z}\left(1-p_{k}^{Z}\right)
\end{array}\right]\left(c_{1}, \cdots, c_{M}\right) \\
& {\left[\begin{array}{lllll}
\sum_{l=1}^{k} c_{1 l} c_{1 l} r_{l} p_{l}^{Z}\left(1-p_{l}^{Z}\right) & \sum_{l=1}^{k} c_{1 l} c_{2 l} r_{l} p_{l}^{Z}\left(1-p_{l}^{Z}\right) & \cdots & \sum_{l=1}^{k} c_{1 l} c_{M l} r_{l} p_{l}^{Z}\left(1-p_{l}^{Z}\right)
\end{array}\right]} \\
& =\begin{array}{cccc}
\sum_{l=1}^{k} c_{2 l} c_{1 l} r_{l} p_{l}^{Z}\left(1-p_{l}^{Z}\right) & \sum_{l=1}^{k} c_{2 l} c_{2 l} r_{l} p_{l}^{Z}\left(1-p_{l}^{Z}\right) & \cdots & \sum_{l=1}^{k} c_{2 l} c_{M l} r_{l} p_{l}^{Z}\left(1-p_{l}^{Z}\right) \\
\vdots & \vdots & \ddots & \vdots
\end{array} \\
& \left.\sum_{l=1}^{k} c_{M l} c_{1 l} r_{l} p_{l}^{Z}\left(1-p_{l}^{Z}\right) \quad \sum_{l=1}^{k} c_{M l} c_{2 l} r_{l} p_{l}^{Z}\left(1-p_{l}^{Z}\right) \quad \cdots \quad \sum_{l=1}^{k} c_{M l} c_{M l} r_{l} p_{l}^{Z}\left(1-p_{l}^{Z}\right)\right] \\
& =\left[\begin{array}{cccc}
\mathbf{c}_{1}^{\prime} V\left(\mathbf{p}^{\mathrm{Z}}\right) \mathbf{c}_{1} & \mathbf{c}_{1}^{\prime} V\left(\mathbf{p}^{\mathbf{Z}}\right) \mathbf{c}_{2} & \cdots & \mathbf{c}_{1}^{\prime} V\left(\mathbf{p}^{\mathrm{Z}}\right) \mathbf{c}_{M} \\
\mathbf{c}_{2}^{\prime} V\left(\mathbf{p}^{Z}\right) \mathbf{c}_{1} & \mathbf{c}_{2}^{\prime} V\left(\mathbf{p}^{Z}\right) \mathbf{c}_{2} & \cdots & \mathbf{c}_{2}^{\prime} V\left(\mathbf{p}^{Z}\right) \mathbf{c}_{M} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{c}_{M}^{\prime} V\left(\mathbf{p}^{Z}\right) \mathbf{c}_{1} & \mathbf{c}_{M}^{\prime} V\left(\mathbf{p}^{Z}\right) \mathbf{c}_{2} & \cdots & \mathbf{c}_{M}^{\prime} V\left(\mathbf{p}^{Z}\right) \mathbf{c}_{M}
\end{array}\right] .
\end{aligned}
$$

Let

$$
B\left(\mathbf{p}^{Z}\right)=\left[\begin{array}{cccc}
\frac{1}{\sqrt{\mathbf{c}_{1}^{\prime} V\left(\mathbf{p}^{\mathrm{Z}}\right) \mathbf{c}_{1}}} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \frac{1}{\sqrt{\mathbf{c}_{2}^{\prime} V\left(\mathbf{p}^{\mathrm{Z}}\right) \mathbf{c}_{2}}} & \cdots & \mathbf{0} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \frac{1}{\sqrt{\mathbf{c}_{M}^{\prime} V\left(\mathbf{p}^{\mathrm{Z}}\right) \mathbf{c}_{M}}}
\end{array}\right]
$$

then

$$
\sqrt{N} B\left(\mathbf{p}^{Z}\right)\left(C^{\prime} \widehat{\mathbf{p}}^{Z}-C^{\prime} \mathbf{p}^{Z}\right)=\left[\begin{array}{c}
\widetilde{Z}_{1} \\
\widetilde{Z}_{2} \\
\vdots \\
\widetilde{Z}_{M}
\end{array}\right]-\left[\begin{array}{c}
\mu_{1}\left(\mathbf{p}^{Z}\right) \\
\mu_{2}\left(\mathbf{p}^{Z}\right) \\
\vdots \\
\mu_{M}\left(\mathbf{p}^{Z}\right)
\end{array}\right] \xrightarrow{d} N(\mathbf{0}, \mathbf{R})
$$

where

$$
\begin{gathered}
\widetilde{Z}_{m}=\frac{\mathbf{c}_{m}^{\prime} \hat{\mathbf{p}}^{Z}}{\sqrt{\mathbf{c}_{m}^{\prime} I_{N}^{-1}\left(\mathbf{p}^{Z}\right) \mathbf{c}_{m}}}=\frac{\sum_{i=1}^{k} c_{m i} \hat{p}_{i}^{Z}}{\sqrt{\sum_{i=1}^{k} p_{i}^{Z}\left(1-p_{i}^{Z}\right) c_{m i}^{2} / n_{i}}}=\frac{\sqrt{N \mathbf{c}_{m}^{\prime}} \hat{\mathbf{p}}^{Z}}{\sqrt{\mathbf{c}_{m}^{\prime} V\left(\mathbf{p}^{Z}\right) \mathbf{c}_{m}}}, \\
\mu_{m}\left(\mathbf{p}^{Z}\right)=\frac{\mathbf{c}_{m}^{\prime} \mathbf{p}^{Z}}{\sqrt{\mathbf{c}_{m}^{\prime} I_{N}^{-1}\left(\mathbf{p}^{Z}\right) \mathbf{c}_{m}}}=\frac{\sum_{i=1}^{k} c_{m i} p_{i}^{Z}}{\sqrt{\sum_{i=1}^{k} p_{i}^{Z}\left(1-p_{i}^{Z}\right) c_{m i}^{2} / n_{i}}}=\frac{\sqrt{N \mathbf{c}_{m}^{\prime} \mathbf{p}^{Z}}}{\sqrt{\mathbf{c}_{m}^{\prime} V\left(\mathbf{p}^{Z}\right) \mathbf{c}_{m}}}, \\
\mathbf{R}=B\left(\mathbf{p}^{Z}\right) C^{\prime} V\left(\mathbf{p}^{Z}\right) C B^{\prime}\left(\mathbf{p}^{Z}\right)=\left(\rho_{i j}\right)_{M \times M} \text { and } \\
\rho_{i j}=\frac{\mathbf{c}_{i}^{\prime} V\left(\mathbf{p}^{Z}\right) \mathbf{c}_{j}}{\sqrt{\mathbf{c}_{i}^{\prime} V\left(\mathbf{p}^{Z}\right) \mathbf{c}_{i}} \sqrt{\mathbf{c}_{j}^{\prime} V\left(\mathbf{p}^{Z}\right) \mathbf{c}_{j}}} \\
=\frac{\sum_{l=1}^{k} c_{i l} c_{j l} r_{l} p_{l}^{Z}\left(1-p_{l}^{Z}\right)}{\sqrt{\sum_{l=1}^{k} c_{i l}^{2} r_{l} p_{l}^{Z}\left(1-p_{l}^{Z}\right)} \sqrt{\sum_{l=1}^{k} c_{j l}^{2} r_{l} p_{l}^{Z}\left(1-p_{l}^{Z}\right)}}
\end{gathered}
$$

Therefore

$$
\widetilde{\mathbf{Z}}=\left[\begin{array}{c}
\widetilde{Z}_{1} \\
\widetilde{Z}_{2} \\
\vdots \\
\widetilde{Z}_{M}
\end{array}\right]=\sqrt{N} B\left(\mathbf{p}^{Z}\right) C^{\prime} \hat{\mathbf{p}}^{Z} \sim A N\left(\boldsymbol{\mu}\left(\mathbf{p}^{Z}\right), \mathbf{R}\right)
$$

where $\boldsymbol{\mu}\left(\mathbf{p}^{Z}\right)=\left[\begin{array}{c}\mu_{1}\left(\mathbf{p}^{Z}\right) \\ \mu_{2}\left(\mathbf{p}^{Z}\right) \\ \vdots \\ \mu_{M}\left(\mathbf{p}^{Z}\right)\end{array}\right]$. Furthermore we have

$$
\begin{aligned}
\mathbf{Z} & =\sqrt{N} B\left(\widehat{\mathbf{p}}^{Z}\right) C^{\prime} \widehat{\mathbf{p}}^{Z}=\left(B\left(\widehat{\mathbf{p}}^{Z}\right) B^{-1}\left(\mathbf{p}^{Z}\right)\right) \sqrt{N} B\left(\mathbf{p}^{Z}\right) C^{\prime} \widehat{\mathbf{p}}^{Z} \\
& =\left(B\left(\widehat{\mathbf{p}}^{Z}\right) B^{-1}\left(\mathbf{p}^{Z}\right)\right) \widetilde{\mathbf{Z}}
\end{aligned}
$$

and the Law of Large Number shows that $\widehat{\mathbf{p}}^{Z} \xrightarrow{P} \mathbf{p}^{Z}$ as $N \rightarrow \infty$, this implies that $B\left(\widehat{\mathbf{p}}^{Z}\right) \xrightarrow{P}$ $B\left(\mathbf{p}^{Z}\right)$ and $B\left(\widehat{\mathbf{p}}^{Z}\right) B^{-1}\left(\mathbf{p}^{Z}\right) \xrightarrow{P} I_{M}$ an $M \times M$ identity matrix by the Continuity Theorem. Since under null hypothesis that $H_{0}: \mathbf{c}_{m}^{\prime} \mathbf{p}^{Z}=0$ for all $m=1, \cdots, M$, we have

$$
\widetilde{\mathbf{Z}} \xrightarrow{d} N(\mathbf{0}, \mathbf{R}) \text { as } N \rightarrow \infty,
$$

multivariate version of Slutsky's Theorem implies that

$$
\mathbf{Z}=\left(B\left(\widehat{\mathbf{p}}^{\mathrm{Z}}\right) B^{-1}\left(\mathbf{p}^{\mathrm{Z}}\right)\right) \widetilde{\mathbf{Z}} \xrightarrow{d} I_{M} N(\mathbf{0}, \mathbf{R})=N(\mathbf{0}, \mathbf{R}) \text { as } N \rightarrow \infty .
$$

This shows that the asymptotic correlation of $Z_{i}$ and $Z_{j}$ is $\rho_{i j}$. Under alternative, we can consider

$$
\boldsymbol{\mu}\left(\mathbf{p}^{Z}\right)=\sqrt{N} B\left(\mathbf{p}^{Z}\right) C^{\prime} \mathbf{p}^{Z}=\left[\begin{array}{c}
\frac{\mathbf{c}_{1}^{\prime} \mathbf{p}^{Z}}{\sqrt{\mathbf{c}_{1}^{\prime} I_{N}^{-1}\left(\mathbf{p}^{Z}\right) \mathbf{c}_{1}}} \\
\frac{\mathbf{c}_{2}^{\prime}}{\sqrt{\mathbf{c}_{2}^{\prime} I_{N}^{-1}\left(\mathbf{p}^{Z}\right) \mathbf{c}_{2}}} \\
\vdots \\
\frac{\mathbf{c}_{M}^{\prime} \mathbf{p}^{Z}}{\sqrt{\mathbf{c}_{M}^{\prime} I_{N}^{-1}\left(\mathbf{p}^{Z}\right) \mathbf{c}_{M}}}
\end{array}\right]
$$

as the approximated non-central parameter for the Wald test statistics $\mathbf{Z}$ and $\mathbf{R}$ the approximated correlation matrix of $\mathbf{Z}$. The large sample assumption is reasonable assumption under the Phase II or III trials since sample sizes are relatively large.

Each single contrast test can be translated into a decision procedure to determine whether a given dose-response shape is statistically significant. Under $H_{0}: \mathbf{c}_{m}^{\prime} \mathbf{p}^{Z}=0$ for all $m=1, \cdots, M$, we have shown that the asymptotic joint distribution of the vector $\mathbf{Z}=\left(Z_{1}, \ldots, Z_{M}\right)^{T} \sim A N(\mathbf{0}, \mathbf{R})$ asymptotic multivariate normal distribution with mean $\mathbf{0}$ and covariance (correlation) matrix $\mathbf{R}=\left(\rho_{i j}\right)_{M \times M}$, where

$$
\rho_{i j}=\frac{\sum_{l=1}^{k} c_{i l} c_{j l} r_{l} p_{l}^{Z}\left(1-p_{l}^{Z}\right)}{\sqrt{\sum_{l=1}^{k} c_{i l}^{2} r_{l} p_{l}^{Z}\left(1-p_{l}^{Z}\right)} \sqrt{\sum_{l=1}^{k} c_{j l}^{2} r_{l} p_{l}^{Z}\left(1-p_{l}^{Z}\right)}} .
$$

The statistic $Z_{\max }=\max \left(Z_{1}, \ldots, Z_{M}\right)$ will be used to make the final decision rule that whether there is a PoC. The computation of the critical value $q_{1-\alpha}$ should account for the multiplicity to control the FWER at a pre-specified level $\alpha$ (Hochberg and Tamhane, 1987), where $q_{1-\alpha}$ satisfies

$$
P\left(\max \left(U_{1}, \ldots, U_{M}\right)>q_{1-\alpha}\right)=\alpha,
$$

with $\mathbf{U}=\left(U_{1}, \ldots, U_{M}\right) \sim N(\mathbf{0}, \mathbf{R})$ a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{R}$.

### 13.2.2 Finding Optimal Model Contrasts

To determine the "best" contrast associated with a given model function $g\left(d_{i}, \theta_{Z}\right)$, when the model is correct, will maximize the chance of rejecting the associated null hypothesis. This can be done by maximizing the approximated non-centrality parameter $\tau=\tau(\boldsymbol{c})$ given $\mathbf{p}^{Z}$. Thus for each model $g\left(d_{i}, \boldsymbol{\theta}_{\mathrm{Z}}\right)$ we should choose $\mathbf{C}_{o p t}(g)$ such that

$$
\mathbf{C}_{\text {opt }}(g)=\underset{\mathbf{c}}{\operatorname{argmax}} h\left(\mathbf{c}, \mathbf{p}^{Z}\right),
$$

where

$$
h\left(\mathbf{c}, \mathbf{p}^{Z}\right)=\frac{\left(\mathbf{c}^{\prime} \mathbf{p}^{Z}\right)^{2}}{\mathbf{c}^{\prime} I_{N}^{-1}\left(\mathbf{p}^{Z}\right) \mathbf{c}}=\frac{\sum_{i=1}^{k} c_{i} p_{i}^{Z}}{\sqrt{\sum_{i=1}^{k} p_{i}^{Z}\left(1-p_{i}^{Z}\right) c_{i}^{2} / n_{i}}}=[\tau(\mathbf{c})]^{2}
$$

and $\mathbf{p}^{Z}$ is determined based on the model $g\left(d_{i}, \boldsymbol{\theta}_{Z}\right)$ with best guess of $\boldsymbol{\theta}_{Z}$ from previous, pilot or other relevant studies.

Without loss of generality, we can assume that the contrast vector $\mathbf{c}=\left(c_{1}, \ldots, c_{k}\right)^{\prime}$ follow the regularity conditions $\sum_{i=1}^{k} c_{i}=0$ and $\sum_{i=1}^{k} c_{i}^{2}=1$. With this condition, the maximization was on the unit sphere (a compact set) in $R^{k}$. This shows the existence of the value $\mathbf{C}_{\text {opt }}(g)=\left(c_{1}^{\max }, \ldots, c_{k}^{\max }\right)^{\prime}$ on the unit sphere that maximizes $h\left(\mathbf{c}, p^{Z}\right)$. The close form of the solution may be hard to find, but numeric calculation can be done with optimization softwares.

## Bibliography

[1] Abelson, R. and Tukey, J. (1963). Efficient utilization of non-numerical information in quantitative analysis: General theory and the case of simple order. Annals of Mathematical statistics, 34, 1347-1369.
[2] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. 2nd international Symposium on Information Theory, 267-281.
[3] Bauer, P. and Hackl, P. (1985). The application of Hunter's inequality in simultaneous testing. Biometrical Journal 27, 25-38.
[4] Beal, s.L., and Sheiner, L.B. (1982). Estimating population Kinetics, CRC Critical Reviews in Biomedical Engineering, 8, 195-222.
[5] Beal, s.L., and Sheiner, L.B. (1988). Heteroscedastic nonlinear regression, Techometics, 30, 327-338.
[6] Bates,D. and Watts D. (1988). Nonlinear regression analysis and its applications. New York: Willey.
[7] Bates,D. and Watts D. (1992). Nonlinear models. In Statistical Models, S.J.M. Chambers and T.J. Hastie(eds), 421-454.
[8] Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988). The New S Language.
[9] Belisle, C. J. P. (1992). Convergence theorems for a class of simulated annealing algorithms on Rd. J Applied Probability, 29, 885-895.
[10] Berry D.A.,Muller P,. Grieve A.P., Smith M., Parke,T., Blazek, R., Mitchard, N., and Krams, M. (2001). Adaptive Bayesian designs for dose-ranging drug trials. Pp99-181 in Gatsonis, C. Carlin,, B., and Carriquiry, A. (eds). Case studies in Bayesian statistics V.Springer, New York.
[11] Bornkamp, B., Bretz, F., Dmitrienko, A., Enas G., Gaydos, B., Hsu, C.H., Konig, F., Krams, M., Liu, Q., Neuenschwander, B., Parke, T., Pinheiro,J., Roy A., Sax, R., and Shen, F (2007). Innovative approaches for designing and analyzing adaptive doseranging trials. Journal of Biopharmaceutical Statistics 17(6), 965-995.
[12] Bretz, F., Pinheiro, J. and Bransosn, M. (2005). Combining multiple comparisons and modeling techniques in dose-response studies. Biometrics 61, 738-748.
[13] Bretz, F., Hsu, J., Pinheiro, J. and Liu, Y. (2008) Dose-finding-A challenge in Statistics. Biometrical Journal 50(4), 480-504.
[14] Buckland, S., burham, K., and Augustion, N. (1997). Model selection: An integral part of inference. Biometrics 53, 603-618.
[15] Clyde, M., and George, E.I. (2004). Model Uncertainty. Statistical Science, 19: 81-94
[16] Cox, D.R. (1972). The analysis of multivariate binary data. Applied Statistics, 21, 113120.
[17] Cox, D.R., and Wermuth, N. (1992). Response models for mixed binary and quantitative variables Biometrika 79, 441-461.
[18] Dluhy RG, Williams GH (2004). Aldosterone-villain or bystander? NEJM, 351(1): 8-10.
[19] Dutt, J.E., Mattes, K.D., Somas, A.P., and Tao, L. C. (1976). An approximation to the maximum modulus of the trivariate T with a comparison to exact values.Biometrics 32, 465-469.
[20] Dunnett, C.W. (1955). A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association 50, 1096-1121.
[21] Efron, B. (1979). Bootstrap. Another look at Jackknife. Ann. Statistics. 7, No.1, 1-26.
[22] Efron, B. (1982). The Jackknife, the Bootstrap and other resampling plans. SIAM, Philadelphia.
[23] Efron, B. (1987). Better bootstrap confidence intervals (with Discussion). Journal of the American Statistical Association, 82, 171-200.
[24] Efron, B. and Tibshirani, R. (1993). An introduction to the Bootstrap. Chapman and Hall, New York.
[25] FDA (2004). Innovation/Stagnation: Challenge and opportunity on the critical path to new medical products. FDA report from March 2004.
[26] FDC Report (1991). FDC Reports from May 6, 1991. The Pink Sheet 53(18), 14-15.
[27] Finner, H. and Strassburger, K. (2002). The partioning principle: A powerful tool in multiple decison theory. Annals of Statistics 30, 1194-1213.
[28] Fitzmaurice, G.M., and Laird, N.M. (1995). Regression models for mixed discrete and continuous responses with potentially missing values. Biometrics, 53, 110-122.
[29] Fitzmaurice, G.M., and Laird, N.M. (1995). Regression models for a bivariate discrete and continuous outcome with clustering. Journal of the American statistical association, 90, 845-852
[30] Freedman LS, Graubard BI, Schatzkin A.(1992). Statistical validation of intermediate endpoints for chronic diseases. Statistics in Medicine, 11, 167-178.
[31] Genz, A. and Bretz, F. (2002). Comparison of methods for the computation of multivariate t-probabilities. Journal of computational and graphical statistics, 11, 950-971.
[32] Gooley, T. A., Martin, P.J., Fisher, L.D,. and Pettinger, M. (1994). Simulation as a design tool for phase I/II clinical trials: An example from bone marrow transplantation. The Annals of Statistics, 15, 450-462.
[33] Hall, D.B., and Severini, T.A. (1998). Extended generalized estimating equations for clustered data, Journal of the American statisticial association, 93, 1365-1375.
[34] Hochberg, Y. and Tamhane, A. (1987). Multiple comparison procedures. New York: Wiley.
[35] Hothorn, T., Bretz, F. and genz, A. (2001). On multivariate $t$ and Gauss probabilities in R. R newsletter 1(2), 27-29.
[36] Epstein M, Williams GH, Weinberger M (2006). Selective aldosterone blockade with eplerenone reduces albuminuria in patients with type 2 diabetes. CJASN, 1(5): 940951.
[37] Flack JM, Oparil S, Pratt H (2003). Efficacy and tolerability of eplerenone and losartan in hypertensive black and white patients. JACC, 41: 1148-1155.
[38] Hochberg,Y. and Tamhane, A.C. (1987). Multiple comparisons procedures. Wiley, New York.
[39] Hoeting,J., Madigan, D., Raftery, A.E., and Volinsky, C.T. 1999. Bayesian model averaging. Statistical Science, 14: 382-417.
[40] Hsu, J.C. (1996). Multiple comparisons. Chapman and Hall, New York.
[41] Krazanowski, W.J. (1988), Principles of multivariate analysis: A user's perspective, Oxford: Clarendon.
[42] Krum H, Nolly H, Workman D (2002) Efficacy of eplerenone added to reninangiostensin blockade in hypertensive patients. Hypertension, 40: 117-123.
[43] Loader, C. (1999). Local regression and likelihood. New York, Springer.
[44] Pharmacological Therapy for Macular Degeneration Study Group. Interferon $\alpha$-IIA is ineffective for patients with choroidal neovascularization secondary to age-related macular degertion. Results of a prospective randomized placebo-controlled clinical trial. Archives of Ophthalmology 1997; 115: 865-892.
[45] Pinheiro, J., Bretz, F., and Branson, M. (2006) Analysis of dose-response studiesmodeling approaches. In: Ting,N., ed. Dose finding in drug development. New York: Springer
[46] Pinheiro, J. and Bates, D. (2000). Mixed-effects models in S and S-PLUS. New York: Springer.
[47] Pinheiro, J.,Bornkamp, B., and Bretz, F. (2006). Design and analysis of dose finding studies combining multiple comparisons and modeling procedures. Journal of Biopharmaceutical Statistics, 16(5), 639-656.
[48] Prentice, R.L., and Zhao, L.P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. Biometrics, 47, 825839
[49] Quigley, J., Hughes, M.D., and Fenton, T.(2001). Dose-finding designs for HIV studies. Biometrics 57, 1018-1029.
[50] Regan, M.M., and Catalano, P.J. (1999). Likelihood models for clustered binary and continuous outcomes: Application to developmental toxicology, Biometrics, 55, 760768
[51] Robertson, T., Wright, F. and Dykstra, R. (1988). Order Restricted Statistical Reference. New York: Wiley.
[52] Rochon, J. (1996). Analyzing bivariate repeated measures for discrete and continuous outcome variables, Biometrics, 41, 1-9.
[53] Rossi S (Ed.). (2004). Australian Medicines Handbook. ISBN 0-9578521-4-2.
[54] Seber,F.A., and Wild,C.J. (1989). Nonlinear Regression. New York: Wiley.
[55] Shih, Weichung Joseph., and Huang, Wei-min (1992). Evaluating correlation with proper bounds. Biometrics 48, 1207-1213.
[56] Shimodaira, H. (1998). An application of multiple comparison techniques to model selection. Annals of the Institute of Statistical Mathematics 50, 1-13.
[57] Schwartz, G. (1978). Estimating the dimension of a model. The Annals of Statistics 6, 461-464.
[58] Stefansson, G., Kim, W.C., and Hsu, J.C. (1988). On confidence sets in multiple comparisons. In: Statistical Decision Thory and Related Topics IV (S.S. Gupta and J.O. Berger, eds.) 2, 89-104. Academic press, New York.
[59] Tamhane, A., Dunnett, C., and hochberg Y. (1996). Multiple testing procedures for dose finding. Biometrics 52, 21-37.
[60] Thall, P.F and Russell, K.T. (1998). A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in Phase I/II clinical trials. Biometrics 54, 251264.
[61] Thall, P.F and Cook, J.D. (2004). Dose-finding based on efficacy-toxicity trade off. Biometrics, 60, 684-693
[62] Thall PF, Nguyen HQ and Estey EH. (2008). Patient-specific dose finding based on bivariate outcomes and covariates. Biometrics 64(4), 1126-1136.
[63] Thomas MC. (2000). Diuretics, ACE inhibitors and NSAIDs - the triple whammy. Med J Aust 2000;172(4):184-185. PMID 10772593
[64] Tukey, J., Ciminera, J., and Heyse, J. (1985). testing the statistical certainty of a response to increasing dose of a drug. Biometrics 41, 295-301.
[65] Vonesh, E.F., and Carter, R.L. (1992). Mixed-effects nonlinear regression for unbalanced repeated measures, Biometrics, 48, 1-17.
[66] Vonesh, E.F., and Chinchilli, V.M. (1997), Linear and nonlinear models for the analysis of repeated measurements. New York: Marcel Dekker.
[67] Vonesh, E.F., Wang, Hao and Majumdar, D.(2001). Generalized least squares, taylor series linearization, and Fisher's scoring in multivariate nonlinear regression. Journal of the American Statistical Association, Vol 96, No. 453, 282-291
[68] Pan, W. (2001). Akaike's information criterion in generalized estimating equations. Biometrics, 57, 120-125.
[69] Williams, D.A. (1971). A test for differences between treatment means when several dose levels are compared with a zero dose level. Biometrics, 27, 103-117.
[70] Williams, D.A. (1972). The comparisons of several dose levels with a zero dose control. Biometrics, 28, 519-531.
[71] Zeger, S.L., and Liang, K.Y. (1991). Feedback models for discrete and continuous time series, Statistica Sinica, 1, 51-64.

Vita
Aiyang Tao

## EDUCATION

1991 MD of Medical Science, Nanjing Medical School, Nanjing, China.
1997 Ph.D. candidate, MS in Molecular Biology, University of Connecticut, Storrs, CT.
1998 MS in Statistics, University of Connecticut, Storrs, CT.
2010 Ph.D. in Biostatistics, School of Public Health, University of Medicine \& Dentistry of New Jersey.

## EXPERIENCE

09/90-09/91 Resident Doctor, Jiangsu General Hospital, Nanjing, P.R. China.
08/93-08/98 Research and Teaching Assistant, University of Connecticut, Storrs, CT.
05/98-08/98 Biostatistician as intern, Bristol-Myers Squibb Company, Wallingford, CT.
08/98-01/99 Research Assistant, University of Connecticut Health Center, Storrs, CT.
02/99-08/99 SAS analyst, IPR and Data Solutions Company, Norwalk, CT.
08/99-07/00 Biostatistician, Kendle, Cranford, NJ.
07/00-04/01 SAS Analyst as consultant, Hoffman Laroche Inc., Nutley, NJ.
04/01-04/03 Senior Biostatistician as consultant, Pfizer Inc., Morris Plains, NJ.
04/03-08/04 Senior Biostatistician as consultant, Novartis Pharma Inc., East Hanover, NJ.
08/04-03/09 Senior Principle Statistician, Novartis Pharma., East Hanover, NJ.
04/09-05/10 Associate Director, Program Statistician, Novartis Pharma., East Hanover, NJ.

