# INDIVIDUAL MEDIATING EFFECTS AND THE CONCEPT OF TERMINAL MEASURES DATA

by

ROSHAN NIRANJALA SERASINGHE

B.Sc., UNIVERSITY OF PERADENIYA, SRI LANKA, 1998
M.Sc., UNIVERSITY OF PERADENIYA, SRI LANKA, 2003

AN ABSTRACT OF A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree
DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

# Abstract

Researches in the fields in science and statistics often go beyond the two-variable cause-and-effect relationship, and also try to understand what connects the causal relationship and what changes the magnitude or direction of the causal relationship between two variables, predictor ( $T$ ) and outcome ( $Y$ ).

A mediator ( $Z$ ) is a third variable that links a cause and an effect, whereby $T$ causes the $Z$ and $Z$ causes $Y$. In general, a given variable may be said to function as a mediator to the extent that it accounts for the relation between the predictor and the outcome (Baron and Kenny, 1986).

The initial question regards the appropriate characterization of a mediation effect. Most studies, when comparing one or more treatments focus on an average mediating effect. This average mediating effect can be misleading when the mediating effects vary from subject to subject in the population. The primary focus of this research is to investigate individual mediating effects in a population, and to define a variance of these individual mediating effects. A concept called subject-mediator (treatment) interaction is presented and its role in evaluating a mediator's behavior on a population of units is studied. This is done using a framework sometimes called a counterfactual model. Some common experimental designs that provide different knowledge about this interaction term are studied. The subgroup analysis is the most common analytic approach for examining heterogeneity of mediating effects.

In mediation analysis, situations can arise where $Z$ and $Y$ cannot both be measured on an individual unit. We refer to such data as terminal measures data. We show a design where a mediating effect cannot be estimated in terminal measures data and another one where it can be, with an assumption. The assumption is linked to the idea of pseudo-replication. These ideas are discussed and a simulation study illustrates the issues involved when analyzing terminal measures data. We know of no methods that are currently available that specifically address terminal measures data.

# INDIVIDUAL MEDIATING EFFECTS AND THE CONCEPT OF TERMINAL MEASURES DATA

by

ROSHAN NIRANJALA SERASINGHE

B.Sc., UNIVERSITY OF PERADENIYA, SRI LANKA, 1998
M.Sc., UNIVERSITY OF PERADENIYA, SRI LANKA, 2003

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree
DOCTOR OF PHILOSOPHY

Department Of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Approved by:
Major Professor
**Dr. Gary Gadbury**

# Copyright

# Abstract

Researches in the fields in science and statistics often go beyond the two-variable cause-and-effect relationship, and also try to understand what connects the causal relationship and what changes the magnitude or direction of the causal relationship between two variables, predictor( $T$ ) and outcome ( $Y$ ).

A mediator ( $Z$ ) is a third variable that links a cause and an effect, whereby $T$ causes the $Z$ and $Z$ causes $Y$. In general, a given variable may be said to function as a mediator to the extent that it accounts for the relation between the predictor and the outcome (Baron and Kenny, 1986).

The initial question regards the appropriate characterization of a mediation effect. Most studies, when comparing one or more treatments focus on an average mediating effect. This average mediating effect can be misleading when the mediating effects vary from subject to subject in the population. The primary focus of this research is to investigate individual mediating effects of a population, and to define a variance of these individual mediating effects. A concept called subject-mediator (treatment) interaction is presented and its role in evaluating a mediator's behavior on a population of units is studied. This is done using a framework sometimes called a counterfactual model. Some common experimental designs that provide different knowledge about this interaction term are studied. The subgroup analysis is the most common analytic approach for examining heterogeneity of mediating effects.

In mediation analysis, situations can arise where $Z$ and $Y$ cannot both be measured on an individual unit. We refer to such data as terminal measures data. We show a design where a mediating effect cannot be estimated in terminal measures data and another one where it can be, with an assumption. The assumption is linked to the idea of pseudo-replication. These ideas are discussed and a simulation study illustrates the issues involved when analyzing terminal measures data. We know of no methods that are currently available that specifically address terminal measures data.

# Table of Contents

# List of Figures

# List of Tables

# Acknowledgements

No words on paper come close to describing my gratitude to the many people who have been part of this PhD journey over the past four-plus years. I would not be writing these words without their love, support, caring, dedication, belief, and sheer patience. I am deeply indebted to my advisor, Dr. Gary Gadbury, for his guidance, patience, support, encouragement, and all the time he dedicated for this dissertation. I am impressed by his insight of the research, devotion to work, teaching, and especially his willingness to help. This dissertation would have been impossible without him. My doctoral committee… Dr. Paul Nelson for his kind assistance, advice, encouragement …Dr. James W. Neill, Dr. David Renter, and Dr. Gabriel Nagy for their helpful suggestions and support.

I would like to express my sincere appreciation to Department of Statistics at Kansas State University for giving me a great opportunity to enhance my knowledge and work. A special thank goes to Dr. John Boyer for offering me a position as a graduate teaching assistantship at the department of statistics, Kansas State University. I would also like to thank all of my friends and collogues for your friendship and support. I am grateful for your love and will cherish all of them.

The most important and my heartfelt appreciation go to my parents and family for a life full of love and support. I always knew that they believed in me and wanted the best for me. They are the source of power and inspiration that keeps me persistent on research work. Without them this wouldn't have been a story of success.

Finally, and most importantly, I would like to thank my loving wife Kumary. Her assistance, encouragement, patience, supports, taking care and unwavering love was undeniably the bedrock upon which the past ten years of my life have been built. Her tolerance of my occasional vulgar moods is a testament itself of her unyielding devotion and love.

# Dedication

This is dedicated to my loving mother for your endless love, support and encouragement. I hope I have made you proud.

To my father who always stood behind me and knew I would succeed. Gone now but never forgotten. I miss you always and love you forever. Thanks for all you did. This work is dedicated to you.

To my beloved wife, Kumary.  To you, I owe more than words can capture.

To my loving Dilshi, Rawisha, Buchi and Himaru, consider my achievement a guiding light to expand your horizons.

# Chapter 1 - Introduction

The cause-and-effect relationship between two variables, predictor ($T$) and outcome ($Y$), has been a focus of much research in fields of science and statistics. The top panel of Figure 1.1 graphically illustrates the causal effect of $T$ on $Y$. Testing causal hypotheses not only verifies researchers' substantive theories around a phenomenon but also answers practical questions about whether an intervention or treatment program has the expected effect if a causal relation exists (Wu et al. 2008). Researches in this field often go beyond the two-variable cause-and-effect relationship and also try to understand what connects the causal relationship and what changes the magnitude or direction of the causal relationship between two variables, $T$ and $Y$. There are several tools that engage with these puzzles.

Mediation in its simplest form represents the addition of a third variable, $Z$, to this $T \rightarrow Y$ relation whereby $T$ causes the mediator, $Z$, and $Z$ causes $Y$, so $T \rightarrow Z \rightarrow Y$. A mediator is a third variable that links a cause and an effect. In general, a given variable may be said to function as a mediator to the extent that it accounts for the relation between the predictor and the outcome (Baron and Kenny, 1986).

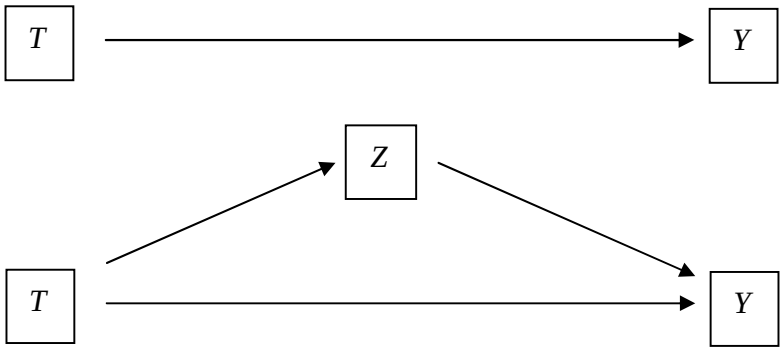



**Figure 1.1 Top panel: The Causal Effect of $T$ on $Y$. Bottom panel: Mediated Relationship among the variables ($T$ = Predictor, $Z$ = Mediator, $Y$ = Outcome)**

The primary focus of this research is to investigate individual effects of mediation that explains the process of ''why'' and ''how'' a cause-and-effect happens (Baron and Kenny 1986; Frazier et al. 2004). Hence, a mediation analysis attempts to ''identify the intermediary process

that leads from the independent variable to the dependent variable'' (Muller et al. 2005, 852-863). In other words, in a simple mediation model, the independent variable is presumed to cause the mediator, and in turn, the mediator causes the dependent variable. For this reason, a mediation effect is also termed an indirect effect, surrogate effect, intermediate effect, or intervening effect (MacKinnon et al. 2002). Figure 1.1 shows the mediated relationship among the variables and possible indirect and direct effects.

## 1.1 Research Examples Involving Potential Mediating Variables

Example (1): The Job Search Intervention Study (JOBS II)

JOBS II is a randomized field experiment that investigates the efficacy of a job training intervention on unemployed workers. The program is designed not only to increase reemployment among the unemployed but also to enhance the mental health of the job seekers.

In the experiment, unemployed workers received a prescreening questionnaire and were then randomly assigned to treatment and control groups. Those in the treatment group participated in job skills workshops in which participants learned job search skills and coping strategies for dealing with setbacks in the job search process. Those in the control condition received a booklet describing job search tips. In follow-up interviews, two key outcome variables were measured: a continuous measure of depressive symptoms based on the Hopkins Symptom Checklist and a binary variable, representing whether the respondent had become employed. Researchers who originally analyzed this experiment hypothesized that workshop attendance leads to better mental health and employment outcomes by enhancing participants' confidence in their ability to search for a job. In the JOBS II data, a continuous measure of job search self-efficacy represents this key mediating variable. The data also include baseline covariates measured before administering the treatment. The most important of these is the pretreatment level of depression, which is measured with the same methods as the continuous outcome variable. (Imai et al. 2010).

$T$ : Those in the "treatment group" participated in job skills workshops in which participants learned job search skills and coping strategies for dealing with setbacks in the job search process. Those in the "control condition" received a booklet describing job search tips.

$Z$: Job search self-efficacy, a continuous variable.

$Y_1$: measure of depressive symptoms based on the Hopkins Symptom Checklist, a continuous variable.

$Y_2$: whether the respondent had become employed, a binary variable.

Example (2): A study of nursing interventions for postoperative pain

This problem is an example about multiple treatments and multiple mediators. The primary objective of this trial was to compare four randomized interventions ('music', 'teaching', 'combination', and 'control') on the reduction of pain after surgery. Patients assigned to the music intervention were provided with a music tape; the teaching intervention involved a tape of instructions for the use of intravenous patient-controlled opioid analgesia; the combination involved the joint use of music and teaching; and patients assigned to control received standard care. Among the primary response variables was level of 'sensation', the patient's report of their sensation of pain at the site of the surgery, using a 100-point visual analog scale. Assessments were made at the beginning and end of five 20-min 'pre–post' tests during the first two days following surgery. Intermediate variables, measured during each pre–post test period, included level of patient-controlled analgesia ('pca'), and whether or not the patient was asleep at the time the nurse came in for the post-experiment assessment ('sleep') (Albert, 2008). Mediation analysis via potential outcomes models. Statistics in Medicine, 27, 1282–1304).

In our usual notation,

$T$: 'music', 'teaching', 'combination', and 'control'

$Z_1$: level of patient-controlled analgesia ('pca'), a continuous variable.

$Z_2$: whether or not the patient was asleep ('sleep'), a binary variable.

$Y$: level of 'sensation', the patient's report of their sensation of pain.

## 1.2 Motivation of Research

This research was motivated in part by Gadbury et al. (2001) who stressed that an evaluation of the effect of treatment relative to a control often focuses on estimating a mean treatment effect; however, the mean treatment effect may be misleading when the effect of the treatment varies widely across subjects. In their paper they proposed methods for evaluating treatment heterogeneity, which was called subject-treatment interaction. It was shown that estimators for subject-treatment interaction are sensitive to an inestimable correlation parameter. In addition to papers by Gadbury and colleagues on subject-treatment interaction, I happened to read two important papers which are related to causal mediation analysis. One paper was titled "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations" (Baron and Kenny, 1986). This paper introduced what is possibly the most popular approach of evaluating mediating effect and is often referred to as the "Baron & Kenny Approach". The other paper is "Mediation analysis via potential outcomes models" (Albert, 2008). This paper defined a mediating effects using a potential outcomes framework which has become a popular methodology for causal inference. More discussion of potential outcomes is in chapter 2. Relating mediation analysis to the concepts of subject treatment interaction led to the core ideas in my research topic.

## 1.3 Outline of the Dissertation

An important issue is the assessment of differences in mediating effects among individuals. Clearly, a necessary condition for a variable $Z$ to be a mediator of the effect of $T$ on $Y$ is that $Z$ precedes the outcome $Y$. Gadbury et al. (2010) had discussed the subject-treatment interaction or individual treatment variation. We extend this discussion to when a mediator is present. That is, the individual treatment effect that includes mediation will be discussed and explored using mediation plots. Some completed work from the preliminary exam will be presented in chapter 3.

Chapter 3 presents individual mediating effects and defines subject-mediator (treatment) interaction, S-M(T). In chapter 6, the terminal measures are described in detailed and the

methods of handling terminal measures are discussed further. Then a simulation study discusses topics such as pseudoreplication that plays a role when analyzing terminal measures data.

In chapter 5, the methods of estimating the variance of individual indirect effects for a two sample completely randomized block design with *s*-disjoint blocks and *n*-individuals for each block are discussed. The estimates of the variability of individual indirect effects are present. Evaluating treatment effects within subsets of patients plays a major part of the analysis of many major clinical trials. In order for the conclusion to be broad based, and to achieve reasonable sample sizes, the study design often includes a number of patient subsets, corresponding to different clinical centers or to patient subsets defined by prognostic factors. As a result, some variation in the estimates of treatment effect among subsets is expected. In chapter 6, we explain the subset analysis concept to evaluate heterogeneity of mediating effects and extend Gail and Simon (1985) test for qualitative interaction(QI) on treatment effects to QI on mediating effects. Chapter 7, summarizes the research and discusses possible future work.

# Chapter 2 - Literature Review

When we search the phrase “mediation analysis”, there are more than 500 papers in web of science search engine. When we use the word “mediating effects”, there are more than 1,000 research papers in many different areas. These researches have used various designs to analyze mediating effects.

Although several design frameworks for studying mediation effects have been proposed to date, the classic works by Kenny and colleagues (Baron and Kenny 1986; Judd and Kenny 1981; Kenny et al. 1998) appear to remain the most popular approaches.

## 2.1 Baron & Kenny Approach

Perhaps the most popular approach to mediation analysis is the *causal steps approach*. Although the method can be traced in some form back to the 1950s, it was made popular in 1980s by a very influential article by Reuben Baron and David Kenny published in the *Journal of Personality and Social Psychology*. For this reason, the causal steps approach has come to be known as the *Baron and Kenny* method. Historically, the vast majority of published mediation analyses are based on the logic of the causal steps approach, and this approach remains widely used today. Baron and Kenny (1986) proposed a multi step approach in which several regression analyses are conducted and significance of the coefficients is examined at each step. In their article, Baron and Kenny presented a simple, regression-based method requiring no specialized software, which has had a huge impact, i.e., to date it has been cited over 19,000 times (Web of Science). However, there are theoretical and empirical reasons for concern about the application of this method of assessing mediation. The 1986 article focused on the distinction between moderation and mediation and did not include extensive discussions about the complexities of path modeling and structural equation modeling (SEM), of which mediation analysis can be considered a special case.

Figure 2.1 shows a case in which $T$ causes $Y$ both directly and indirectly (through $Z$). In more complex models of such relations (e.g., with baseline covariates for $Z$ or $Y$), the paths between $T$, $Z$, and $Y$ are not estimated by the same simple and partial regression coefficients

depicted, but in every case there are coefficients that correspond to $\alpha, \beta, \gamma$ and $\gamma'$, so we use this notation to represent either case. Because of the linear model framework and resulting algebraic relations among regression coefficients (Wright, 1934), $\gamma = \alpha\beta + \gamma'$. If a model such as that depicted in Figure 2.1 is true, the empirical coefficients describe causal effects as follows: $\gamma'$ describes the direct effect of $T$ on $Y$, not acting through $Z$, and $\alpha\beta$ describes the indirect effect of $T$ on $Y$ through $Z$. The total effect of $T$ on $Y$ is defined as the sum of the direct and indirect effects; $\gamma$ describes the total effect of $T$ on $Y$ ( Lois, 2009 ).

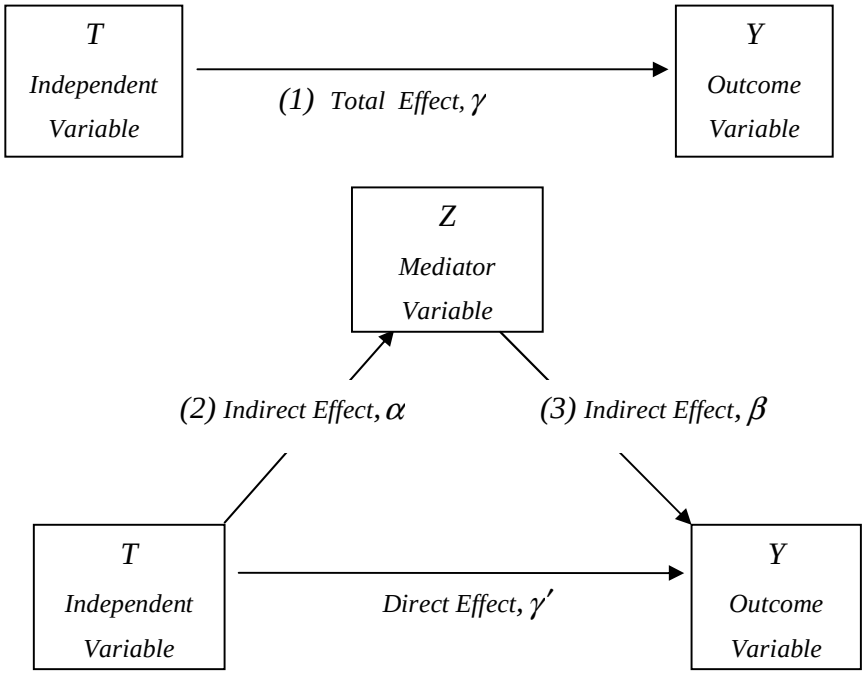



**Figure 2.1 The path diagram of the single mediation model.**

The Baron & Kenny (1986) regression equations are as follows:

$$Y_i = c_Y + \gamma T_i + \varepsilon_Y \qquad (1) \quad (\textit{Total Effect})$$

$$\left.\begin{aligned} Z_i &= c_Z + \alpha T_i + \varepsilon_Z \quad (2) \\ Y_i &= c'_Y + \gamma' T_i + \beta Z_i + \varepsilon'_Y \quad (3) \end{aligned}\right\} \ (\textit{Mediation Model})$$

where the $c$'s are the intercepts; the epsilons are the model errors, and the $\alpha, \beta, \gamma$ and $\gamma'$ terms are the regression coefficients capturing the relationships between the three focal variables. The

parameters $\gamma$ and $\gamma'$ relate the independent variable to the outcome variable, but $\gamma'$ is a partial effect adjusted for the effect of the mediator, $Z$. The parameter $\alpha$ relates the independent variable to the mediating variable, and $\beta$ relates the mediator to the dependent variable adjusted for the effects of the independent variable.

Baron and Kenny (1986) described mediation analysis in four steps. Step 1 involves testing the significance of $\gamma$ to determine that there is a relation to be mediated. If significant, one tests the significance of $\alpha$ (step 2) to demonstrate a relation between $T$ and $Z$. In step 3, a significant $\beta$ shows that there is a relation between $Z$ and $Y$ not accounted for by $T$. Once steps 2 and 3 are passed, evidence consistent with a nonzero indirect effect has been obtained; the model is consistent with either partial mediation, complete mediation, or suppression. In the first part of Step 4, the observed values of $\gamma$ and $\gamma'$ are compared; if $\gamma'$ is smaller than $\gamma$, the data are consistent with mediation; if $\gamma'$ is larger than $\gamma$, the data are consistent with suppression. No significance test is necessary for this step. If the data suggest that either mediation or suppression is present, one examines the significance of $\gamma'$ to determine if the data are consistent with partial versus complete mediation (the second part of step 4); if $\gamma'$ is smaller than $\gamma$ but significantly different from 0, the data are consistent with partial mediation. If $\gamma'$ is smaller than $\gamma$ but not significantly different from 0, the data are consistent with complete mediation.

MacKinnon et al. (2002) assembled, from a variety of disciplines, 14 methods for testing a mediating effect of an intermediate variable. They organized these approaches into three categories: 'causal steps', difference in coefficients, $\gamma - \gamma'$, and product of coefficients, $\alpha\beta$. While MacKinnon et al. made some recommendations based on their simulation study of coverage and power, they noted that the 'diversity of methods . . . indicates that there is no firm consensus across disciplines as to the definition of an intervening variable effect'. (Albert, 2008, 1282–1304).

Among the methods assembled in MacKinnon et al. (2002), there are two popular ways of calculating the mediated effect; the difference in coefficients & the product of coefficients.

*Difference in Coefficients:* Mediating Effect = $\gamma - \gamma'$; reflects how much of the relation between the independent variable and the dependent variable is explained by the mediation.

*Product of Coefficients:* Mediating Effect = $\alpha\beta$; reflects how much a one unit change in $T$ affects $Y$ indirectly through $Z$. By solving the equations $(1), (2)$ and $(3)$, a mediating effect can be estimated.

One of the assumptions of the mediation equations described above is that the relation from the mediator to the dependent variable is the same across levels of the independent variable, that is, the $Z$ to $Y$ relation does not differ across levels of $T$. A nonzero $TZ$ interaction effect suggests that the independent variable alters the relation between the mediator and the dependent variable. Under the assumption of no interaction between the independent variable, $T$, and the mediator, $Z$, $\alpha\beta = \gamma - \gamma'$. The existence of the interaction also means that the relation of the independent variable to the dependent variable differs across levels of the mediator. If the $TZ$ interaction is statistically significant, it is important to explore the source of the significant interaction with contrasts including simple effects and plots. If the $TZ$ interaction is statistically significant, the mediating effect is different from $\alpha\beta$, and this will be discussed later.

The standard errors of estimated mediating effects were estimated using Goodman's (1960) unbiased solution that seeks to estimate $\sqrt{\alpha^2\sigma_\beta^2 + \beta^2\sigma_\beta^2 - \sigma_\alpha^2\sigma_\beta^2}$. Let $a$, $b$, $c$ and $c'$ be the estimators of $\alpha, \beta, \gamma$ and $\gamma'$ respectively. The estimated mediating effect would be $ab$ (or $c - c'$) and the estimated standard error of estimates would be $\sqrt{a^2S_b^2 + b^2S_a^2 - S_a^2S_b^2}$ (McKinnon et al., 2002 ), where $S_a^2$ and $S_b^2$ are the estimated variances of parameter estimates $a$ and $b$ respectively.

The above equations $(1), (2)$ and $(3)$ can be modified when a single baseline covariate, $X$, is present as follows:

$$Y_i = c_Y + \gamma T_i + \eta X + \varepsilon_Y \qquad (4) \qquad (\textit{Total Effect})$$

$$\left.\begin{aligned} Z_i &= c_Z + \alpha T_i + \eta X + \varepsilon_Z \qquad (5) \\ Y_i &= c'_Y + \gamma' T_i + \beta Z_i + \eta X + \varepsilon'_Y \qquad (6) \end{aligned}\right\} \quad (\textit{Mediation Model})$$

The formula for the mediating effect does not change, but the value of the mediating effect may change when a single baseline covariate is present (Imai, 2010).

The appeal of this Baron & Kenny approach is that it clearly lays out the intuitive, structural relationships among treatment, mediator, and outcome, and can be applied with straightforward regression analyses. What has not always been clear, nor explicitly stated, is whether traditional mediation analyses lead to causal interpretations of mediation parameters. With certain assumptions the traditional approach to mediation can lead to causal interpretations (Imai et al. 2010), but these assumptions are quite stringent. In light of these assumptions, MacKinnon (2008) commented on traditional mediation analysis: "In many situations, the results of a mediation analysis are descriptive rather than implying causal relations" (p. 67). Over the last 20 years there has been a parallel line of work on mediation in the biostatistical and epidemiological literatures that is firmly rooted within Rubin's causal model framework (Rubin, 1974), including principal stratification (Gallop et al. 2009) and structural mean model, SMM (Robins & Greenland, 1992). Moreover, Pearl (2001) has made notable contributions to causal graphs and the identifiability of direct and indirect effects. Notably, these alternative approaches provide alternative estimation methods that relax key assumptions necessary for the traditional approach, while at the same time introducing their own set of assumptions. This study uses the traditional approach of Kenny and colleagues using Rubin's potential outcomes (i.e., counterfactual framework), which is widely used in the statistical literature and is described below.

## 2.2 Potential Outcomes Framework

Based on the early work on experimental design by Neyman (1923), Rubin (1974, 1986) formalized the counterfactual model for causal analysis of data from randomized experiments and observational studies. In statistics, the model is often referred to as the potential outcomes framework which is now commonly used in statistics for causal inference as well as in the fields of epidemiology, sociology, psychology and political science. The potential outcomes framework is sometimes called the 'Rubin Causal Model' (RCM) (Holland, 1986), but it has roots in the context of randomized experiments with randomization-based inference in the work of Neyman (1923) and Fisher (1925) (Rubin, 2004 ). Causal inference using potential outcomes is sometimes called causal inference involving counterfactuals.

The core of the counterfactual model is as follows. Suppose that each individual in a population of interest can be exposed to two alternative treatments which are referred to as treatment and control. The key assumption of the counterfactual framework is that each individual in the population of interest has a potential outcome under each treatment state, even though each individual can be observed in only one treatment state at any one particular point in time. These outcomes are termed counterfactuals because only one outcome can be observed for a given individual.

As formalized by Rubin (1974), in the potential outcomes framework, the effect of some treatment, $T_i = 1$ (vs. a control, $T_i = 0$) on an outcome, $Y_i$ or an individual can be expressed as the difference between two potential outcomes, $Y_i^{(1)} - Y_i^{(0)}$ where $Y_i^{(1)}$ is the value of the outcome the individual would experience if exposed to the treatment, and $Y_i^{(0)}$ is the outcome the individual would experience if exposed to the control. The fundamental problem of causal inference (Holland, 1986) is that only one potential outcome can be observed for each person at a given time.

Because only either $Y_i^{(0)}$ or $Y_i^{(1)}$ is observable, even randomized experiments cannot identify this individual-level causal effect. Thus, researchers often focus on the identification and estimation of the average causal effect (ACE), which is defined as $E(Y_i^{(1)} - Y_i^{(0)})$, where the expectation is taken with respect to the random sampling of units from a target population. If the treatment is randomized, then $T_i$ is statistically independent of potential outcomes; formally, we write $Y_i^{(1)}, Y_i^{(0)} \perp T_i$. When this is true, the average causal effect can be estimated by the observed mean difference between the treatment and control groups, $E(Y_i^{(1)} - Y_i^{(0)}) = E(Y_i^{(1)} \mid T_i = 1) - E(Y_i^{(0)} \mid T_i = 0)$, which is the familiar result that the difference- in-means estimator is unbiased for the average causal effect in randomized experiments ( Imai, 2010).

## 2.3 The Extension of Causal Effects to Causal Mediation Effects

In this section, the counterfactual framework of causal inference, which is widely used in the statistical literature, is explained for the mediation analysis. Although there are many situations where the treatment, mediator and the response can be either binary or continuous, we use a dichotomous treatment variable ($T = 0,1$), a continuous mediator ($Z$), and a continuous response ($Y$) throughout this study.

Most of the published research in causal inference focuses on relationships between two variables, $T$ (Treatment) and $Y$ (response), and much has been written about two-variable relations, including conditions under which $T$ can be considered a possible cause of $Y$. These conditions include randomization of units to values of $T$ and independence of units across and within values of $T$. Mediation represents the addition of a third variable to this $T \rightarrow Y$ relation as shown in Figure 2.1.

The counterfactual framework and notation have been extended to define causal mediation effects (Albert, 2008). We use $Z_i$ to denote the observed level of a mediating variable. There exists two potential values, $Z_i^{(0)}$ and $Z_i^{(1)}$, only one of which will be observed. In causal inference, the potential outcomes were only a function of the treatment, but in causal mediation analysis the potential outcomes depend on the mediators as well as the treatment variable. Therefore, we use $Y_i(T = t,\ Z = z(t)) = Y^{(t)}(Z^{(t)})$ to denote the potential outcome that would result if the treatment and mediating variables equal $T = t,\ Z = z(t)$ respectively. The observable potential outcome is either $Y^{(1)}(Z^{(1)})$ or $Y^{(0)}(Z^{(0)})$ which would be realized under treatment or control conditions, respectively. There are two unobservable potential outcomes, $Y^{(0)}(Z^{(1)})$, which refer to the potential outcome for individual *i* who is assigned to the control condition but takes on a value of the mediator that would be realized under the treatment condition and $Y^{(1)}(Z^{(0)})$, which refer to the potential outcome for individual *i* who is assigned to the treatment condition but takes on a value of the mediator that would be realized under the control condition.

We now define the individual level causal mediation effect as $Y^{(t)}(Z^{(1)}) - Y^{(t)}(Z^{(0)})$ for $t = 0,1$. Thus, the causal mediation effect represents the indirect effect of the treatment on the outcome through the mediating variable. The average causal mediation

effect (ACME) can be defined as $E[(Y_i^{(t)}(Z_i^{(1)}) - Y_i^{(t)}(Z_i^{(0)})]$ for $t = 0,1$ (Imai K, 2010 ). The potential outcomes framework makes it clear that this quantity involves counterfactual outcomes that can never be observed such as $Y^{(1)}(Z^{(0)})$.

Some assumptions are needed when estimating mediation effects.

(1) Stable Unit Treatment Value Assumption (SUTVA)

A common assumption for causal inference is Rubin's stable unit treatment value assumption (SUTVA), which includes two sub-assumptions: a) "non-interference" which means that one unit's intervention will not affect other units potential outcomes, and b) "no treatment variation" which means there are not multiple versions of the same treatment. Thus, there is only one set of potential outcomes for each unit (Rubin, 2005).

(2) Sequential Ignorability

The second assumption needed is the randomization assumption (or ignorability assumption in an observational study). This assumption means that the observed intervention assignment is independent of the potential mediator defined by different levels of intervention and all potential outcomes defined by different levels of intervention and mediator. The ignorability version of this assumption assumes conditional independence given baseline covariates (*X*).

Based on the potential outcomes we describe the total, direct and indirect effects and their relationships.

## 2.4 Total, Direct and Indirect Effects

Robins and Greenland (1992) used the terminology "pure" and "total" for direct and indirect effects because there are different ways of decomposing an overall effect into direct and indirect effect components. Albert (2008) used the concepts of a potential outcome framework to develop a causal or manipulation model framework for mediation analysis. Using this framework, he provided new definitions and measures of mediation for multiple treatments and mediators based on the concepts of "pure" and "total" effects though he did not specifically use those terms. Effects of manipulations are modeled via the linear structural model. The methods

were applied to data from a study of nursing interventions for postoperative pain. They have discussed the cases of more than two treatment groups, and an interaction among mediators. But, they have not considered the interactions between treatment and the mediator or individual and the mediator.

It is challenging to understand the concepts behind the relationship among direct, indirect and total effects without a valid decomposition technique. My research being presented here was helped by a paper published by VanderWeele (2011) who defined the total, direct and indirect effects and decomposed total effect into direct and indirect effect and also identified the important characteristics of those effects. In our usual notation, total effect of $T$ on $Y$ can be defined as $TE = Y^{(1)}(Z^{(1)}) - Y^{(0)}(Z^{(0)})$. Direct effects: that part of the exposure effect which is not mediated by a given set of potential mediators.

The natural direct effects can be defined as

Total Direct Effect, $TDE = Y^{(1)}(Z^{(1)}) - Y^{(0)}(Z^{(1)})$

Pure Direct Effect, $PDE = Y^{(1)}(Z^{(0)}) - Y^{(0)}(Z^{(0)})$

Indirect / mediated effects: that part of the exposure effect which is mediated by a given set of potential mediators.

The natural indirect effects can be defined as

Total Indirect Effect, $TIE = Y^{(1)}(Z^{(1)}) - Y^{(1)}(Z^{(0)})$

Pure Indirect Effect, $PIE = Y^{(0)}(Z^{(1)}) - Y^{(0)}(Z^{(0)})$.

If there is no interaction between the treatment and the mediator then we can show that

$TDE = PDE$ and $TIE = PIE$ whereas if there is an interaction then

$TDE \neq PDE$ and $TIE \neq PIE$.

Decomposing the total effect into indirect effects and the direct effects gives,

$$\begin{aligned} TE = Y^{(1)}(Z^{(1)}) - Y^{(0)}(Z^{(0)}) &= Y^{(1)}(Z^{(1)}) - Y^{(1)}(Z^{(0)}) + Y^{(1)}(Z^{(0)}) + Y^{(0)}(Z^{(0)}) \\ &= \left[Y^{(1)}(Z^{(1)}) - Y^{(1)}(Z^{(0)})\right] + \left[Y^{(1)}(Z^{(0)}) + Y^{(0)}(Z^{(0)})\right] \\ &= TIE + PDE \end{aligned}$$

Alternatively,

$$\begin{aligned} TE = Y^{(1)}(Z^{(1)}) - Y^{(0)}(Z^{(0)}) &= Y^{(1)}(Z^{(1)}) - Y^{(0)}(Z^{(1)}) + Y^{(0)}(Z^{(1)}) + Y^{(0)}(Z^{(0)}) \\ &= \left[Y^{(1)}(Z^{(1)}) - Y^{(0)}(Z^{(1)})\right] + \left[Y^{(0)}(Z^{(1)}) + Y^{(0)}(Z^{(0)})\right] \\ &= TDE + PIE \end{aligned}$$

Regardless of whether a $T \times Z$ interaction term is present or not,

$TE = TIE + PDE = TDE + PIE$.

## 2.5 Mediation Plot

The mediation plot is used to present diagrams which show the structural relationship between the independent, mediator, and dependent variables. Here we discuss how to construct plots of the mediated effect for a study with a dichotomous treatment variable (e.g., random assignment to a treatment and control group), a continuous mediator, and a continuous outcome variable.

The data are plotted with $Y$ on the vertical axis and $Z$ on the horizontal axis, as shown in Figure 2.2. Next, Equation (1) is plotted for the values of $T$ (i.e., 0 and 1), such that a horizontal line (black colored) is placed at $Y = c_Y$, corresponding to $T = 0$, and a second horizontal line is placed at $Y = c_Y + \gamma$, corresponding to $T = 1$. The distance between the horizontal lines represents the total effect of $T$ on $Y$, $\gamma$. Then, Equation (2) is plotted for both values of $T$, resulting in a vertical line (blue colored) at $Z = c_Z$ and a second vertical line at $Z = c_Z + \alpha$. The distance between the two vertical lines represents the treatment effect on the mediator and is equal to $\alpha$. Finally, Equation (3) is plotted at both values of $T$. The slopes of these parallel simple regression lines (red colored) are equal to $\beta$. Thus, the mean indirect effect can be written as $\alpha\beta = \beta(\mu_{Z^{(1)}} - \mu_{Z^{(0)}})$. Plots of the mediated effect may be useful to investigate the distributions of data for outliers and to improve understanding of relations among variables in the mediation model (Fritz et al, 2010). Fritz et. al (2010) had discussed how to draw a mediation plot when the treatment variable is continuous and when the interaction term is present.

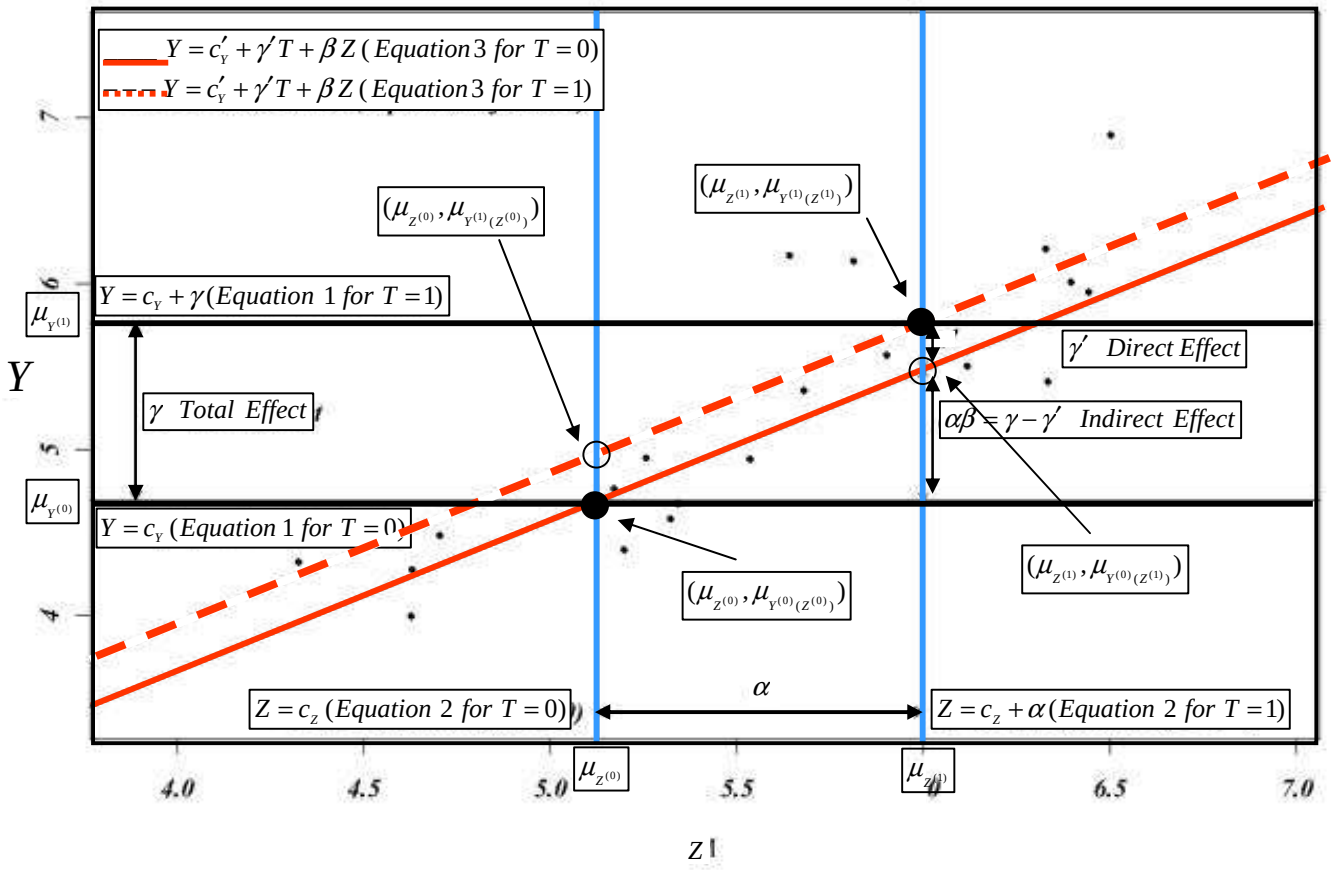



**Figure 2.1 Plot of the mediating effect for a dichotomous treatment variable. The open circles in the plot represent means for which no data are available to directly estimate, and the closed circles do have such data observable.**

## 2.6 Individual Mediating Effects

The published research on heterogeneity of mediating effects seems rare and we did not find any that discuss the concepts in this dissertation. Before we discuss heterogeneity of mediating effects, the past work done on the heterogeneity of treatment effects should be recalled. Gadbury, Allison, and Albert have past work focused on defining true individual effects of a treatment an estimating quantities describing the variability of a treatment's effect across individuals in a population.

### *2.6.1 Treatment Heterogeneity*

When we compare a treatment, $T=1$, and a control, $T=0$, under a two sample completely randomized design, the value of an outcome variable, $Y$, is measured at a particular point in time. This outcome may be quantitative or dichotomous (i.e., success or failure) and assume that the outcome variable s a quantitative one. For a subject receiving treatment T = 1, $Y^{(1)}$ is observable, and for a subject receiving control, $T=0$, the outcome $Y^{(0)}$ is observable. At a particular time, only one of two potential outcomes (Rubin, 1974) in the bivariate pair $(Y^{(1)}, Y^{(0)})$ is observable because of the "fundamental problem of causal inference"(Holland, 1986).

Figure 2.3 shows the potential outcomes for $N$ subjects in a study and the observed outcomes post treatment assignment.

$$\begin{pmatrix} Y_1^{(1)} & Y_1^{(0)} \\ Y_2^{(1)} & Y_2^{(0)} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ Y_N^{(1)} & Y_N^{(0)} \end{pmatrix} \xrightarrow{\text{Treatment Assignment}} \begin{pmatrix} Y_1^{(1)} & ? \\ ? & Y_2^{(0)} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ ? & Y_{N-1}^{(0)} \\ Y_N^{(1)} & ? \end{pmatrix}$$

**Figure 2.1 Potential outcomes and the observed outcomes for $N$ subjects for two sample completely randomized design.**

The set of $N$ potential outcomes has the form given in Figure 2.3 (left) that, after treatment assignment produces observed outcomes of the form shown (right), and where the "?" represents an unobserved potential outcomes.

The true individual treatment effect is defined as $D = Y^{(1)} - Y^{(0)}$ which cannot be observed for any subject. If evaluating $k$ treatment levels, the potential outcomes would be a vector containing $k$ outcomes (rather than two), and only one of the $k$ outcomes would be observable for a given subject at a particular time. It is the average treatment effect, $E(D)$, which is usually of interest. But, we are interested in the variance, $Var(D)$, that quantifies the degree of variability of individual treatment effects.

If the $\sigma_D^2 = Var(D) > 0$ then the individual treatment heterogeneity is present Gadbury (2001). This is also called Subject-Treatment Interaction which cannot be directly estimated because $\sigma_D^2 = \sigma_{Y^{(1)}}^2 + \sigma_{Y^{(0)}}^2 - 2\sigma_{Y^{(1)}}\sigma_{Y^{(0)}}\rho_{Y^{(1)}Y^{(0)}}$ and there is no available information about $\rho_{Y^{(1)}Y^{(0)}}$ in observable data; that is, it is nonestimable.

Though $\sigma_D^2$ cannot be directly estimated using observed data, bounds for it can be derived, and these can be estimated. Observed pretreatment moderating variables can be used to tighten bounds or to impute missing potential outcomes, thereby providing naïve estimates of individual effects. Gadbury and Iyer (2000) and Gadbury et al., (2001) considered a two sample completely randomized design with covariates, Gadbury et al., (2004) considered a matched-pairs design, and Albert et al. (2005) used a blocked design to estimate or bound individual treatment heterogeneity parameters in randomized trials. Gadbury (2010) also presented the framework and some initial results for a two period cross-over design where, conceptually, there were two individual treatment effect variables, $D$ as defined above, one for each time period. Poulson et al., (2012) added results comparing individual treatment heterogeneity to subset interaction, cross-over designs, and overlap of marginal density curves.

In chapter 3, we extend the concepts of treatment heterogeneity to mediating heterogeneity when the intermediate variable is present in the causal relation. This is a main objective of my research. There is another aspect of mediation that was also considered as a part of this work.

## 2.7 Terminal Measures

Situations can arise where $Z$ and $Y$ cannot both be measured on an individual unit. Mouse and plant experiments are two examples where measurement of $Z$ requires terminating the animal or plant and $Y$ is to be measured at a later time. We refer to such data as terminal measures data. Another situation may be where one experiment focused on measurement of $Z$, and the second on measurement of $Y$, and interest is in combining the data sets to evaluate the mediating effect of $Z$ on $Y$.

In chapter 4, the Terminal Measures are described in detail and the methods of handling Terminal Measures are discussed further. Then the simulation study discusses the issues such as pseudoreplication involved in analyzing terminal measures data. The term *pseudoreplication* was

coined by Hurlbert (1984, 190-191) to refer to "the use of inferential statistics to test for treatment effects with data from experiments where either treatments are not replicated (though samples may be) or replicates are not statistically independent." The context of his paper was ecological field experiments, but pseudoreplication can occur in other contexts as well.

When we search the term “terminal measures” or “terminal data” in mediation analysis there was only one paper published by Makowsky et al (2011). They have mentioned “In psychiatric research, the costs of measuring the putative mediator or the outcome can be prohibitive. The illustrated some extreme sampling designs as methods for reducing study costs by increasing power per subject measured on the more expensive variable when assessing bivariate relationships. However, these designs did require that the mediator and outcome variable be jointly observable on at least some subjects in the study. In terminal measures data discussed herein, the mediator and outcome variable are never observable on the same unit in the study.

# Chapter 3 - Defining Heterogeneity of Individual Indirect Effects

Although both individual and average causal effects are defined in the potential outcomes framework, most papers focus on learning about the average causal effects. As mentioned in the literature review, Gadbury, Allison, and Albert have past work focused on defining true individual effects of a treatment and estimating quantities describing the variability of a treatment's effect across individuals in a population. In 2005, Steyer analyzed individual and average causal effects via structural equation models.

But when we review the literature on individual causal mediating effects, although both individual and average causal mediation effects are defined in Rubin's framework to causality, almost no efforts deal with developing designs and models to learn about variability of individual mediating effects. In fact, it is not clear how to conceptualize heterogeneity in individual mediating effects.

This study takes a first step in this direction. In the first and general part, Gadbury et al 's concept of individual causal effects is extended, replacing causal effects by causal mediating effects. Based on this extension, in the second and main part, the designs, assumptions and models are introduced which can allow identification of (1) the variance of the individual causal mediating effects, (2) the nonestimable quantities in the variance formula and (3) in some cases the bounds for the variance of the individual causal mediating effects.

According to Gadbury et al. (2010), that the effect of a treatment will vary among subjects is not surprising, nor is it a recent concept. Subject-treatment (S-T) interaction is, as the term implies, an interaction of specific subjects with applied treatment(s). The result of such interaction is a variability of "individual treatment effects" or "individual treatment heterogeneity" in a population of interest. Similarly, we can describe subject-mediator (treatment) (S-M(T)) interaction as heterogeneity of mediated ( or indirect ) effects of a treatment across individuals in a population.

Before we discuss the S-M(T) interaction, an important relationship between the mediating effects obtained by the Baron and Kenny approach and the individual indirect effects based on potential outcomes should be derived.

## 3.1 Individual Mediation Plot

The individual mediation plot is used to present the structural relationship between the independent, mediator, and dependent variables for an individual in a potential outcomes framework. Again we consider a dichotomous treatment variable (e.g., random assignment to a treatment and control group), a continuous mediator, and a continuous outcome variable.

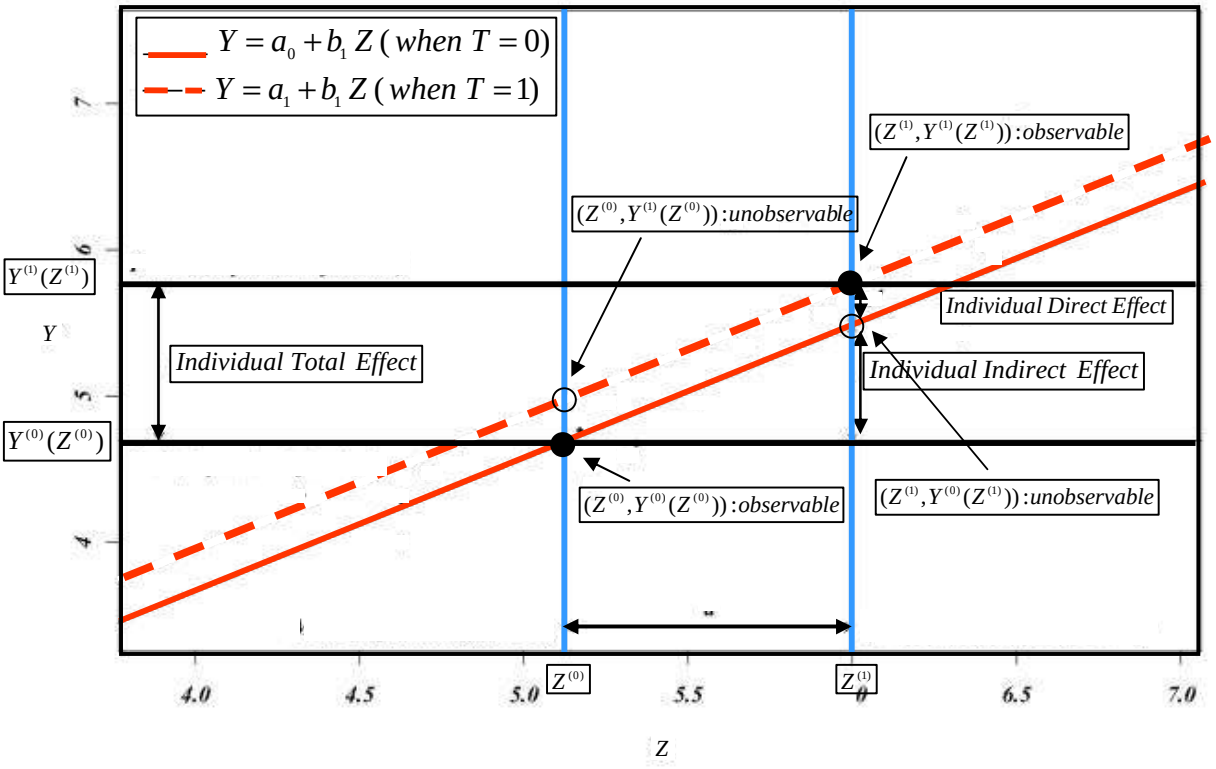



**Figure 3.1 Individual Mediation Plot.**

Figure 3.1 shows the mediation plot for an individual $i$ with the independent variable, $T$, the mediator $Z$ on the horizontal axis, and the dependent variable $Y$ on the vertical axis. The four points are the potential outcomes and two of them are potentially observable and the other two will never be observed for an individual. Among the observable outcomes, only one is actually observed at a given time for an individual. Each individual has a particular slope and intercept as shown in the plot. When we consider the individual plots, the slopes and the intercepts are considered to be random variables with a probability distribution.

## 3.2 The Relationship Between the Average Mediating Effects Obtained by the Baron & Kenny Approach and the Average Indirect Effect defined for Potential Outcomes

Now consider a structural relationship between the independent, mediator, and dependent variables for an individual in potential outcomes framework. For an individual *i*

$Y_i^{(1)}(Z_i^{(1)}) = A_{1i} + B_i Z_i^{(1)}$ where $A_{1i} \sim (c_Y^{/} + \gamma', \sigma_{A_1}^2)$, $B_i \sim (\beta, \sigma_B^2)$ and $A_{1i} \perp B_i$

$Y_i^{(0)}(Z_i^{(0)}) = A_{0i} + B_i Z_i^{(0)}$ where $A_{0i} \sim (c_Y^{/}, \sigma_{A_0}^2)$, $B_i \sim (\beta, \sigma_B^2)$ and $A_{0i} \perp B_i$

$Y_i^{(1)}(Z_i^{(0)}) = A_{1i} + B_i Z_i^{(0)}$ where $A_{1i} \sim (c_Y^{/} + \gamma', \sigma_{A_1}^2)$, $B_i \sim (\beta, \sigma_B^2)$ and $A_{1i} \perp B_i$

$Y_i^{(0)}(Z_i^{(1)}) = A_{0i} + B_i Z_i^{(1)}$ where $A_{0i} \sim (c_Y^{/}, \sigma_{A_0}^2)$, $B_i \sim (\beta, \sigma_B^2)$ and $A_{0i} \perp B_i$

Based on this set up, the averages of individual potential outcomes can be derived.

$$E(Y_i^{(1)}(Z_i^{(1)})) = E(Y_i^{(1)} \mid Z_i^{(1)}) = E(A_{1i}) + E(B_i Z_i^{(1)} \mid Z_i^{(1)}) = (c_Y' + \gamma') + \beta Z_i^{(1)}$$

$$E(Y_i^{(1)}(Z_i^{(0)})) = E(Y_i^{(1)} \mid Z_i^{(0)}) = E(A_{1i}) + E(B_i Z_i^{(0)} \mid Z_i^{(0)}) = (c_Y' + \gamma') + \beta Z_i^{(0)}$$

$$E(Y_i^{(0)}(Z_i^{(0)})) = E(Y_i^{(0)} \mid Z_i^{(0)}) = E(A_{0i}) + E(B_i Z_i^{(0)} \mid Z_i^{(0)}) = c_Y' + \beta Z_i^{(0)}$$

$$E(Y_i^{(0)}(Z_i^{(1)})) = E(Y_i^{(0)} \mid Z_i^{(1)}) = E(A_{0i}) + E(B_i Z_i^{(1)} \mid Z_i^{(1)}) = c_Y' + \beta Z_i^{(1)}$$

Note that the above expectations were taken conditional on $Z$ as is typically done in a regression setting. In doing this, the above four equations represent the expectation of individual potential outcomes that are then the same as the equations obtained from equation (3) in Chapter 2 for the Baron and Kenny approach under the treatment conditions, $T = 0, T = 1$, and where there expectations were taken with respect to the distribution of a random error term in a regression model. Thus the approach used here agrees with past approaches as far as means are concerned. The above framework, however, allows flexibility in modeling sources of heterogeneity in mediating effects, namely because $Z$ is not a covariate but an outcome affected by treatment.

For each individual, the relationship among the individual total effect (*ITE)*, individual indirect effect (*IIE)* and the individual direct effect (*IDE)* can be written in terms of potential outcomes as follows:

$$\begin{aligned} ITE = Y^{(1)}(Z^{(1)}) - Y^{(0)}(Z^{(0)}) &= Y^{(1)}(Z^{(1)}) - Y^{(1)}(Z^{(0)}) + Y^{(1)}(Z^{(0)}) - Y^{(0)}(Z^{(0)}) \\ &= \left[Y^{(1)}(Z^{(1)}) - Y^{(1)}(Z^{(0)})\right] + \left[Y^{(1)}(Z^{(0)}) - Y^{(0)}(Z^{(0)})\right] \\ &= B(Z^{(1)} - Z^{(0)}) + (A_1 - A_0) \\ &= B(Z^{(1)} - Z^{(0)}) + G', \\ &\qquad \text{where } B = \frac{Y^{(1)}(Z^{(1)}) - Y^{(1)}(Z^{(0)})}{Z^{(1)} - Z^{(0)}} \text{ and } G' = A_1 - A_0. \\ &= IIE + IDE \end{aligned}$$

where $B$ is the "slope" random variable and the $G'$ is the difference between the two intercepts, which is the individual direct effect. The individual indirect effect is $B(Z^{(1)} - Z^{(0)})$ and the mean of an individual indirect effect would be $\beta(\mu_{Z^{(1)}} - \mu_{Z^{(0)}})$, which is the effect shown in Figure 2.2 that illustrates a mean effect.

## 3.3 Plots for Illustrating the Variability of Total Effect of Individuals

As in section 3.2, in terms of potential outcomes the Individual Total Effect (*ITE)* can be decomposed into the Individual Indirect Effect (*IIE*) and the Individual Direct Effect (*IDE*). Thus, the mediating effect for an individual depends on the slope $(B)$, the intercepts ($G'$=the difference between the intercepts) of its regression lines and the treatment effect on the mediator, $Z^{(1)} - Z^{(0)}$. We investigate how the mediation plots can behave for different individuals as follows: Out of four quantities we fix two and vary one, and then observe the behavior of the other. Next we find the functional relationship between what we vary and observe. The following table gives all the cases to be considered and the roll of each parameter.

As we are interested in observing the variation of total effect we focus only on the cases (1), (5) and (6). Table 3.2 shows hypothetical numerical examples of calculating total effects, indirect effects and direct effects for each case.

**Table 3.1 All cases for mediation plot when two parameters were fixed at a time**

| | **Cases** | | | | | |
|---|---|---|---|---|---|---|
| **Parameters** | **1** | 2 | 3 | 4 | **5** | **6** |
| Slopes | fix | vary | vary | fix | vary | fix |
| Intercepts | fix | fix | vary | vary | fix | vary |
| Total Effect | vary | fix | fix | fix | vary | vary |
| $Z^{(1)} - Z^{(0)}$ | vary | vary | fix | vary | fix | fix |

**Table 3.2 Hypothetical values for individual mediation plots for two individuals.**

| | $A_0$ | $A_1$ | $B_1$ | $Z^{(1)} - Z^{(0)}$ | *TE* | *IE* | *DE* |
|---|---|---|---|---|---|---|---|
| Case 1 | **1** | **2** | **1** | 8-6 = 2 | 3 | 2 | 1 |
| | **1** | **2** | **1** | 9-6 = 3 | 4 | 3 | 1 |
| Case 5 | **1** | **2** | 0.5 | **8-6 = 2** | 2 | 1 | 1 |
| | **1** | **2** | 1 | **8-6 = 2** | 3 | 2 | 1 |
| Case 6 | 1 | 2 | **1** | **8-6 = 2** | 3 | 2 | 1 |
| | -14 | -18 | **1** | **8-6 = 2** | -2 | 2 | -4 |

In the Figure 3.2, two black colored horizontal lines represent the total effect, two blue colored vertical lines represent the treatment effect on the mediator, $Z^{(1)} - Z^{(0)}$ and the two sloped red lines represent the linear relationship between *Y* and *Z* under the treatment assignment and the four blue colored dots are the four potential outcomes. Each row in Figure 3.2 illustrates cases 1, 5, and 6, respectively. The two columns represent effects for two individuals.

In case (1); Since

$$ITE = Y^{(1)}(Z^{(1)}) - Y^{(0)}(Z^{(0)}) = IIE + IDE = slope * (Z^{(1)} - Z^{(0)}) + (difference\ of\ \text{int}\, ercepts),$$

the total effect is a function of $Z^{(1)} - Z^{(0)}$, the treatment effect on the mediator because intercepts are fixed (that is direct effect is fixed ). So variability in the total effects is a due to variability in indirect effects, $slope * (Z^{(1)} - Z^{(0)})$, resulting from variance in $Z^{(1)} - Z^{(0)}$. In case (5), variability in total effects is due to variability in slopes as intercepts are fixed (that is direct effect

is fixed) and $Z^{(1)} - Z^{(0)}$ is fixed. In case (6), variability in total effects is due to variability in direct effects.

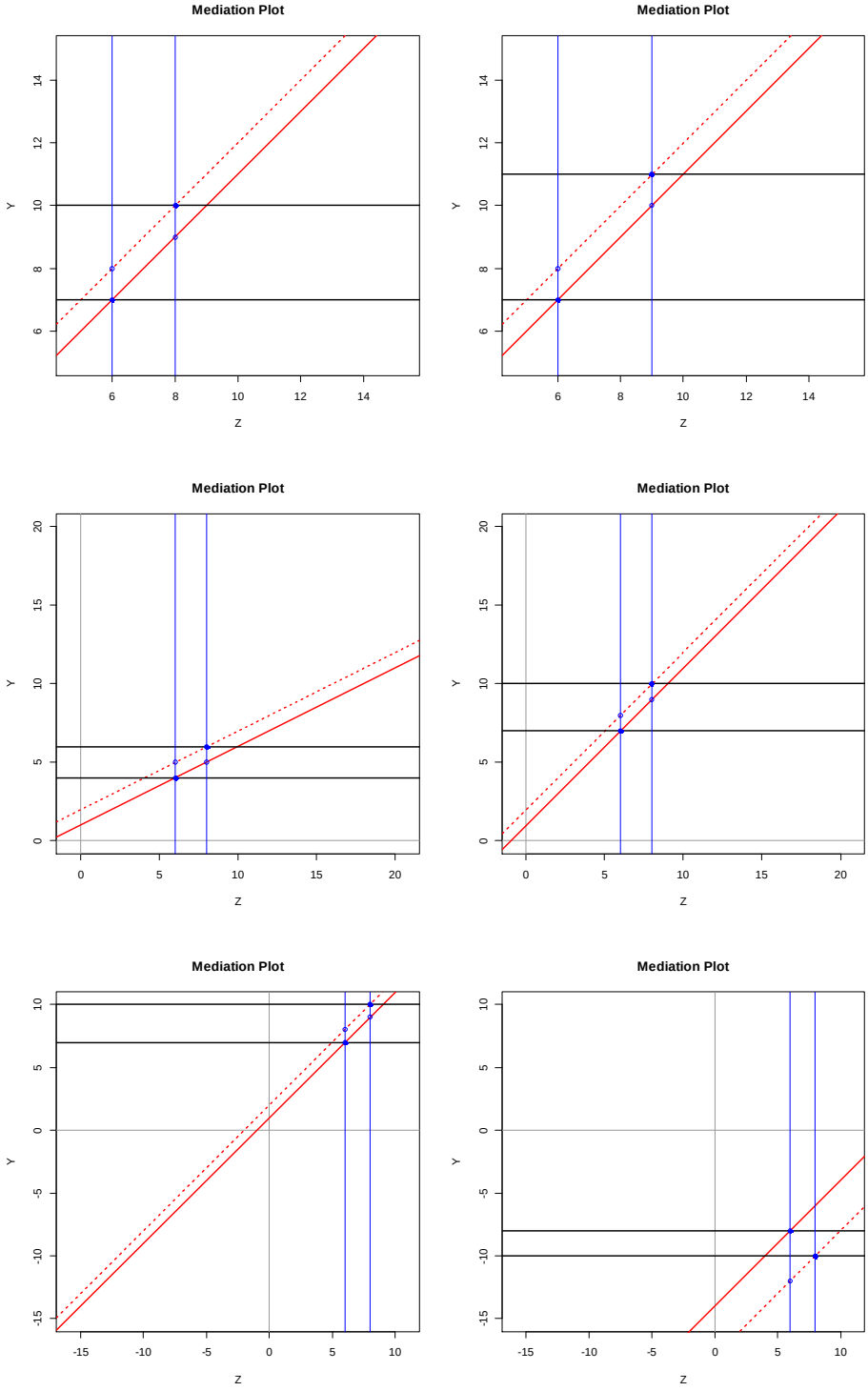



**Figure 3.2 Mediation plots when the total effect was observed ( Ref: Table 3.2).**

Each row illustrates cases 1, 5, and 6, respectively. The two columns represent effects for two individuals.

Next we investigate how the mediation plots behave for different situations when three quantities were varied and one was fixed. The following table gives all the cases to be considered and the roll of each quantity.

**Table 3.3 All cases for a mediation plot when one parameter was fixed at a time.**

| | **Parameters** | | | |
|---|---|---|---|---|
| **Cases** | Slope | Intercept | Total Effect | $Z^{(1)} - Z^{(0)}$ |
| Case1 | fix | vary | vary | vary |
| Case2 | vary | fix | vary | vary |
| Case3 | vary | vary | vary | fix |

The Figure 3.3 shows the variability of total effect for a different individual with different slopes and intercepts in the regression model. Interpreting the plots: the larger the distance between the horizontal lines, the larger the overall effect of $T$ on $Y$, the amount of change in $Y$ for a one unit change in $T$. The larger the distance between the vertical lines, the greater the effect of $T$ on $Z$. The steeper the slope of the regression lines, the larger the effect, of $Z$ on $Y$, adjusting for $T$. Again, each row of figure 3.3 illustrates cases 1, 2, and 3, respectively. Each column represents the individual effects for a particular subject.

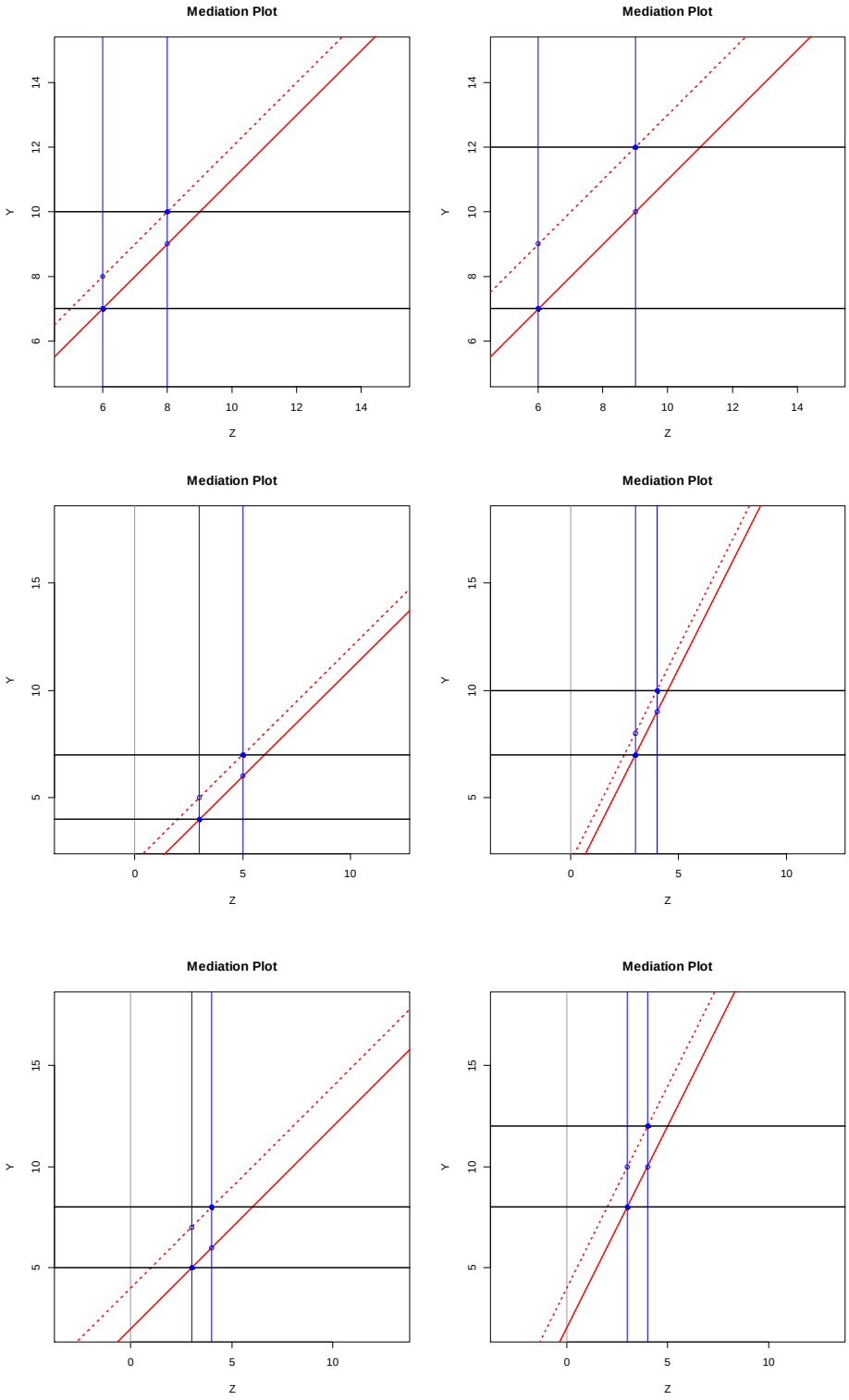



**Figure 3.3 Mediation plots when the total effect was observed ( Ref: Table 3.3).**

Each row illustrates cases 1, 2, and 3, respectively. Each column represents the individual effects for a particular subject.

## 3.4 The Distribution of Potential Outcomes and the Indirect Effects for an Individual

The prior section used tables and figures to illustrate sources of variability in total effects, direct effects, and indirect effects. The distribution of individual indirect effects will be assumed to be of the form $B(Z^{(1)} - Z^{(0)}) \sim (\beta(\mu_{Z^{(1)}} - \mu_{Z^{(0)}}), \sigma^2_{IIE})$, where $E(B) = \beta$ is the population mean of slopes that was used as the slopes of the mediation plot in chapter 2 and $\sigma^2_{IIE}$ is the variance of individual mediating effects which are of interest in this study and will be discussed in detail in a later section.

Recall the following equations in section 3.2 for an individual *i*,

$Y_i^{(1)}(Z_i^{(1)}) = A_{1i} + B_i Z_i^{(1)}$ where $A_{1i} \sim (c_Y' + \gamma', \sigma^2_{A_1})$, $B_i \sim (\beta, \sigma^2_B)$ and $A_{1i} \perp B_i$

$Y_i^{(0)}(Z_i^{(0)}) = A_{0i} + B_i Z_i^{(0)}$ where $A_{0i} \sim (c_Y', \sigma^2_{A_0})$, $B_i \sim (\beta, \sigma^2_B)$ and $A_{0i} \perp B_i$

$Y_i^{(1)}(Z_i^{(0)}) = A_{1i} + B_i Z_i^{(0)}$ where $A_{1i} \sim (c_Y' + \gamma', \sigma^2_{A_1})$, $B_i \sim (\beta, \sigma^2_B)$ and $A_{1i} \perp B_i$

$Y_i^{(0)}(Z_i^{(1)}) = A_{0i} + B_i Z_i^{(1)}$ where $A_{0i} \sim (c_Y', \sigma^2_{A_0})$, $B_i \sim (\beta, \sigma^2_B)$ and $A_{0i} \perp B_i$

The expected values were derived in 3.2 and now the variances of individual potential outcomes are derived. The variances of these individual potential outcomes as follows:

$$\begin{aligned} Var(Y_i^{(1)}(Z_i^{(1)})) &= Var(A_{1i}) + E(Var(B_i Z_i^{(1)} | Z_i^{(1)})) + Var(E(B_i Z_i^{(1)} | Z_i^{(1)})) \\ &= \sigma^2_{A_1} + \sigma^2_B E((Z_i^{(1)})^2) + \beta^2 Var(Z_i^{(1)}) \\ &= \sigma^2_{A_1} + \sigma^2_B [Var(Z_i^{(1)}) + (E(Z_i^{(1)}))^2] + \beta^2 Var(Z_i^{(1)}) \\ &= \sigma^2_{A_1} + \sigma^2_B [\sigma^2_{Z^{(1)}} + \mu^2_{Z^{(1)}}] + \beta^2 \sigma^2_{Z^{(1)}}. \end{aligned}$$

$$\begin{aligned} Var(Y_i^{(1)}(Z_i^{(0)})) &= Var(A_{1i}) + E(Var(B_i Z_i^{(0)} | Z_i^{(0)})) + Var(E(B_i Z_i^{(0)} | Z_i^{(0)})) \\ &= \sigma^2_{A_1} + \sigma^2_B E((Z_i^{(0)})^2) + \beta^2 Var(Z_i^{(0)}) \\ &= \sigma^2_{A_1} + \sigma^2_B [Var(Z_i^{(0)}) + (E(Z_i^{(0)}))^2] + \beta^2 Var(Z_i^{(0)}) \\ &= \sigma^2_{A_1} + \sigma^2_B [\sigma^2_{Z^{(0)}} + \mu^2_{Z^{(0)}}] + \beta^2 \sigma^2_{Z^{(0)}}. \end{aligned}$$

$$
\begin{aligned}
Var(Y_i^{(0)}(Z_i^{(0)})) &= Var(A_{0i}) + E(Var(B_i Z_i^{(0)} \big| Z_i^{(0)})) + Var(E(B_i Z_i^{(0)} \big| Z_i^{(0)})) \\
&= \sigma_{A_0}^2 + \sigma_B^2 E((Z_i^{(0)})^2) + \beta^2 Var(Z_i^{(0)}) \\
&= \sigma_{A_0}^2 + \sigma_B^2 [Var(Z_i^{(0)}) + (E(Z_i^{(0)}))^2] + \beta^2 Var(Z_i^{(0)}) \\
&= \sigma_{A_0}^2 + \sigma_B^2 [\sigma_{Z^{(0)}}^2 + \mu_{Z^{(0)}}^2] + \beta^2 \sigma_{Z^{(0)}}^2 .
\end{aligned}
$$

$$
\begin{aligned}
Var(Y_i^{(0)}(Z_i^{(1)})) &= Var(A_{0i}) + E(Var(B_i Z_i^{(1)} \big| Z_i^{(1)})) + Var(E(B_i Z_i^{(1)} \big| Z_i^{(1)})) \\
&= \sigma_{A_0}^2 + \sigma_B^2 E((Z_i^{(1)})^2) + \beta^2 Var(Z_i^{(1)}) \\
&= \sigma_{A_0}^2 + \sigma_B^2 [Var(Z_i^{(1)}) + (E(Z_i^{(1)}))^2] + \beta^2 Var(Z_i^{(1)}) \\
&= \sigma_{A_0}^2 + \sigma_B^2 [\sigma_{Z^{(1)}}^2 + \mu_{Z^{(1)}}^2] + \beta^2 \sigma_{Z^{(1)}}^2 .
\end{aligned}
$$

In this study we are interested in estimating the variance of the individual indirect effects, $\sigma_{IIE}^2 = Var(B(Z^{(1)} - Z^{(0)}))$.

## 3.5 Individual Mediating Effects and Mediating Heterogeneity

In this section, the individual mediating effect, the variability of individual mediating effect and subject-mediator (treatment) interaction are introduced. Then, we discuss the problems we face when estimating the variability of individual mediating effects.

### *3.5.1 Defining Subject-Mediator (Treatment) Interaction*

Let $\left(Z^{(1)}, Z^{(0)}, Y^{(1)}(Z^{(1)}), Y^{(0)}(Z^{(0)})\right)$ be a set of potentially observable potential outcomes for an individual subject in an investigation to compare the effect of treatment $T^{(1)}$ with respect to a control treatment $T^{(0)}$. The four values are imagined to be measured at the same moment of time. But, in practice only the values corresponding to the treatment *t*, $Z^{(t)}$ and $Y^{(t)}(Z^{(t)})$, actually assigned can be observed for a particular subject. The two potential outcomes for the mediator help to conceptualize a true treatment effect on the mediator for a subject that we define to be $D_Z = Z^{(1)} - Z^{(0)}$. The expectation and the variance of $D_Z$ are given by the formula $E(D_Z) = \mu_{D_Z} = \mu_{Z1} - \mu_{Z0}$ and $Var(D_Z) = \sigma_{D_Z}^2 = \sigma_{Z^{(1)}}^2 + \sigma_{Z^{(0)}}^2 - 2\rho_{Z^{(1)}Z^{(0)}}\sigma_{Z^{(1)}}\sigma_{Z^{(0)}}$. Estimating $\mu_{D_Z}$ is straightforward in common randomized experiments. But are interest is in estimating $\sigma_{D_Z}^2$. There is a subject-treatment interaction on the mediating variable if $\sigma_{D_Z}^2 > 0$. The subject-mediator

(treatment) interaction is a combination of subject-treatment interaction on $(Z^{(1)} - Z^{(0)})$ and the subject-treatment interaction on $(Y^{(1)} - Y^{(0)})$. In particular, we consider subject-mediator (treatment) interaction to be present when $\sigma^2_{IIE} = Var(B(Z^{(1)} - Z^{(0)})) > 0$.

Figure 3.4 shows the potential mediators and potential outcomes for $N$ subjects in a study and the observed outcomes post treatment assignment. The set of $N$ potential mediators and the potential outcomes has the form given at top that, after treatment assignment produces observed mediators and the observed outcomes of the form shown at bottom, and where the "?" represents an unobservable potential outcome at a particular time.

$$\begin{pmatrix} Z_1^{(1)} & Z_1^{(0)} & Y_1^{(1)}(Z_1^{(1)}) & Y_1^{(0)}(Z_1^{(1)}) & Y_1^{(1)}(Z_1^{(0)}) & Y_1^{(0)}(Z_1^{(0)}) \\ Z_2^{(1)} & Z_2^{(0)} & Y_2^{(1)}(Z_2^{(1)}) & Y_2^{(0)}(Z_2^{(1)}) & Y_2^{(1)}(Z_2^{(0)}) & Y_2^{(0)}(Z_2^{(0)}) \\ . & . & & & & \\ . & . & & & & \\ . & . & & & & \\ Z_N^{(1)} & Z_N^{(0)} & Y_N^{(1)}(Z_N^{(1)}) & Y_N^{(0)}(Z_N^{(1)}) & Y_N^{(1)}(Z_N^{(0)}) & Y_N^{(0)}(Z_N^{(0)}) \end{pmatrix}$$

Treatment Assignment

$$\begin{pmatrix} Z_1^{(1)} & ? & Y_1^{(1)}(Z_1^{(1)}) & ? & ? & ? \\ ? & Z_2^{(0)} & ? & ? & ? & Y_2^{(0)}(Z_2^{(0)}) \\ . & . & & & & \\ . & . & & & & \\ . & . & & & & \\ ? & Z_{N-1}^{(0)} & ? & ? & ? & Y_{N-1}^{(0)}(Z_{N-1}^{(0)}). \\ Z_N^{(1)} & ? & Y_N^{(1)}(Z_N^{(1)}) & ? & ? & ? \end{pmatrix}$$

**Figure 3.4 Potential mediators and potential outcomes and the observed mediators and observed outcomes.**

Consider only the potential outcomes that are potentially observable, $\left(Z_i^{(1)}, Z_i^{(0)}, Y_i^{(1)}(Z^{(1)}), Y_i^{(0)}(Z^{(0)})\right)$, $i = 1,2,3,\ldots,$ , and assume these are independent and identically distributed(*iid*) random variables from a multivariate normal distribution with mean $(\mu_{Z^{(1)}}, \mu_{Z^{(0)}}, \mu_{Y^{(0)}(Z^{(0)})}, \mu_{Y^{(1)}(Z^{(1)})})'$ and variance matrix

$$\begin{pmatrix} \sigma^2_{Z^{(1)}} & \rho_{Z^{(0)}Z^{(1)}}\sigma_{Z^{(0)}}\sigma_{Z^{(1)}} & \rho_{Y^{(1)}Z^{(1)}}\sigma_{Y^{(1)}}\sigma_{Z^{(1)}} & \rho_{Y^{(0)}Z^{(1)}}\sigma_{Y^{(0)}}\sigma_{Z^{(1)}} \\ \rho_{Z^{(0)}Z^{(1)}}\sigma_{Z^{(0)}}\sigma_{Z^{(1)}} & \sigma^2_{Z^{(0)}} & \rho_{Y^{(1)}Z^{(0)}}\sigma_{Y^{(1)}}\sigma_{Z^{(0)}} & \rho_{Y^{(0)}Z^{(0)}}\sigma_{Y^{(0)}}\sigma_{Z^{(0)}} \\ \rho_{Y^{(1)}Z^{(1)}}\sigma_{Y^{(1)}}\sigma_{Z^{(1)}} & \rho_{Y^{(1)}Z^{(0)}}\sigma_{Y^{(1)}}\sigma_{Z^{(0)}} & \sigma^2_{Y^{(1)}} & \rho_{Y^{(0)}Y^{(1)}}\sigma_{Y^{(0)}}\sigma_{Y^{(1)}} \\ \rho_{Y^{(0)}Z^{(1)}}\sigma_{Y^{(0)}}\sigma_{Z^{(1)}} & \rho_{Y^{(0)}Z^{(0)}}\sigma_{Y^{(0)}}\sigma_{Z^{(0)}} & \rho_{Y^{(0)}Y^{(1)}}\sigma_{Y^{(0)}}\sigma_{Y^{(1)}} & \sigma^2_{Y^{(0)}} \end{pmatrix}$$

**Figure 3.5 Variance-Covariance matrixes for the potentially observable outcomes.**

The parameters of this distribution, except $\rho_{Z1Z0}, \rho_{Y1Y0}, \rho_{Z1Y0}, \rho_{Z0Y1}$, can be estimated from the marginal distributions of $\left(Z^{(1)}, Z^{(0)}, Y^{(1)}(Z^{(1)}), Y^{(0)}(Z^{(0)})\right)$ and from the bivariate normal distributions of $\left(Z^{(1)}, Y^{(1)}\right)$ and $\left(Z^{(0)}, Y^{(0)}\right)$.

Because the direct and indirect effects are counterfactual quantities, in general we will not be able to compute these for any individual in the population. But under certain assumptions we can estimate them on average.

We assume that there is no interaction between treatment and the mediator, the direct effects are independent of the indirect effects. Then the expected value and the variance of total effect depend on the expectation and variance of the indirect effects. So we are interested in variance of an indirect effect for each subject. The following are the cases to be considered:

**Case (1):** $B$ *is* fixed, $Z^{(1)} - Z^{(0)}$ varies,

where $P(B = \beta) = 1$ and $\begin{pmatrix} Z^{(1)} \\ Z^{(0)} \end{pmatrix} \sim BVN\left(\begin{pmatrix} \mu_{Z^{(1)}} \\ \mu_{Z^{(0)}} \end{pmatrix}, \begin{pmatrix} \sigma^2_{Z^{(1)}} & \rho_{Z^{(1)}Z^{(0)}}\sigma_{Z^{(1)}}\sigma_{Z^{(0)}} \\ \rho_{Z^{(1)}Z^{(0)}}\sigma_{Z^{(1)}}\sigma_{Z^{(0)}} & \sigma^2_{Z^{(0)}} \end{pmatrix}\right)$.

Consider the individual indirect effect, $IIE = \beta(Z^{(1)} - Z^{(0)})$.

Then $E(IE) = E[\beta(Z^{(1)} - Z^{(0)})] \Rightarrow \mu_{IE} = \beta(\mu_{Z^{(1)}} - \mu_{Z^{(0)}}) = \beta\mu_{D_Z}$ which can be estimated from the observable data and $Var(IE) = Var[\beta(Z^{(1)} - Z^{(0)})]$ implies

$$\begin{aligned} \sigma^2_{IIE} &= \beta^2 Var(Z^{(1)} - Z^{(0)}) \\ &= \beta^2 \sigma^2_{D_Z} \\ &= \beta^2(\sigma^2_{Z^{(1)}} + \sigma^2_{Z^{(0)}} - 2\sigma_{Z^{(1)}}\sigma_{Z^{(0)}}\rho_{Z^{(1)}Z^{(0)}}) \\ &= \beta^2\left[(\sigma_{Z^{(1)}} - \sigma_{Z^{(0)}})^2 + 2\sigma_{Z^{(1)}}\sigma_{Z^{(0)}}(1 - \rho_{Z^{(1)}Z^{(0)}})\right] \end{aligned}$$

$\sigma^2_{IIE}$ cannot be directly estimated because $\sigma^2_{D_Z} = \sigma^2_{Z^{(1)}} + \sigma^2_{Z^{(0)}} - 2\sigma_{Z^{(1)}}\sigma_{Z^{(0)}}\rho_{Z^{(1)}Z^{(0)}}$ and there is no available information about $\rho_{Z^{(1)}Z^{(0)}}$ in observable data; that is, it is nonestimable.

Since $\sigma^2_{IIE} = \beta^2\left[(\sigma_{Z^{(1)}} - \sigma_{Z^{(0)}})^2 + 2\sigma_{Z^{(1)}}\sigma_{Z^{(0)}}(1-\rho_{Z^{(1)}Z^{(0)}})\right]$, there is a variability in indirect effects or S-M interaction present unless $\sigma_{Z^{(1)}} = \sigma_{Z^{(0)}}$ and $\rho_{Z^{(1)}Z^{(0)}} = 1$, and the former condition can be tested using observed data but the latter cannot.

In practice, one will not know the true value of $\rho_{Z^{(1)}Z^{(0)}}$ and, so, will be unable to directly evaluate the size of $\sigma^2_{IIE}$ with respect to $\mu_{IIE}$. Little else can be done with observable data except to note that letting $\rho_{Z^{(1)}Z^{(0)}} = 1$ and - 1 produces bounds for $\sigma^2_{IIE}$ that can be estimated.

If $\rho_{Z^{(1)}Z^{(0)}} = 1$, then $\sigma^2_{IIE} = \beta^2\left[(\sigma_{Z^{(1)}} - \sigma_{Z^{(0)}})^2\right]$ and if $\rho_{Z^{(1)}Z^{(0)}} = -1$, then $\sigma^2_{IE} = B^2\left[(\sigma_{Z^{(1)}} + \sigma_{Z^{(0)}})^2\right]$ implies $\beta^2\left[(\sigma_{Z^{(1)}} - \sigma_{Z^{(0)}})^2\right] \le \sigma^2_{IIE} \le \beta^2\left[(\sigma_{Z^{(1)}} + \sigma_{Z^{(0)}})^2\right]$ which are the estimable lower and upper bounds for the variability of mediating effects for an individual.

**Case (2):** $B$ varies, $Z^{(1)} - Z^{(0)}$ fixed,

where $P((Z^{(1)} - Z^{(0)}) = \mu_{D_Z}) = 1$ and $B \sim N(\beta, \sigma^2_B)$.

Then

$E(IIE) = E[B\mu_{D_Z}] \Rightarrow \mu_{IIE} = \mu_{D_Z}E(B) = \mu_{D_Z}\beta$ which is estimable as $\beta$, the expected value of the slope variable, can directly be estimated from the observed data using Baron & Kenny approach and $\mu_{D_Z}$ is fixed.

$Var(IIE) = Var[B\mu_{D_Z}]$

$$\begin{aligned}\sigma^2_{IIE} &= \mu^2_{D_Z}Var(B)\\ &= \mu^2_{D_Z}\sigma^2_B\end{aligned}$$

$\sigma^2_{IIE}$ cannot be directly estimated because $\sigma^2_{IIE} = \mu^2_{D_Z}\sigma^2_B$ , a special method is needed to estimate $\sigma^2_B$, the variance of the slope variable, which we will discuss later.

**Case (3):** $B$ varies, $Z^{(1)} - Z^{(0)}$ varies,

where $B \sim N(\beta, \sigma_B^2)$ and $\begin{pmatrix} Z^{(1)} \\ Z^{(0)} \end{pmatrix} \sim BVN\left(\begin{pmatrix} \mu_{Z^{(1)}} \\ \mu_{Z^{(0)}} \end{pmatrix}, \begin{pmatrix} \sigma^2_{Z^{(1)}} & \rho_{Z^{(1)}Z^{(0)}}\sigma_{Z^{(1)}}\sigma_{Z^{(0)}} \\ \rho_{Z^{(1)}Z^{(0)}}\sigma_{Z^{(1)}}\sigma_{Z^{(0)}} & \sigma^2_{Z^{(0)}} \end{pmatrix}\right)$.

Consider $IIE = B(Z^{(1)} - Z^{(0)})$.

$E(IIE) = E[B(Z^{(1)} - Z^{(0)})] \Rightarrow \mu_{IIE} = E[E[B(Z^{(1)} - Z^{(0)})|B]] = E(B\mu_{D_Z}) = \mu_{D_Z}E(B) = \mu_{D_Z}\beta$ where $\mu_{D_Z}$ can be estimated and $\beta$ also can be estimated from the observed data using the Baron & Kenny approach.

$Var(IIE) = Var[B(Z^{(1)} - Z^{(0)})]$ implies

$$\begin{aligned}
\sigma^2_{IIE} &= Var[E(B(Z^{(1)} - Z^{(0)})|B)] + E[Var(B(Z^{(1)} - Z^{(0)})|B)] \\
&= Var[B(\mu_{Z^{(1)}} - \mu_{Z^{(0)}})] + E[\sigma^2_B Var(Z^{(1)} - Z^{(0)})] \\
&= \mu^2_{D_Z}\sigma^2_B + \sigma^2_{D_Z}E(B^2) \\
&= \mu^2_{D_Z}\sigma^2_B + \sigma^2_{D_Z}(Var(B) + (E(B))^2) \\
&= \mu^2_{D_Z}\sigma^2_B + \sigma^2_{D_Z}\sigma^2_B + \sigma^2_{D_Z}\mu^2_B \\
&= \mu^2_{D_Z}\sigma^2_B + (\sigma^2_B + \mu^2_B)\sigma^2_{D_Z} \\
&= \mu^2_{D_Z}\sigma^2_B + (\sigma^2_B + \mu^2_B)\left[(\sigma_{Z^{(1)}} - \sigma_{Z^{(0)}})^2 + 2\sigma_{Z^{(1)}}\sigma_{Z^{(0)}}(1 - \rho_{Z^{(1)}Z^{(0)}})\right]
\end{aligned}$$

$\sigma^2_{IIE}$ cannot be directly estimated because there is no available information about $\rho_{Z^{(1)}Z^{(0)}}$ in observable data and one cannot estimate $\sigma^2_B$ without assumptions and design considerations.

If an estimate of $\sigma^2_B$ were available, then if $\rho_{Z^{(1)}Z^{(0)}} = 1$, $\sigma^2_{IIE} = \mu^2_{D_Z}\sigma^2_B + (\sigma^2_B + \mu^2_B)\left[(\sigma_{Z^{(1)}} - \sigma_{Z^{(0)}})^2\right]$ and if $\rho_{Z^{(1)}Z^{(0)}} = -1$, then $\sigma^2_{IIE} = \mu^2_{D_Z}\sigma^2_B + (\sigma^2_B + \mu^2_B)\left[(\sigma_{Z^{(1)}} + \sigma_{Z^{(0)}})^2\right]$ implies

$\mu^2_{D_Z}\sigma^2_B + (\sigma^2_B + \mu^2_B)\left[(\sigma_{Z^{(1)}} - \sigma_{Z^{(0)}})^2\right] \le \sigma^2_{IIE} \le \mu^2_{D_Z}\sigma^2_B + (\sigma^2_B + \mu^2_B)\left[(\sigma_{Z^{(1)}} + \sigma_{Z^{(0)}})^2\right]$ which would be estimable lower and upper bounds for the variability of mediating effects for an individual.

This chapter has explored variability in mediating effects and quantified them. In doing so, estimable and nonestimable quantities were identified. Bounds for some of these quantities are estimable but estimating or bounding the variance in slopes is problematic. In Chapter 6, some initial ideas for terminal measures data is presented that involved different designs that can facilitate at least a naïve estimate of $\sigma^2_B$. First, however, the issues involved with terminal measures data will be discussed and two designs considered for analyzing terminal measures

data. One design does not work, but the other does with an assumption that may or may not be plausible, depending on the application.

# Chapter 4 - Estimating Heterogeneity of Individual Indirect Effects

In chapter 3, individual indirect effects were introduced. The mediation plot was extended to the individual mediating plot to describe the relationship among the individual total, direct, and indirect effects. Using the individual mediation plots, sources of variability of the total effect across individuals was demonstrated. The individual indirect effect was defined and the variability of individual indirect effects was formulated based on the estimable and non estimable quantities. The upper and lower bounds of the variability of individual mediating effects was shown for three different cases. Then the mediating heterogeneity or subject-mediator (treatment) interaction was defined.

In this chapter, we discuss the problems of estimating the variance of individual indirect effects or subject-mediator (treatment) interaction for a two sample completely randomized block design with *s*-disjoint blocks and *n*-individuals for each block. Following an example used in chapter 4 when illustrating terminal measures data, we use the term 'strain' for block for effect strains of mice as is done there. For the most general situation (that is, when both $B$ and $D_Z$ vary), the variability of individual indirect effects is given by the formula:

$$\sigma^2_{IIE} = \mu^2_{D_Z}\sigma^2_B + (\sigma^2_B + \mu^2_B)\sigma^2_{D_Z} = \mu^2_{D_Z}\sigma^2_B + (\sigma^2_B + \mu^2_B)\left[(\sigma_{Z^{(1)}} - \sigma_{Z^{(0)}})^2 + 2\sigma_{Z^{(1)}}\sigma_{Z^{(0)}}(1-\rho_{Z^{(1)}Z^{(0)}})\right],$$

where $\mu_B$ : the mean of the slope variable, estimable from the observed data.

$\sigma^2_B$ : the variance of the slope variable, nonestimable from the observed data.

$\sigma^2_{Z^{(1)}}$ : the variance of the mediator when the treatment assignment, $T = 1$, estimable.

$\sigma^2_{Z^{(0)}}$ : the variance of the mediator when the treatment assignment, $T = 0$, estimable.

$\rho_{Z^{(1)}Z^{(0)}}$ : the unconditional correlation between $Z^{(1)}$ and $Z^{(0)}$, nonestimable.

It is assumed that,

$B \sim N(\beta, \sigma^2_B)$ and $\begin{pmatrix} Z^{(1)} \\ Z^{(0)} \end{pmatrix} \sim BVN\left(\begin{pmatrix} \mu_{Z^{(1)}} \\ \mu_{Z^{(0)}} \end{pmatrix}, \begin{pmatrix} \sigma^2_{Z^{(1)}} & \rho_{Z^{(1)}Z^{(0)}}\sigma_{Z^{(1)}}\sigma_{Z^{(0)}} \\ \rho_{Z^{(1)}Z^{(0)}}\sigma_{Z^{(1)}}\sigma_{Z^{(0)}} & \sigma^2_{Z^{(0)}} \end{pmatrix}\right)$. The key problems of estimating the variability of individual indirect effects, $\sigma^2_{IIE}$, are based on the

estimates of two terms: $\sigma^2_{D_Z}$ which cannot be estimated because there is no available information about $\rho_{Z^{(1)}Z^{(0)}}$, and $\sigma^2_B$ cannot be estimated because individual slopes that are not observable.

## 4.1 Bounds of $\sigma^2_{IIE}$:

In practice, one will not know the true value of $\rho_{Z^{(1)}Z^{(0)}}$, but little else can be done with observable data except to note that letting $\rho_{Z^{(1)}Z^{(0)}} = 1$ and $\rho_{Z^{(1)}Z^{(0)}} = -1$ produces bounds for $\sigma^2_{D_Z}$ that can be estimated. The two quantities,

$\sigma^2_{IIE_L} = \mu^2_{D_Z}\sigma^2_B + (\sigma^2_B + \beta^2)\left[(\sigma_{Z^{(1)}} - \sigma_{Z^{(0)}})^2\right]$ and $\sigma^2_{IIE_U} = \mu^2_{D_Z}\sigma^2_B + (\sigma^2_B + \beta^2)\left[(\sigma_{Z^{(1)}} + \sigma_{Z^{(0)}})^2\right]$ would be the estimable lower and upper bounds for the variability of mediating effects for an individual, but the challenge is to estimate $\sigma^2_B$ which is the variability of individual slopes.

We proceed by assuming there is a blocking variable, i.e., strain that groups individuals into 'nearly' homogeneous subsets. Observe that

$\sigma^2_B = Var(B) = E\left(Var(B|s = j)\right) + Var\left(E(B|s = j)\right) = E(\sigma^2_{B_e}) + Var(\beta_j)$, where the conditional expectations and variances are 'within strain' and the outer expectations and variances are across strains, meaning averages and variances across a given set of strains. Note that $Var(\beta_j)$ can be estimated by estimating the mean slope for each strain and then taking the variance of these estimates across strain. The within strain variance, $\sigma^2_{B_e}$, remains a problem for estimation. It is proposed that, for a lower bound, homogeneity of slopes within strains is assumed so that $\sigma^2_{B_e} = 0$. For an upper bound, it is assumed that $\sigma^2_{B_e} = Var(\beta_j)$, it is argued that, for an effective blocking variable for a design investigating variability in treatment effects, the within subset variation should be smaller than the variation across subsets.

Given the above assumptions, proposed lower and upper bounds are now

$\sigma^2_{IIE_L} = \mu^2_{D_Z}Var(\beta_j) + \left(Var(\beta_j) + \beta^2\right)\left[(\sigma_{Z^{(1)}} - \sigma_{Z^{(0)}})^2\right]$ and

$\sigma^2_{IIE_U} = 2\mu^2_{D_Z}Var(\beta_j) + \left(2Var(\beta_j) + \beta^2\right)\left[(\sigma_{Z^{(1)}} + \sigma_{Z^{(0)}})^2\right]$

### *4.1.1 Estimating Bounds of $\sigma^2_{IIE}$ :*

Assume a balanced two sample randomized complete block design, so there are *n* units per strain for a total of *ns* individuals. We assume that within each strain $j$ , $n/2$ units are randomly assigned to treatment $T=1$ and the other $n/2$ units assigned to $T=0$ . Without loss of generality, assume that the first $n/2$ units in the *j*th subset receive $T=1$ and the second $n/2$ receive $T=0$. Let $Z_{ij}^{(1)}$ *for* $i=1,2,....,n/2$ ; $j=1,2,...,s$ be the observed mediators when a subject recieves $T=1$ and let $Z_{ij}^{(0)}$ *for* $i=(n/2+1),...,n$ ; $j=1,2,...,s$ be the observed mediators when subject recieves $T=0$. Let $B_{ij}$ *for* $i=1,2,....,n$ ; $j=1,2,...,s$ be the individual slopes for each individual *i* (not available from the observed data).

Let $\overline{Z}_j^{(1)}=\frac{1}{(n/2)}\sum_{i=1}^{n/2}Z_{ij}^{(1)}$ , $\overline{Z}_j^{(0)}=\frac{1}{(n/2)}\sum_{i=n/2+1}^{n}Z_{ij}^{(0)}$ , $\hat{\sigma}^2_{Z_j^{(1)}}=\frac{\sum_{i=1}^{n/2}(Z_{ij}^{(1)}-\overline{Z}_j^{(1)})^2}{((n/2)-1)}$, and

$\hat{\sigma}^2_{Z_j^{(0)}}=\frac{\sum_{i=n/2+1}^{n}(Z_{ij}^{(0)}-\overline{Z}_j^{(0)})^2}{((n/2)-1)}$; $j=1,2,...,s$. Further, let $\overline{D}_{Z_j}=\overline{Z}_j^{(1)}-\overline{Z}_j^{(0)}$ and define $\hat{\mu}_{D_Z}=\frac{1}{s}\sum_j\overline{D}_{Z_j}$ as a proposed estimator for the mean effect of the treatment on the mediating variable. Now define $\left[\overline{(\hat{\sigma}_{Z^{(1)}}+\hat{\sigma}_{Z^{(0)}})^2}\right]=\frac{1}{s}\sum_j\left(\hat{\sigma}_{Z_j^{(1)}}+\hat{\sigma}_{Z_j^{(0)}}\right)^2$ and $\left[\overline{(\hat{\sigma}_{Z^{(1)}}-\hat{\sigma}_{Z^{(0)}})^2}\right]=\frac{1}{s}\sum_j\left(\hat{\sigma}_{Z_j^{(1)}}-\hat{\sigma}_{Z_j^{(0)}}\right)^2$, and let $\hat{\beta}_j$ denote the Barron and Kenny estimate of the slope parameter for a particular strain *j*, so that $\hat{\beta}=\frac{1}{s}\sum_j\hat{\beta}_j$ becomes a population estimate of the slope parameter. For interpretability of this estimate, it is assumed that there is no qualitative interaction on the slope parameters for each strain and, similarly, on the mean treatment effect for each strain. The bounds are still valid at the population level even without this assumption, but the interpretability is lacking. A qualitative interaction on indirect effects is the subject of chapter 5. Finally, let $\hat{Var}(\beta_j)=\frac{1}{s-1}\sum_{j=1}^{s}(\hat{\beta}_j-\hat{\beta})^2$ denote the estimated between strain variance on the slope parameters. The proposed estimated lower and upper bounds of the variance of individual indirect effects is given by,

$\hat{\sigma}^2_{IIE_L} = \hat{\mu}^2_{D_Z} \widehat{Var}(\hat{\beta}_j) + (\widehat{Var}(\hat{\beta}_j) + \hat{\beta}^2)\left[\overline{(\hat{\sigma}_{Z^{(1)}} - \hat{\sigma}_{Z^{(0)}})^2}\right]$ and

$\hat{\sigma}^2_{IIE_U} = 2\hat{\mu}^2_{D_Z} \widehat{Var}(\hat{\beta}_j) + (2\widehat{Var}(\hat{\beta}_j) + \hat{\beta}^2)\left[\overline{(\hat{\sigma}_{Z^{(1)}} + \hat{\sigma}_{Z^{(0)}})^2}\right]$.

## 4.2 An illustration using simulated data

The potential outcomes were generated for *s*-strains with *n*-individuals in each strain so that the total number of mice was *ns*. The observed data were generated by removing counterfactuals. Then from the observed data, the individual indirect effects, $IIE = BD_Z$, were generated for each strain. Thus, a two stage hierarchical approach is used to simulate data. The next section shows the steps and parameter setting for this simulation study.

### *4.2.1 Steps of Generating Potential Outcomes, Observed Data and Individual Indirect Effects*

To generate potential outcomes of $Z^{'}$, we use the following steps:

(1). Let $\mu_{Z_j^{(0)}}$ and $\mu_{Z_j^{(1)}}$ be the mean of $Z^{(0)}$ and $Z^{(1)}$ for strains in the control and experimental groups, respectively. The statistical model corresponding to $\mu_{Z_j^{(1)}}$ and $\mu_{Z_j^{(0)}}$ for each strain, $j = 1,2,...,s$ is:

$$\begin{pmatrix} \mu_{Z_j^{(1)}} \\ \mu_{Z_j^{(0)}} \end{pmatrix} \sim MVN\left( \begin{pmatrix} \mu_{Z^{(1)}} \\ \mu_{Z^{(0)}} \end{pmatrix}, \begin{pmatrix} \sigma^2_{\mu_{Z^{(1)}}} & 0 \\ 0 & \sigma^2_{\mu_{Z^{(0)}}} \end{pmatrix} \right) \text{ for } j = 1,2,...,s\,.$$

Here, $\mu_{Z^{(0)}}$ and $\mu_{Z^{(1)}}$ represent the means of mediator across individuals in the populations from which control group and the experiment group were sampled. The mean treatment effect on the mediator for each strain, $\mu_{D_{Z_j}} = \mu_{Z_j^{(1)}} - \mu_{Z_j^{(0)}}$, is a random variable with respect to the superpopulation from which parameters for each strain are generated. The superpopulation parameters are then $E(\mu_{D_{Z_j}}) = E(\mu_{Z_j^{(1)}}) - E(\mu_{Z_j^{(0)}}) = \mu_{Z^{(1)}} - \mu_{Z^{(0)}}$ and $Var(\mu_{D_{Z_j}}) = Var(\mu_{Z_j^{(1)}} - \mu_{Z_j^{(0)}}) = \sigma^2_{\mu_{Z_j^{(1)}}} + \sigma^2_{\mu_{Z_j^{(1)}}} - 2Cov(\mu_{Z_j^{(1)}}, \mu_{Z_j^{(0)}})$. We set $\rho_{\mu_{Z^{(1)}}\mu_{Z^{(0)}}} = 0$, so that

the strain level means are assumed to be independent across strains, and $\mu_{Z^{(1)}} = 17$ , $\mu_{Z^{(0)}} = 22$ and $\sigma^2_{\mu_j^{(0)}} = 4.5,\ \sigma^2_{\mu_j^{(1)}} = 4.5$. Thus, $\mu_{D_{Z_j}} \sim N(-5,\ 9)$.

(2) Let $Z_{ij}^{(1)}$ and $Z_{ij}^{(0)}$ be the mediator variables for all individuals for all strains in the treatment and control groups, respectively. The statistical model corresponding to $Z_{ij}^{(1)}$ and $Z_{ij}^{(0)}$ for $i = 1,2,....,n$ ; $j = 1,2,...,s$ is:

$$\begin{pmatrix} Z_{ij}^{(1)} \\ Z_{ij}^{(0)} \end{pmatrix} \sim BVN\left( \begin{pmatrix} \mu_{Z_j^{(1)}} \\ \mu_{Z_j^{(0)}} \end{pmatrix}, \begin{pmatrix} \sigma^2_{Z^{(1)}} & \rho_{Z^{(1)}Z^{(0)}}\sigma_{Z^{(1)}}\sigma_{Z^{(0)}} \\ \rho_{Z^{(1)}Z^{(0)}}\sigma_{Z^{(1)}}\sigma_{Z^{(0)}} & \sigma^2_{Z^{(0)}} \end{pmatrix} \right),\ for\ i = 1,2,....,n\ ;\ \ j = 1,2,...,s$$

We set $\rho_{Z^{(1)}Z^{(0)}} = 0.5$ , $\sigma^2_{Z^{(0)}} = 4,\ \sigma^2_{Z^{(1)}} = 5$. Thus, $D_{Z_{ij}} \sim N(-5,\ 13.2)$.

To generate intercepts when *T*=0 and when *T*=1 for each individual, we use the following steps:

(3) We repeat steps (1) and (2) with different parameter setting:

Set the mean direct effects, $DE_j = \mu_{A_1} - \mu_{A_0}$ for each strain $j = 1,2,...,s$ and also the individual direct effect, $DE_{ij} = (A_{1,ij} - A_{0,ij})$ within each strain for each individual, $i = 1,2,....,n$ ; $j = 1,2,...,s$ . Here, $\mu_{A_1} - \mu_{A_0} = 10 - 3 = 7.$ Set some specific values for correlations and the variances. We do not pay attention to these intercepts here as individual indirect effects, $IIE = B(Z^{(1)} - Z^{(0)})$, do not depend on intercepts and it is *IIE* is the focus here.

To generate potential slopes for each individual, $B_{ij}$ , we use the following steps:

(4) The distribution of the mean slopes for each strain is assumed to be $\beta_j \sim N(\beta, \sigma^2_{B_\mu})$ , where the specified values are $\beta = -1$ $\sigma^2_{B_\mu} = 0.16$. Generate the individual slopes so that $B_{ij} = \beta_j + \varepsilon_{ij}$ , where $\varepsilon_{ij} \sim N(0, \sigma^2_{B_e})$ for each individual within each strain, $i = 1,2,....,n$ ; $j = 1,2,...,s$ . The distribution of the slopes for all individuals: $B_{ij} \sim N(\beta, \sigma^2_B)$ , where $\beta = -1,\ \sigma^2_B = \sigma^2_{B_\mu} + \sigma^2_{B_e} = 0.16 + 0.04 = 0.20.$

To generate potential outcomes of $Y^{'}$s given $Z^{'}$s we use the following steps:

(5) For individual *i:* generate $Y_{ij}^{(1)}(Z_{ij}^{(1)}) = A_{1ij} + B_{ij}\, Z_{ij}^{(1)}$ , $Y_{ij}^{(0)}(Z_{ij}^{(0)}) = A_{0ij} + B_{ij} Z_{ij}^{(0)}$, $Y_{ij}^{(1)}(Z_{ij}^{(0)}) = A_{1ij} + B_{ij} Z_{ij}^{(0)}$ and $Y_{ij}^{(0)}(Z_{ij}^{(1)}) = A_{0ij} + B_{ij} Z_{ij}^{(1)}$ which are mentioned in section (3.2).

(6) From steps (1)-(5), potential outcomes are generated and then, removing counterfactuals, we generate the observed data.

### *4.2.2 The distributions of Average of Individual Indirect Effects and the Indirect Effects for each strain*

If we can observe $B_{ij}$ *and* $D_{Z_{ij}}$ for each individual for all strains , then we would be able to calculate indirect effects for each individual within a strain and estimate the average indirect effect for the particular strain. The distributions of average of individual indirect effects for strain $j$, $\overline{IIE}_j = \overline{B_j D_{Z_j}} = \frac{1}{n}\sum_{i=1}^{n} B_{ij} D_{Z_{ij}}$ . The average indirect effect ( which is similar to the Baron & Kenny Approach) for strain $j$, $IE_j = \overline{B}_j \overline{D}_{Z_j} = \left(\frac{1}{n}\sum_{i=1}^{n} B_{ij}\right)\left(\frac{1}{n}\sum_{i=1}^{n} D_{Z_{ij}}\right)$; $j = 1,2,...,s$. Now, we discuss the distributions of these two estimates.

Since $\overline{B}_j \sim N(\beta, \sigma^2_{\overline{B}})$ and $\overline{D}_{Z_j} \sim N(\mu_{D_Z}, \sigma^2_{\overline{D}_Z})$, and $B_{ij}$ *and* $D_{Z_{ij}}$ are independent, we have

$$E(IE_j) = E(\overline{B}_j \overline{D}_{Z_j}) = E(\overline{B}_j)E(\overline{D}_{Z_j}) = \beta\mu_{D_Z} = (-1)(-5) = 5,$$

$$\begin{aligned} Var(IE_j) &= Var(\overline{B}_j \overline{D}_{Z_j}) = E(\overline{B}_j \overline{D}_{Z_j})^2 - \left[E(\overline{B}_j \overline{D}_{Z_j})\right]^2 = E(\overline{B}_j^2).E(\overline{D}_{Z_j}^2) - \left[E(\overline{B}_j).E(\overline{D}_{Z_j})\right]^2 \\ &= E(\overline{B}_j^2).E(\overline{D}_{Z_j}^2) - \left[\beta.\mu_{D_Z}\right]^2 \\ &= \left[Var(\overline{B}_j) + (E(\overline{B}_j))^2\right]\left[Var(\overline{D}_{Z_j}) + (E(\overline{D}_{Z_j}))^2\right] - \left[\beta.\mu_{D_Z}\right]^2 \\ &= \left[\sigma^2_{B_\mu} + \frac{\sigma^2_{B_e}}{n} + \beta^2\right]\left[\sigma^2_{D_{Z_\mu}} + \frac{\sigma^2_{D_{Z_e}}}{n} + \mu_{D_Z}\right] - \left[\beta.\mu_{D_Z}\right]^2 \\ &= (0.162+1)(9.23+25) - 25 = 14.78. \end{aligned}$$

$$\begin{aligned} E(\overline{IIE}_j) = E(\overline{B_{ij} D_{Z_{ij}}}) &= E\left[\frac{1}{n}\sum_{i=1}^{n} B_{ij} D_{Z_{ij}}\right] = \frac{1}{n}E\left(\sum_{i=1}^{n} B_{ij} D_{Z_{ij}}\right) = \frac{1}{n}E\left(E\left(\sum_{i=1}^{n} B_{ij} D_{Z_{ij}} \,|\, j\right)\right) \\ &= \frac{1}{n}E\left(nE(B_{ij}|j)E(D_{Z_{ij}}|j)\right) = E\left(\beta_j \mu_{D_{Z_j}}\right) = \beta\mu_{D_Z} = (-1)(-5) = 5, \end{aligned}$$

$$Var(\overline{IIE}_j) = Var(\overline{B_{ij}D_{Z_{ij}}}) = Var\left[\frac{1}{n}\sum_{i=1}^{n} B_{ij}D_{Z_{ij}}\right] = \frac{1}{n^2}Var\left(\sum_{i=1}^{n} B_{ij}D_{Z_{ij}}\right)$$

$$= \frac{1}{n^2}\left[Var\left(E\left(\sum_{i=1}^{n} B_{ij}D_{Z_{ij}} \mid j\right)\right) + E\left(Var\left(\sum_{i=1}^{n} B_{ij}D_{Z_{ij}} \mid j\right)\right)\right] \rightarrow (A)$$

$$Var\left(E\left(\sum_{i=1}^{n} B_{ij}D_{Z_{ij}} \mid j\right)\right) = Var(n\beta_j\mu_{D_{Z_j}}) = n^2\left[E(\beta_j\mu_{D_{Z_j}})^2 - \left(E(\beta_j\mu_{D_{Z_j}})\right)^2\right]$$

$$= n^2\left[\sigma^2_{B_\mu}\sigma^2_{D_{Z_\mu}} + \sigma^2_{B_\mu}\mu^2_{D_Z} + \beta^2\sigma^2_{D_{Z_\mu}}\right] \rightarrow (A1)$$

$$E\left(Var\left(\sum_{i=1}^{n} B_{ij}D_{Z_{ij}} \mid j\right)\right) = E\left[n(B_{ij}D_{Z_{ij}} \mid j)\right] = nE\left[E(B_{ij}D_{Z_{ij}} \mid j)^2 - \left(E(B_{ij}D_{Z_{ij}} \mid j)\right)^2\right]$$

$$= n\left[\sigma^2_{B_e}\sigma^2_{D_{Z_e}} + \sigma^2_{B_e}(\sigma^2_{D_{Z_\mu}} + \mu^2_{D_Z}) + \sigma^2_{D_{Z_e}}(\sigma^2_{B_\mu} + \beta^2)\right] \rightarrow (A2)$$

By Substituting (A1) and (A2) in to (A), we get

$$Var(\overline{IIE}_j) = (\sigma^2_{D_{Z_\mu}} + \mu^2_{D_Z})(\sigma^2_{B_\mu} + \frac{1}{n}\sigma^2_{B_e}) + \beta^2(\sigma^2_{D_{Z_\mu}} + \frac{1}{n}\sigma^2_{D_{Z_e}}) + \frac{1}{n}\sigma^2_{D_{Z_e}}(\sigma^2_{B_e} + \beta^2) = 14.78$$

Thus, the distributions of average of individual indirect effects for strain *j*, $\overline{IIE}_j = \overline{B_j D_{Z_j}}$ , and the average indirect effect ( which is similar to Baron & Kenny Approach) for strain $j$ , $IE_j = \overline{B}_j\overline{D}_{Z_j}$; $j = 1,2,...,s$ are nearly the same. This means, for homogeneous strains the variability of individual indirect effects can be estimated using the variance of estimated mediating effects obtained from the Baron and Kenny approach across strains. In the simulation study, we obtain density plots of $\overline{IIE}_j = \overline{B_j D_{Z_j}}$ and $IE_j = \overline{B}_j\overline{D}_{Z_j}$ which are shown in Figure 4.1.

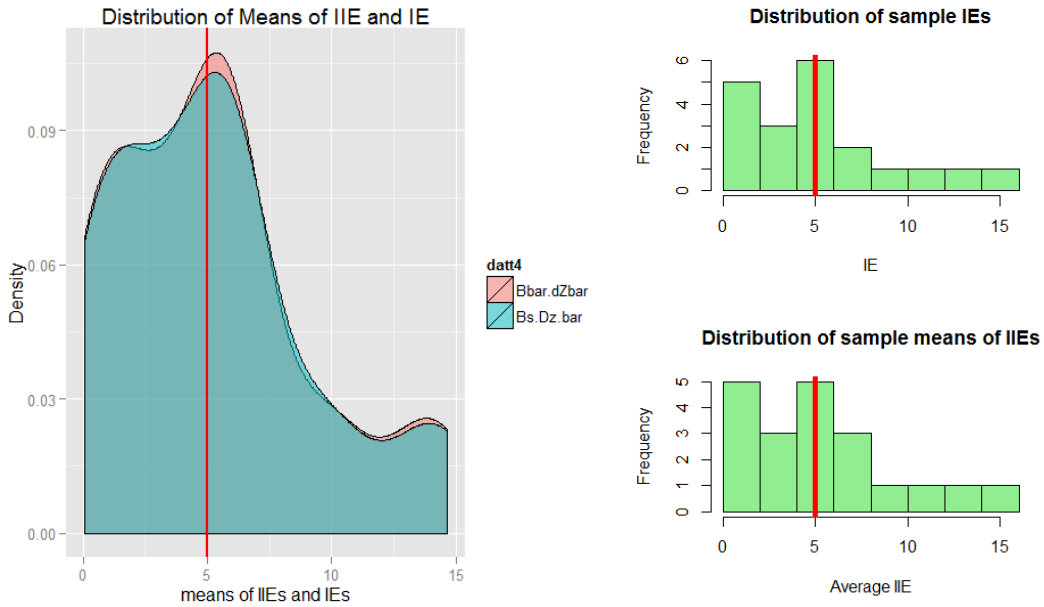



**Figure 4.1 The Distributions of Average of Individual Indirect Effects and the Indirect Effects for each strain.**

### *4.2.3 Estimates for the bounds using simulated data.*

Based on the parameter settings in the simulation study, the true variability of the individual indirect effects and, lower and upper bounds for the individual indirect effects are,

$$\sigma^2_{IIE} = \mu^2_{D_Z}\sigma^2_B + (\sigma^2_B + \beta^2)\sigma^2_{D_Z} = (-5)^2(0.20) + (0.20+1)(13.53) = 21.236$$

$$\sigma^2_{IIE_L} = \mu^2_{D_Z}\sigma^2_B + (\sigma^2_B + \beta^2)\left[(\sigma_{Z^{(1)}} - \sigma_{Z^{(0)}})^2\right]$$
$$= (25)(0.20) + (0.20+1)[\sqrt{5} - \sqrt{4}]^2 = 5.067 \text{ and}$$

$$\sigma^2_{IIE_U} = \mu^2_{D_Z}\sigma^2_B + (\sigma^2_B + \beta^2)\left[(\sigma_{Z^{(1)}} + \sigma_{Z^{(0)}})^2\right]$$
$$= (25)(0.20) + (0.20+1)[\sqrt{5} + \sqrt{4}]^2 = 26.533$$

Now, using the observed data, the estimates of the lower and upper bounds of heterogeneity of mediating effects as follows:

$$\hat{\sigma}^2_{IIE_L} = \hat{\mu}^2_{D_Z}\hat{\sigma}^2_{B_L} + (\hat{\sigma}^2_{B_L} + \hat{\beta}^2)\left[\overline{(\hat{\sigma}_{Z^{(1)}} - \hat{\sigma}_{Z^{(0)}})^2}\right] \text{ and } \sigma^2_{IIE_U} = \hat{\mu}^2_{D_Z}\hat{\sigma}^2_{B_U} + (\hat{\sigma}^2_{B_U} + \hat{\beta}^2)\left[\overline{(\hat{\sigma}_{Z^{(1)}} + \hat{\sigma}_{Z^{(0)}})^2}\right].$$

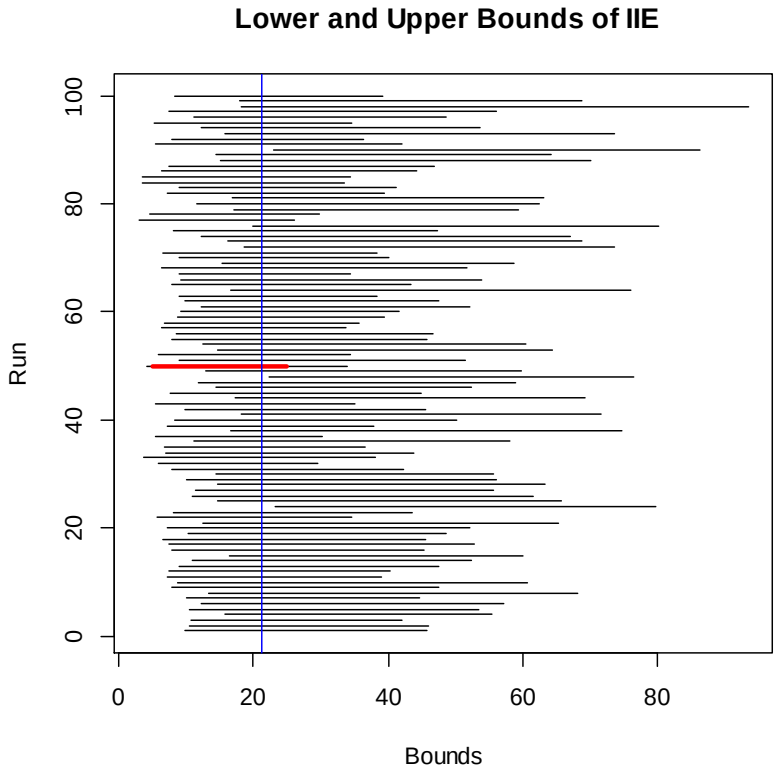



**Figure 4.2 The estimated lower and upper bounds for the variability of individual indirect effects for 100 Monte Carlo simulations**

Figure 4.2 shows the estimated lower and upper bounds for the individual indirect effects which are calculated using 100 Monte Carlo simulations. For each run, the lower and upper bounds are calculated based on 20 strains. Most of the intervals covered the true variability of individual indirect effects and two of them (2%) did not cover the true variability of individual indirect effects. Standard errors of lower and upper bounds can be found using non parametric bootstrapping .

### *4.2.4 Inverse Coefficient of Variation for the individual indirect effects :*

The coefficient of variation (CV), which is the ratio of the standard deviation to the mean, is a dimensionless measure of dispersion found to be very useful in many situations. Sometimes, it might be interpretable to work with the reciprocal of the CV, denoted ICV. The ICV has special applications in parametric inference problems for some important lifetime

distributions. The CV is often estimated by the ratio of the sample standard deviation to the sample mean, called the sample CV. We calculate the Ratio (ICV) between the estimates of indirect effects and their bounds: $\frac{\hat{\mu}_{IIE}}{\hat{\sigma}_{IIE_L}}, \frac{\hat{\mu}_{IIE}}{\hat{\sigma}_{IIE_U}}$ . The shapes of the distributions are shown in the Figure 4.3

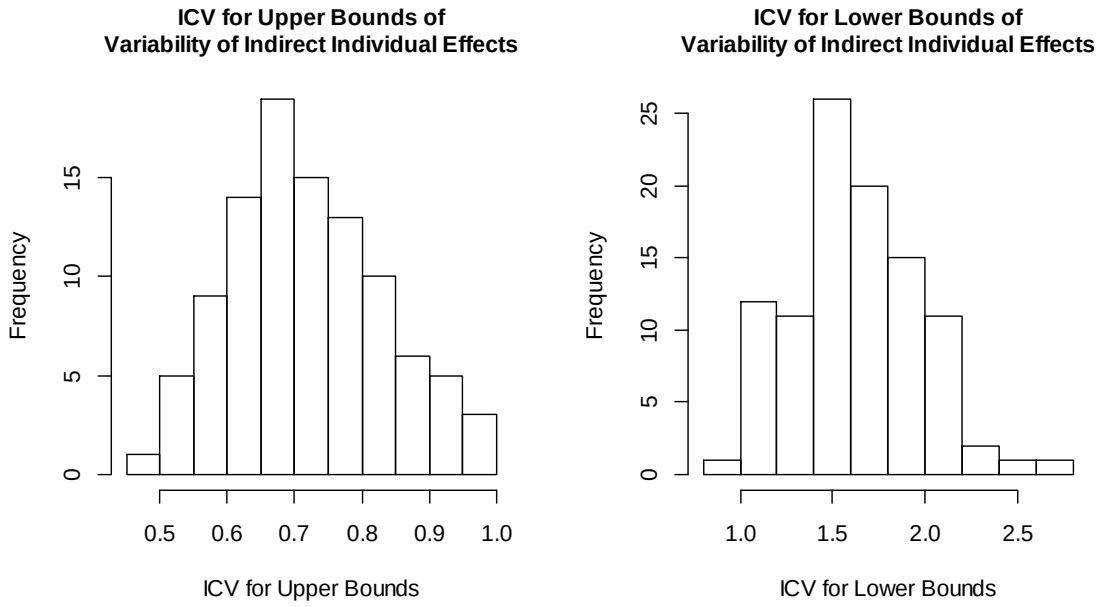



**Figure 4.3 The Estimated ICVs**

Assuming without loss of generality that, $\mu_{IIE} > 0$, then the proportion of the population with an unfavorable mediating effect is given by $P(IIE < 0)$. If $IIE$ is normally distributed, we would be able to estimate $P(IIE < 0)$ (Gadbury et.al.,2001).

The estimated mean of indirect individual effects is $\hat{\mu}_{IIE} = 4.95$ and $\hat{\sigma}_{IIE_L} = 2.52,\ \hat{\sigma}_{IIE_U} = 95.7$. Then the ICVs of lower and upper bounds are $\frac{\hat{\mu}_{IIE}}{\hat{\sigma}_{IIE_L}} = \frac{4.95}{2.52} \approx 2$ $\frac{\hat{\mu}_{IIE}}{\hat{\sigma}_{IIE_U}} = \frac{4.95}{95.7} \approx 0.05$, respectively. That is, the estimated ICV bounds are ( 0.05, 2) and this implies that the true mean of individual indirect effects, $\mu_{IIE}$ is estimated to be less than 2 (standard deviations) from 0. That is, heterogeneity of individual effects may be present and there may be qualitative interactions too.

# Chapter 5 - Testing for Qualitative Interaction on Indirect Effects

Evaluating treatment effects within subsets of patients plays a major part of the analysis of many major clinical trials. In order for the conclusion to be broad based, and to achieve reasonable sample sizes, the study design often includes a number of patient subsets, corresponding to different clinical centers or to patient subsets defined by prognostic factors. As a result, some variation in the estimates of treatment effect among subsets is expected. Assessment of the variation in treatment effect among subsets is important for several reasons. For example, the pattern of variation may lead to specific hypotheses concerning the relationship between efficacy and certain patient characteristics.

According to Poulson (2012, p16-17), "treatment heterogeneity is present when the effect of a treatment, say *T*, with respect to a treatment, *R*, varies across subsets or individuals in a population. At the individual level, this variability is called subject-treatment interaction (Gadbury 2010). A consequence of this heterogeneity is that the effect of a treatment *T* with respect to *R* may be in opposite directions across different individuals or subsets, with treatment *T* having higher efficacy for some and treatment *R* having higher efficacy for others."

The term Qualitative Interaction (QI) has been used to describe this situation at the subset (or subgroup) level (Peto, 1982; Gail and Simon 1985). Note that, as Yusuf (1991) mentioned, a proper subgroup is a group of patients characterized by a common set of 'baseline' parameters and in contrast, an improper subgroup is a group of patients characterized by a variable measured after randomization and potentially affected by treatments. A test has been developed to detect a QI (Gail and Simon 1985). A "quantitative" interaction (Peto 1982) exists when the magnitudes of the difference between treatment and control differ across subsets, but are in the same direction. In the presence of a quantitative treatment by subset interaction, it is easy to assess and interpret the overall treatment effect. In the presence of qualitative treatment by subset interaction, however, it is even hard to define a measure of the overall treatment effect.

The statistical problem of determining whether observed variability in treatment effects represents a qualitative interaction was addressed by Gail and Simon (1985). These authors, who

use the more visual terms “crossover and noncrossover” interaction for Peto’s “qualitative and quantitative” interaction, developed a normal theory likelihood ratio test to detect the presence of a qualitative interaction. Other tests followed including those of Berger (1989, Section 6), Zelterman (1990), the pushback tests of Ciminera *et al.* (1993), and the range test of Piantadosi and Gail (1993). Unfortunately, however, these tests perform poorly; the pushback tests can be liberal, and results in Piantadosi and Gail (1993) indicate that unreasonably large sample sizes may be needed to obtain adequate power with both the range and Gail and Simon tests.

Since we are interested in causal mediation analysis, the qualitative and quantitative interactions are defined as follows: qualitative interaction (QI) arises when the direction of the true indirect effects varies among subsets of individuals where as quantitative interactions arises when the magnitude, but not the direction, of the true indirect effect varies among subsets of individuals. This chapter adopts the work of Gail and Simon (1985) to the context of mediation analysis and indirect effects.

## 5.1 Testing QI on mediating effects across patient subsets

The concepts of no interactions, quantitative interactions and qualitative interactions are illustrated in Figure 5.1 for two patient subsets in mediation analysis.

Let $\delta_i = \beta_i \mu_{D_{Z_i}}$, $i = 1,2$ be the true indirect effects for two treatments in subsets 1 and 2, respectively. The origin represents the hypothesis of no indirect effect for either subset. The line $\delta_1 = \delta_2$ represents the locus of points for which there are same indirect effects (except for the origin) but no mediator by subset interactions. Any point which does not fall on this line defines a quantitative interaction. qualitative interactions consist of points only in the second ( $\delta_1 < 0, \delta_2 > 0$ ) , and forth ( $\delta_1 > 0, \delta_2 < 0$ ) quadrants. Non cross over interactions consist of those points in the first and third quadrants which do not lie on the line $\delta_1 = \delta_2$ .

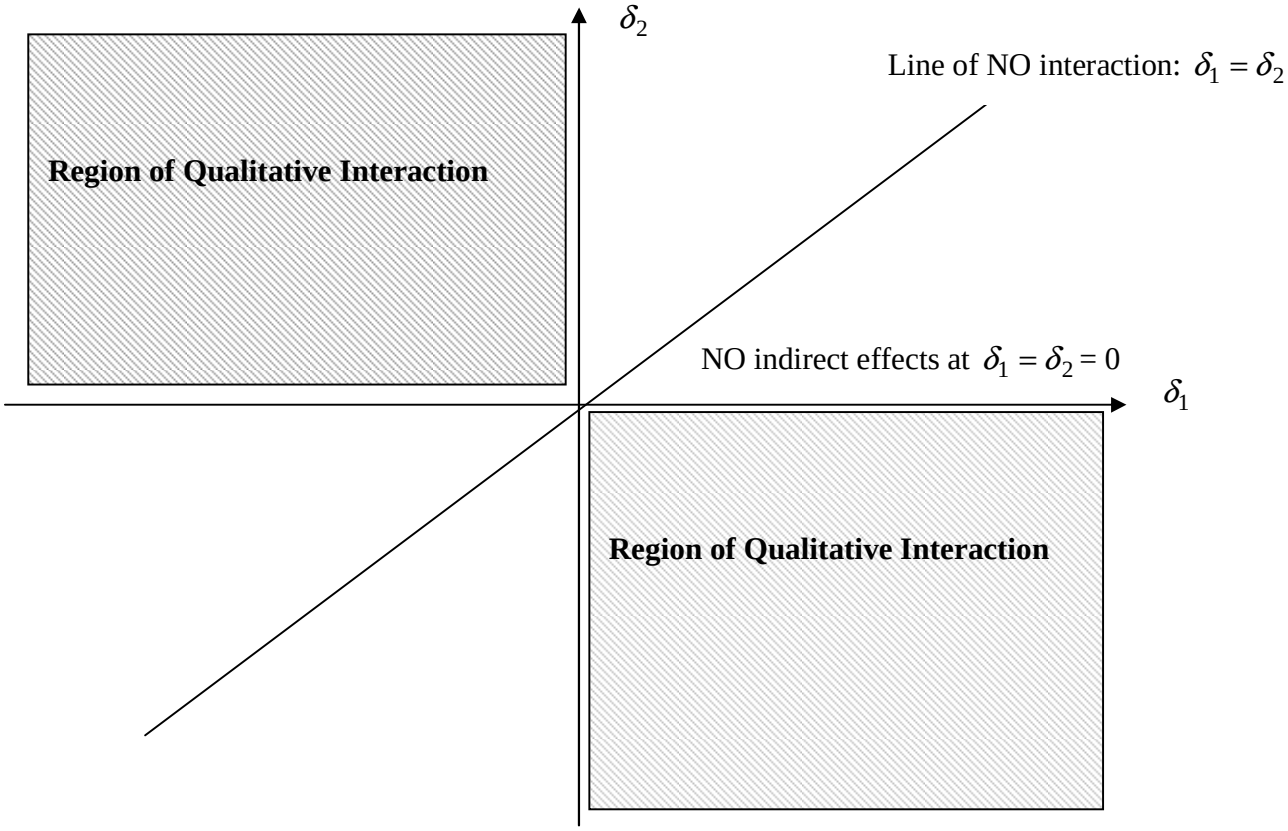



**Figure 5.1 The space of Indirect Effects for *s* =2 subsets**

## 5.2 Notation and Assumptions

Let $\delta_i = \beta_i \mu_{D_{Z_i}}$ for $i = 1, 2, ..., s$, be the true mean indirect effect in the $i^{th}$ subset, where $\beta_i$ is true slope for the $i^{th}$ subset and $\mu_{D_{Z_i}} = \mu_{Z_i^{(1)}} - \mu_{Z_i^{(0)}}$ is the true mean treatment effect on the mediator variable for the $i^{th}$ subset. We also assume that we have an estimate of the indirect effect in the $i^{th}$ subset, $\hat{\delta}_i = \hat{\beta}_i \hat{\mu}_{D_{Z_i}} = \hat{\beta}_i (\hat{\mu}_{Z_i^{(1)}} - \hat{\mu}_{Z_i^{(0)}})$ for $i = 1, 2, ..., s$.

The tests for qualitative interactions assume that the $\{\hat{\beta}_i\}$ are independent and normally distributed with mean $\beta_i$ and variance $\sigma^2_{\hat{\beta}_i}$ and $\{\hat{\mu}_{D_{Z_i}}\}$ are also independent and normally distributed with mean $\mu_{D_{Z_i}}$ and variance $\sigma^2_{\hat{\mu}_{D_{Z_i}}}$. We also assume that $\hat{\beta}_i$ and $\hat{\mu}_{D_{Z_i}}$ are

independent statistics. Though this later assumption may not hold in all situations, we use it here to begin the process of investigating a QI on indirect effects across subsets.

## 5.3 The Likelihood Ratio Tests (LRT)

We discuss the test for the null hypothesis of no qualitative interaction for three different cases.

(1). Assume that there is no qualitative interaction on $\beta_i$ and the qualitative interaction on $\delta_i = \beta_i \mu_{D_{Z_i}}$ depends on the qualitative interaction on $\mu_{D_{Z_i}}$ across subsets.

(2). Assume that there is no qualitative interaction on $\mu_{D_{Z_i}}$, and the qualitative interaction on $\delta_i = \beta_i \mu_{d_{Z_i}}$ depends on the qualitatve interaction on $\beta_i$ across subsets.

(3). Test whether there is a qualitative interaction on $\delta_i = \beta_i \mu_{D_{Z_i}}$ across subsets.

We discuss each case as follows:

### *5.3.1 Case (1): The QI on $\mu_{D_{Z_i}}$ across subsets*

Based on the likelihood ratio test of Gail and Simon (1985), here we use it for the treatment effect on the mediator variable. We assume that within each subset $i$, $n_{i1}$ units are randomly assigned to treatment $T=1$ and the other $n_{i2}$ units assigned to $T=0$ so that $n_{i1}+n_{i2}=n$.

Without loss of geneality, we assume that the first $n_{i1}$ units in the $i$ th subset recieves $T=1$ and the second recieves $T=0$. The treatment effect on mediator in each subset is $\mu_{D_{Z_i}} = \mu_{Z_i^{(1)}} - \mu_{Z_i^{(0)}}$, which is estimated by $\hat{\mu}_{D_{Z_i}} = (\overline{Z}_i^{(1)} - \overline{Z}_i^{(0)})$, $\overline{Z}_i^{(1)} = \frac{1}{n_{i1}} \sum_{k=1}^{n_{i1}} Z_{ik}^{(1)}$ , $\overline{Z}_i^{(0)} = \frac{1}{n_{i2}} \sum_{k=n_{i1}+1}^{n} Z_{ik}^{(0)}$ and the variance of $\hat{\mu}_{D_{Z_i}}$ is $\sigma^2_{\hat{\mu}_{D_{Z_i}}} = \sigma_i^2 \left( \frac{1}{n_{i1}} + \frac{1}{n_{i2}} \right)$.

As in Gail and Simon (1985), we assume that $\sigma_i^2$ is known in order to simplify exposition. In order to carryout the test, $\sigma_i^2$ is replaced by a consistent estimate, that is, it is estimated by the sample pooled variance for each subset, $s_i^2$ where

$s_i^2 = \dfrac{(n_{i1}-1)S_{i1}^2 + (n_{i2}-1)S_{i2}^2}{(n_{i1}+n_{i2}-2)}$, where $S_{i1}^2 = \dfrac{1}{(n_{i1}-1)}\sum_{k=1}^{n_{i1}}(Z_{ik}^{(1)} - \overline{Z}_i^{(1)})^2$

and $S_{i2}^2 = \dfrac{1}{(n_{i2}-1)}\sum_{k=n_{i1}+1}^{n}(Z_{ik}^{(0)} - \overline{Z}_i^{(0)})^2$ for $i = 1,2,...,s.$

The null hypothesis of no QI is $H_0 : \Delta \in O^+ \cup O^-$, where $\Delta = (\mu_{D_{Z_1}}, \mu_{D_{Z_2}}, ..., \mu_{D_{Z_s}})$,

$O^+ = \{\Delta : \mu_{D_{Z_i}} \geq 0 \text{ for all } i\}$ and $O^- = \{\Delta : \mu_{D_{Z_i}} \leq 0 \text{ for all } i\}$, $O^+$ and $O^-$ are the positive and negative orthants respectively, so $H_0$ specifies that all $\mu_{D_{Z_i}}$ are positive or they are all negative. The Likelihood Ratio Test (LRT) of this hypothesis is based on the statistic

$$\lambda(\underline{\hat{\mu}_{D_{Z_i}}}) = \frac{\max\limits_{\Delta \in O^+ \cup O^-} \exp \sum_{i=1}^{s}\left[ -\left(\hat{\mu}_{D_{Z_i}} - \mu_{D_{Z_i}}\right)^2 \Big/ 2\sigma^2_{\hat{\mu}_{D_{Z_i}}} \right]}{\max\limits_{\Delta} \exp \sum_{i=1}^{s}\left[ -\left(\hat{\mu}_{D_{Z_i}} - \mu_{D_{Z_i}}\right)^2 \Big/ 2\sigma^2_{\hat{\mu}_{D_{Z_i}}} \right]}.$$

In this denominator the maximization is unrestricted and therefore, the denomenator maximum occurs at $\mu_{D_{Z_i}} = \hat{\mu}_{D_{Z_i}}$ for $i = 1,2,...,s$, and the denominator equals 1. The likelihood ratio test is thus, $\max\limits_{\Delta \in O^+ \cup O^-} \exp \sum_{i=1}^{s}\left[ -\left(\hat{\mu}_{D_{Z_i}} - \mu_{D_{Z_i}}\right)^2 \Big/ 2\sigma^2_{\hat{\mu}_{D_{Z_i}}} \right] < k,$ (1)

where $k$ is chosen to ensure that the rejection region does not exceed significance level $\alpha$ for any point in the null space, $O^+ \cup O^-$. Expression (1) can be further simplified. The inequality (1) is equivelent to the simultaneous inequalities

$$\min_{\Delta \in O^-} \sum_{i=1}^{s}\left(\hat{\mu}_{D_{Z_i}} - \mu_{D_{Z_i}}\right)^2 / \sigma^2_{\hat{\mu}_{D_{Z_i}}} > c \quad (2)$$

and

$$\min_{\Delta \in O+} \sum_{i=1}^{s}\left(\hat{\mu}_{D_{Z_i}} - \mu_{D_{Z_i}}\right)^2 / \sigma^2_{\hat{\mu}_{D_{Z_i}}} > c, \quad (3)$$

where $c = -2\log(k)$. Thus, one rejects the null hypothesis if $\hat{\mu}_{D_Z} = (\hat{\mu}_{D_{Z_1}}, \hat{\mu}_{D_{Z_2}}, ..., \hat{\mu}_{D_{Z_s}})$ is far away from both $O^+$ and $O^-$, with distance defined by the inverse variance metric. The minimum value in (2) occurs for $\mu_{D_{Z_i}} = \hat{\mu}_{D_{Z_i}}$ if $\hat{\mu}_{D_{Z_i}} \leq 0$ and for $\mu_{D_{Z_i}} = 0$ otherwise for all $i = 1,2,...,s.$

Similarly, the minimum value in (3) occurs for $\mu_{D_{Z_i}} = \hat{\mu}_{D_{Z_i}}$ if $\hat{\mu}_{D_{Z_i}} \geq 0$ and for $\mu_{D_{Z_i}} = 0$ otherwise for all $i = 1,2,...,s$. Consequently, the likelihood ratio test of Gail and Simon (1985) rejects $H_0$ if both

$$Q^- = \sum \left( \hat{\mu}^2_{D_{Z_i}} \Big/ \sigma^2_{\hat{\mu}_{D_{Z_i}}} \right) I(\hat{\mu}_{D_{Z_i}} > 0) > c \tag{4}$$

and

$$Q^+ = \sum \left( \hat{\mu}^2_{D_{Z_i}} \Big/ \sigma^2_{\hat{\mu}_{D_{Z_i}}} \right) I(\hat{\mu}_{D_{Z_i}} < 0) > c, \tag{5}$$

where $I(\hat{\mu}_{D_{Z_i}} > 0) = 1$ if $\hat{\mu}_{D_{Z_i}} > 0$ and 0 otherwise, and $I(\hat{\mu}_{D_{Z_i}} < 0) = 1$ if $\hat{\mu}_{D_{Z_i}} < 0$ and 0 otherwise. The quantities $Q^+$ and $Q^-$ are minimum values of $\sum (\hat{\mu}_{D_{Z_i}} - \mu_{D_{Z_i}})^2 / \sigma^2_{\hat{\mu}_{D_{Z_i}}}$ over $O^+$ and $O^-$ respectively, and the LRT can be expressed as $\min(Q^+, Q^-) > c$. The challenge is to calculate value of $c$ such that for all $\Delta \in O^+ \cup O^-$, the probability that (4) and (5) are both satisfied is no greater than the significance level, $\alpha$.

The critical value $c$, which may be obtained from Table 1 of Gail and Simon (1985), is chosen so that the test has level at most $\alpha$ under any $\Delta$ for which there is no crossover interaction. The Table 5.1 is extracted from a table in Gail and Simon (1985) and it shows the values of $c$ corresponding to significance levels 0.05 and 0.1 for the test. If Gail and Simon's test rejects the null hypothesis, then no conclusion about the overall treatment effect on mediator difference can be made. If Gail and Simon's test does not reject the null hypothesis, then we do not have strong evidence to conclude whether the subset interaction is quantitative or qualitative.

**Table 5.1 A part of Gail and Simon (1985) table 1 ( Critical values ( c ) for the likelihood ratio test** $\min(Q^+, Q^-) > c$ **)**

| Number of groups | Significance level | |
|---|---|---|
| | **0.1** | **0.05** |
| 2 | 1.64 | 2.71 |
| 3 | 2.95 | 4.23 |
| 6 | 5.84 | 7.48 |
| 10 | 9.02 | 10.99 |
| 16 | 13.33 | 15.66 |
| 20 | 16.04 | 18.57 |
| 25 | 19.34 | 22.09 |
| 30 | 22.55 | 25.50 |

This table was constructed by solving the formula $\sum_{j=1}^{s-1} B(j;\ n = s-1,\ p = 0.5)[1 - F_j(c)]$, where $B(j; n, p)$ is the binomial probability mass function with index *n* and parametre *p*, and $F_j(*)$ is the central chi-square distribution with *j* degrees of freedom, for the value of *c* such that, for fixed number of groups *s* and significance level $\alpha$, the quantity $\sum_{j=1}^{s-1} B(j;\ n = s-1,\ p = 0.5)[1 - F_j(c)] = \alpha$ ( Gail and Simon, 1985).

### *5.3.2 Case (2): The QI on $\beta_i$ across subsets*

We assume that there is no QI on $\mu_{D_{Z_i}}$. The distribution of the estimated mean slope of a subset, $\hat{\beta}_i$, is $\hat{\beta}_i \sim N(\beta_i, \sigma^2_{\hat{\beta}_i})$. Since $\hat{\beta}_i$ satisfies all the assumptions in Gail and Simon's test, we adopt Gail and Simon (1985) LRT to check whether there is a qualitative interaction on mean slopes, $\beta_i$ or not.

Typically, as in Baron & Kenny the $\hat{\beta}_i$ are obtained in a clinical trial by estimating the simple regression equation in *s* subsets :

$Y_{ij} = \alpha_i + \beta_i Z_{ij} + \gamma_i T + \varepsilon'_{ij}$ *for* $i = 1,2,...,s$ ; $j = 1,2,...,n_i$, where $\beta_i$ is the mean slope from the *i*th subset and the errors $\varepsilon_{ij} \sim N(0, \sigma_i^2)$. The null hypothesis of no QI is $H_0 : \Delta \in O^+ \cup O^-$, where $\Delta = (\beta_1, \beta_2, ..., \beta_s)$ , $O^+ = \{\Delta : \beta_i \geq 0 \text{ for all } i\}$ and $O^- = \{\Delta : \beta_i \leq 0 \text{ for all } i\}$, $O^+$ and $O^-$ are the positive and negative orthants respectively.

The LRT of this hypothesis is based on the statistic

$$\lambda(\underline{\hat{\beta}}) = \frac{\max\limits_{\Delta \in O^+ \cup O^-} \exp \sum_{i=1}^{s} \left[ -\left(\hat{\beta}_i - \beta_i\right)^2 \Big/ 2\sigma^2_{\hat{\beta}_i} \right]}{\max\limits_{\Delta^-} \exp \sum_{i=1}^{s} \left[ -\left(\hat{\beta}_i - \beta_i\right)^2 \Big/ 2\sigma^2_{\hat{\beta}_i} \right]}.$$

In the denominator, the maximization is unrestricted. Consequently the denominator maximum occurs $\beta_i = \hat{\beta}_i$ for $i = 1, 2, ..., s$ and the denominator equals 1. The likelihood ratio test of Gail and Simon (1985) rejects $H_0$ if both

$$Q^- = \sum \left( \hat{\beta}_i^2 \Big/ \sigma^2_{\hat{\beta}_i} \right) I(\hat{\beta}_i > 0) > c \tag{6}$$

and

$$Q^+ = \sum \left( \hat{\beta}_i^2 \Big/ \sigma^2_{\hat{\beta}_i} \right) I(\hat{\beta}_i < 0) > c), \tag{7}$$

where the quantities $Q^+$ and $Q^-$ are minmum values of $\sum (\hat{\beta}_i - \beta_i)^2 / \sigma^2_{\bar{B}_i}$ over $O^+$ and $O^-$ respectively, and the LRT can be expressed as $\min(Q^+, Q^-) > c$. This test and the distributional assumptions are similar to what Gail and Simon discussed and therefore we use Table 5.1 to find c such that for all $\Delta \in O^+ \cup O^-$, the probability that (6) and (7) are both satisfied is no greater than the significance level, $\alpha$.

### *5.3.3 Case (3): The QI on $\delta_i = \beta_i \mu_{d_{Z_i}}$ across subsets*

Before we construct the LRT to check whether there is a qualitative interaction on $\delta_i = \beta_i \mu_{D_{Z_i}}$, we need to establish the probability distribution of $\hat{\delta}_i = \hat{\beta}_i \hat{\mu}_{D_{Z_i}}$. Since we assume that $\hat{\mu}_{D_{Z_i}} \sim N(\mu_{D_{Z_i}}, \sigma^2_{\hat{\mu}_{D_{Z_i}}})$ and $\hat{\beta}_i \sim N(\beta_i, \sigma^2_{\hat{\beta}_i})$ are independent, the probability density function

of the product of these two non-zero mean normal variables, $\hat{\delta}_i = \hat{\beta}_i \hat{\mu}_{D_{Z_i}}$, is not known. If this distribution was normal, we would be able to use the above LRT to test the null hypotheses that there is no QI on $\delta_i = \beta_i \mu_{D_{Z_i}}$. Now, we discuss another method of testing the null hypotheses that there is no QI on $\delta_i = \beta_i \mu_{D_{Z_i}}$.

Before introducing an appropriate test, we discuss the possibilities that when and where the QI on $\delta_i = \beta_i \mu_{D_{Z_i}}$ across subsets occurs. If there are no QIs on both $\beta_i$ and $\mu_{D_{Z_i}}$ then there is no QI on $\delta_i = \beta_i \mu_{D_{Z_i}}$. But no QI on $\delta_i = \beta_i \mu_{D_{Z_i}}$ does not imply that there are no QIs on $\beta_i$ and $\mu_{D_{Z_i}}$. That is,

$$\left.\begin{matrix} No\ \ QI\ \ on\ \beta_i \\ No\ \ QI\ \ on\ \mu_{D_{Z_i}} \end{matrix}\right\} \Rightarrow No\ \ QI\ \ on\ \ \beta_i \mu_{D_{Z_i}} \text{ but } No\ \ QI\ \ on\ \ \beta_i \mu_{D_{Z_i}} \not\Rightarrow \begin{cases} No\ \ QI\ \ on\ \beta_i \\ No\ \ QI\ \ on\ \mu_{D_{Z_i}} \end{cases}.$$

Thus, instead of testing the null hypotheses that there is no QI on $\beta_i \mu_{D_{Z_i}}$ we discuss the test for the null hypotheses that there is no QI on both $\beta_i$ and $\mu_{D_{Z_i}}$. Hence, we formulate the required hypotheses as follows: $H_0 : (No\ \ QI\ \ on\ \ \mu_{D_Z}) \cap (No\ \ QI\ \ on\ \ \beta)$ versus $H_1 : QI\ \ on\ \ either\ \mu_{D_Z}\ \ or\ \ \beta\ \ or\ \ both.$

Roy's (1953) well-known Union-Intersection method (UIT) can be used to test the null hypotheses given above. Roy's principle of construction of tests for the case when the null hypothesis $H_0$ consists of the simultaneous occurrence of several disjoint sub-hypotheses and is represented as $\bigcap_{i=1}^{g} H_{0i}$ is the Union-Intersection (UI) principle (Roy, 1953), where $g$ is the number of individual tests. Here $g = 2$. Cassella and Berger (2002) have proved the following relationships between the overall LRT and the UIT .

Consider a UIT test $H_0 : \theta \in \Theta_0$ *versus* $H_1 : \theta \in \Theta_0^c$, where $\Theta_0 = \bigcap_{\gamma \in \Gamma} \Theta_\gamma$. Let $\lambda_\gamma(\underline{x})$ be the LRT statistic for testing $H_0 : \theta \in \Theta_\gamma$ *versus* $H_1 : \theta \in \Theta_\gamma^c$ and let $\lambda(\underline{x})$ be the LRT statistic for

$H_0 : \theta \in \Theta_0$ *versus* $H_1 : \theta \in \Theta_0^c$. Define $T(\underline{x}) = \inf_{\gamma \in \Gamma} \lambda_\gamma(\underline{x})$, and form the UIT with rejection region $\{\underline{x} : \lambda_\gamma(\underline{x}) < c \text{ for some } \gamma \in \Gamma\} = \{\underline{x} : T(\underline{x}) < c\}$. Also consider the usual LRT with rejection region $\{\underline{x} : \lambda_\gamma(\underline{x}) < c\}$. Then

*a.* $T(\underline{x}) > \lambda(\underline{x})$ *for every* $\underline{x}$.

*b. If* $\beta_T(\theta)$ *and* $\beta_\lambda(\theta)$ *are the power functions for the test based on T and* $\lambda$*, respectively, then* $\beta_T(\theta) \leq \beta_\lambda(\theta)$ *for every* $\theta \in \Theta$.

*c. If the LRT is a level* $\alpha$ *test, then the UIT is a level* $\alpha$ *test*.

Based on Roy's UI principle below is the procedure of testing our null hypothesis $H_0 : (No\ QI \text{ on } \mu_{D_Z}) \cap (No\ QI \text{ on } \beta)$ versus $H_1 : QI$ on either $\mu_{D_Z}$ or $\beta$ or both. Based on Gail and Simon (1985), the LRT of $H_{0D} : No\ QI$ on $\mu_{D_Z}$ vs $H_{1D} : QI$ on $\mu_{D_Z}$ is rejected if $\min(Q_D^+, Q_D^-) > c$, where $Q_D^- = \sum \left( \hat{\mu}_{D_{Z_i}}^2 / \sigma_{\hat{\mu}_{D_{Z_i}}}^2 \right) I(\hat{\mu}_{D_{Z_i}} > 0) > c$ and $Q_D^+ = \sum \left( \hat{\mu}_{D_{Z_i}}^2 / \sigma_{\hat{\mu}_{D_{Z_i}}}^2 \right) I(\hat{\mu}_{D_{Z_i}} < 0) > c$. Similarly, the LRT of $H_{0B} : No\ QI$ on $\beta$ vs $H_{1B} : QI$ on $\beta$ is rejected if $\min(Q_B^+, Q_B^-) > c$, where $Q_B^- = \sum \left( \hat{\beta}_i^2 / \sigma_{\hat{\beta}_i}^2 \right) I(\hat{\beta}_i > 0) > c$ and $Q_B^+ = \sum \left( \hat{\beta}_i^2 / \sigma_{\hat{\beta}_i}^2 \right) I(\hat{\beta}_i < 0) > c$. Let $Q_{D\min} = \min(Q_D^-, Q_D^+)$ and $Q_{B\min} = \min(Q_B^-, Q_B^+)$. Thus, the Union-Intersection test of $H_0$ *vs* $H_1$ formed from these two LRTs is reject $H_0$ if $Q_{D\min} > c$ or $Q_{B\min} > c$.

## 5. 4 The Simulation Study for Qualitative Interactions

We test the null hypotheses using the generated observed data for three cases discussed in the previous section. Since we do not have a real data set, a data set was generated to illustrate the QI across subsets assuming a balanced study design with $n = 400/2s$ subjects per treatment per subset, where $s = 20$ is the number of disjoint subsets which consist of observed *Y* , *Z* and *T*.

The first three plots in Figure 5.2 show how QI occurs on $\mu_{D_{Z_i}}$ across subsets (top left), on $\beta_i$ cross subsets (top right) and on $\delta_i = \beta_i \mu_{D_{Z_i}}$ across subsets(bottom left). The plots of 20 subgroups are represented.

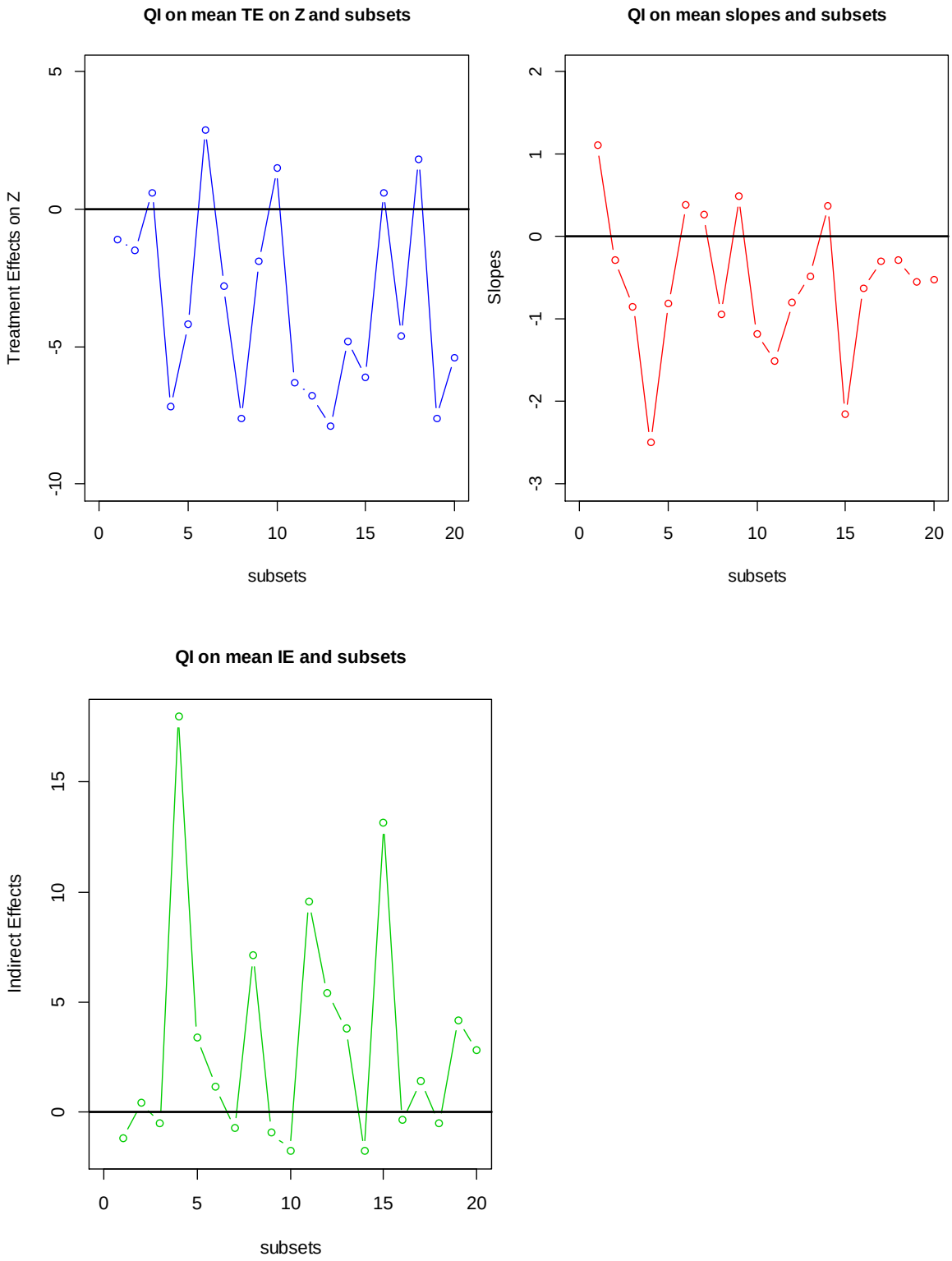



**Figure 5.2 QI on slopes, treatment effects on the mediator, indirect effect and subsets**

To check whether there is a significant qualitative interaction we use Gail and Simon (1985) test for the hypotheses mentioned in the section 5.3 at the 5% level of significance. Results are given in the table 5.2.

(1) First we discuss the hypothesis test in section 5.3.1, that is

$H_0$ : there is no $QI$ on $\mu_{D_Z}$ versus $H_0$ : there is $QI$ on $\mu_{D_Z}$.

In our simulation study, we compute $Q_D^- = \sum \left( \hat{\mu}_{D_{Z_i}}^2 \Big/ \sigma_{\hat{\mu}_{D_{Z_i}}}^2 \right) I(\hat{\mu}_{D_{Z_i}} > 0) = 31.8$ and $Q_D^+ = \sum \left( \hat{\mu}_{D_{Z_i}}^2 \Big/ \sigma_{\hat{\mu}_{D_{Z_i}}}^2 \right) I(\hat{\mu}_{D_{Z_i}} < 0) = 626$, where $I(\hat{\mu}_{D_{Z_i}} > 0) = 1$ *if* $\hat{\mu}_{D_{Z_i}} > 0$ and 0 otherwise, $I(\hat{\mu}_{D_{Z_i}} < 0) = 1$ *if* $\hat{\mu}_{D_{Z_i}} < 0$ and 0 otherwise, and the LRT can be expressed as $\min(Q_D^+, Q_D^-) = 31.8$. From Table 5.1, we get the critical value at 5% level of significance is 18.57. That is, reject the null hypotheses since $\min(Q_D^+, Q_D^-) = 31.8 > 18.57$ implies there is a significant qualitative interaction on $\mu_{D_Z}$.

(2) Then we discuss the hypothesis test in section 5.3.2, that is

$H_0$ : there is no $QI$ on $\beta$ versus $H_0$ : there is $QI$ on $\beta$.

In our simulation study, we compute $Q_B^- = \sum \left( \hat{\beta}^2 \Big/ \sigma_{\hat{\beta}}^2 \right) I(\hat{\beta} > 0) = 9.26$ and $Q_B^+ = \sum \left( \hat{\beta}^2 \Big/ \sigma_{\hat{\beta}}^2 \right) I(\hat{\beta} < 0) = 576$, where $I(\hat{\beta} > 0) = 1$ *if* $\hat{\beta} > 0$ and 0 otherwise, and $I(\hat{\beta} < 0) = 1$ *if* $\hat{\beta} < 0$ and 0 otherwise, and the LRT can be expressed as $\min(Q_B^+, Q_B^-) = 9.26$. From Table 5.1, we get the critical value at 5% level of significance is 18.57. That is, do not reject the null hypotheses since $\min(Q_B^+, Q_B^-) = 9.26 < 18.57$ implies there is no significant qualitative interaction on $\beta$.

(3) The third test would be the union-intersection test, $H_0 : (No\ QI\ on\ \mu_{D_Z}) \cap (No\ QI\ on\ \beta)$ versus $H_1 : QI\ on\ either\ \mu_{D_Z}\ or\ \beta\ or\ both.$ The LRT of $H_{0D} : No\ QI\ on\ \mu_{D_Z}$ vs $H_{1D} : QI\ on\ \mu_{D_Z}$ is rejected since $\min(Q_D^+, Q_D^-) = 31.8 > 18.57$ implies there is a significant qualitative interaction on $\mu_{D_Z}$.

**Table 5.2 QI on slopes, TE on the mediator, indirect effect across subsets for 20 subsets**

| *Subset* | $\hat{\mu}_{D\,Z_i}$ | $s^2_{\hat{\mu}_{D\,Z_i}}$ | $\hat{\mu}^2_{D_{Z_i}} \big/ s^2_{\hat{\mu}_{D\,Z_i}}$ | $\hat{\beta}_i$ | $s^2_{\hat{\beta}_i}$ | $\hat{\beta}_i^2 \big/ s^2_{\hat{\beta}_i}$ |
|---|---|---|---|---|---|---|
| 1 | -0.3878 | 0.563 | 0.267 | 0.938 | 0.4314 | 2.038 |
| 2 | -2.0137 | 0.667 | 6.08 | -1.180 | 0.2254 | 6.182 |
| 3 | 2.9754 | 0.708 | 12.50 | 0.436 | 0.2526 | 0.754 |
| 4 | -5.4451 | 0.757 | 39.20 | -3.076 | 0.1194 | 79.257 |
| 5 | -4.7614 | 0.956 | 23.70 | -1.832 | 0.1247 | 26.931 |
| 6 | 2.2361 | 0.565 | 8.85 | 0.986 | 0.1995 | 4.872 |
| 7 | -4.1573 | 0.596 | 29.00 | -0.810 | 0.1983 | 3.312 |
| 8 | -7.3039 | 0.504 | 106.00 | -1.991 | 0.0422 | 93.948 |
| 9 | -2.2101 | 0.373 | 13.10 | -0.219 | 0.3542 | 0.136 |
| 10 | 1.4775 | 1.419 | 1.54 | -1.002 | 0.1851 | 5.429 |
| 11 | -5.8544 | 1.627 | 21.10 | -2.026 | 0.0975 | 42.114 |
| 12 | -6.3345 | 0.634 | 63.20 | -2.005 | 0.0463 | 86.771 |
| 13 | -6.2663 | 0.803 | 48.90 | -0.871 | 0.0904 | 8.394 |
| 14 | -4.4055 | 1.026 | 18.90 | -0.333 | 0.0329 | 3.384 |
| 15 | -6.0050 | 0.476 | 75.80 | -3.094 | 0.0852 | 112.373 |
| 16 | 0.0284 | 1.334 | 0.0006 | -1.146 | 0.2867 | 4.579 |
| 17 | -4.9267 | 0.711 | 34.20 | -1.345 | 0.0730 | 24.790 |
| 18 | 2.5840 | 0.751 | 8.89 | 0.661 | 0.2731 | 1.600 |
| 19 | -7.4993 | 0.484 | 116.00 | -1.298 | 0.0399 | 42.253 |
| 20 | -6.3374 | 1.319 | 30.50 | -1.237 | 0.0426 | 35.906 |
| Test for QI on slopes: $Q_B^- = 9.26$ | | | $Q_B^+ = 576$ | | $Q_{B\min} = \min(Q_B^-, Q_B^+) = 9.26$ | |
| Test for QI on TE on Z: $Q_D^- = 31.8$ | | | $Q_D^+ = 626$ | | $Q_{D\min} = \min(Q_D^-, Q_D^+) = 31.8$ | |

In contrast, the LRT of $H_{0B}: No\ QI\ on\ \beta$ vs $H_{1B}: QI\ on\ \beta$ is not rejected as $\min(Q_B^+, Q_B^-) = 9.26 < 18.57$ implies there is no significant qualitative interaction on $\beta$. Thus, the Union-Intersection test of $H_0: there\ is\ no\ QI\ on\ \beta$ versus $H_0: there\ is\ QI\ on\ \beta$ is rejected as $\min(Q_D^+, Q_D^-) > c$.

### *5.4.1 Type I Error Rate*

Monte Carlo simulation was conducted 1000 times to estimate the type I error rate for the above UIT test under the null hypotheses, $H_0: (No\ QI\ on\ \mu_{D_Z}) \cap (No\ QI\ on\ \beta)$. Two LRT tests , $H_0: No\ QI\ on\ \mu_{D_Z}\ versus\ H_1: QI\ on\ \mu_{D_Z}$ and $H_0: No\ QI\ on\ \beta\ versus\ H_1: QI\ on\ \beta$ were conucted and found the number of times we reject the null hypotheses of UIT test. Figure 5.3 (a) shows the test statistics for both LRT test for 1000 Monte Carlo simulations. At 5% level of significance with 20 subgroups, the critical value would be 18.57 , vertical line in the figure 5.3(a). Type I error rate can be estimated. Since there are 14 and 40 unique LRT tests are rejected for $H_0: No\ QI\ on\ \mu_{D_Z}\ versus\ H_1: QI\ on\ \mu_{D_Z}$ and

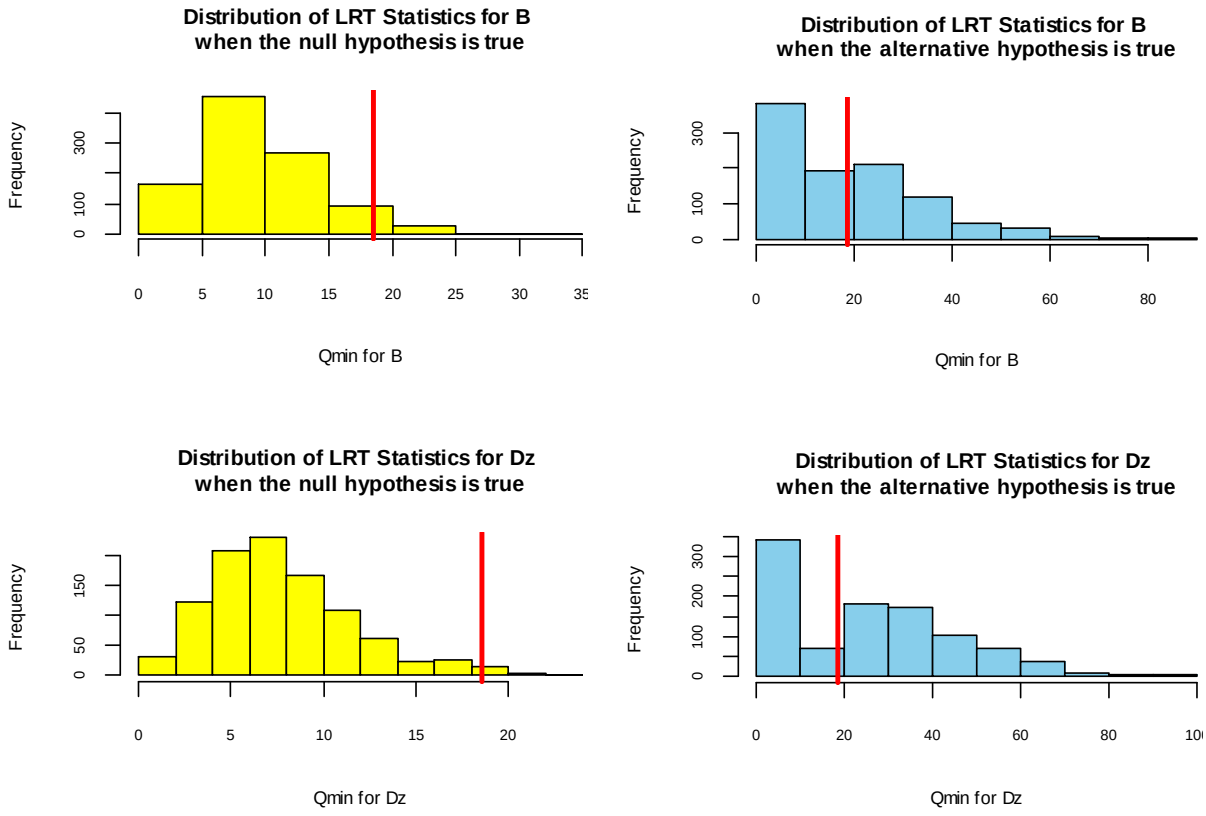
$H_0: No\ QI\ on\ \beta\ versus\ H_1: QI\ on\ \beta$, the estimated type I error rates are 0.014 and 0.040, respectively. Thus, the type I error rate for the test $H_0: (No\ QI\ on\ \mu_{D_Z}) \cap (No\ QI\ on\ \beta)$ under UIT would be 0.054.

### *5.4.2 Power of the test*

Under the alternative hypothesis the data were generated so that there is QI on $\mu_{D_Z}$ and $\beta$ The signal indicating the alternative hypothesis is as sown in Figure 5.2. Monte Carlo simulation was conducted 1000 times to estimate the power of the UIT test. Two LRT tests $, H_0: No\ QI\ on\ \mu_{D_Z}\ versus\ H_1: QI\ on\ \mu_{D_Z}$ and

$H_0: No\ QI\ on\ \beta\ versus\ H_1: QI\ on\ \beta$ were conucted and found the number of times we reject the null hypotheses of UIT test. Figure 5.3 (b) shows the test statistics for both LRT test when the alternative is true, for 1000 Monte Carlo simulations. At the 5% level of significance with 20 subgroups, the critical value would be 18.57, the vertical line in the figure 5.3(b). Since there are 607 and 460 LRT tests are rejected for $H_0: No\ QI\ on\ \mu_{D_Z}\ versus\ H_1: QI\ on\ \mu_{D_Z}$ and

$H_0 : No\ \ QI\ \ on\ \beta\ versus\ \ H_1 : QI\ \ on\ \beta$, the estimated power of the tests are 0.607 and 0.460, respectively. But, the total number of tests that reject the null hypothesis of UIT is 645. Thus, the power of the test for the test $H_0 : (No\ \ QI\ \ on\ \ \mu_{D_Z}) \cap (No\ \ QI\ \ on\ \ \beta)$ would be 0.645.




**(a)** **(b)**

**Figure 5.3 (a): LRT test statistics when the null hypothesis is true (b): LRT test statistics when the alternative hypothesis is true**

# Chapter 6 - Mediating Effects for Terminal Measures

In mediation analysis, situations can arise where the mediator, $Z$, and the outcome, $Y$, cannot both be measured on an individual unit. Mouse and plant experiments are two examples where measurement of $Z$ requires terminating the animal or plant and $Y$ is to be measured at a later time. We refer to such data as terminal measures data. Another situation may be where one experiment focused on measurement of $Z$, and the second on measurement of $Y$, and interest is in combining the data sets to evaluate the mediating effect of $Z$ on $Y$. This simulation study discusses the issues involved in analyzing terminal measures data.

| Proposed mechanism |
|---|
| Growth retardation |
| Reduction of body fat content |
| Reduction of metabolic rate |
| Attenuation of oxidative damage |
| Enhancement of apoptosis |
| Inhibition of apoptosis |
| Enhancement of autophagy |
| Attenuation of inflammation |
| Reduction of body temperature |
| Elevation of physical activity |
| Reduction of plasma [glucose] |
| Attenuation of IGF1 signaling |
| Enhancement of IGF1 signaling |
| Enhancement of insulin signaling |
| Attenuation of insulin signaling |
| Attenuation of TOR signaling |
| Hormesis |

**Table 6.1 The mechanism of life extension by CR.**

There are many studies exploring the mediators of Calorie Restriction (CR) effects on longevity (Stephen et al., 2003). The table 6.1 (Masoro, 2009) provides a listing of proposed mechanisms of action of CR. Some, like reduction in body temperature, can be measured without harming an animal or person. Others, like attenuation of oxidative damage or attenuation of inflammation in specific tissues, typically require killing an animal to take the measurement in the relative tissue.

This is also true of tissue-specific gene expression measurements and gene-methylation measurements. Hence it is vital to develop methods that can evaluate mediators requiring terminal measurements. The following are the two examples for terminal measures experiments:

Example (1): (Stephen et al., 2003)

A moderate CR regimen ; 20% reduction in calorie intake (treatment) decreases IGF-1; Insulin-like growth factor-1 (mediator), increases the ratio of apoptotic versus proliferating preneoplastic urothelial cells, and suppresses p-cresidine–induced bladder carcinogenesis (outcome) in p53-deficient mice. Weanling p53C mice were fed a diet containing 0.5% p-cresidine for 16 weeks

and then randomized to two groups: (1) ad libitum (AL) control diet plus implanted saline osmotic minipump; (2) CR (80% of AL diet intake) plus implanted saline osmotic minipump. After 4 weeks of diet treatment with implants, half of the mice were killed (to measure the mediator) , blood was collected, and their bladders were excised and analyzed histologically. The other half of the mice was used to observe the outcome.

Example (2): (Water Stress in Plants: Causes, Effects and Responses, Seyed Y. S. Lisar )
Water deficits in trees have an adverse effect on many of the tree's growth processes. Plant water stress, often times caused by drought, can have major impacts on plant growth and development. Plant growth under drought is influenced by altered photosynthesis, respiration, translocation, ion uptake, carbohydrates, nutrient metabolism, and hormones. Photosynthesis is particularly sensitive to the effects of water deficiency. Photosynthesis of higher plants decreases with the reduction in the relative water content (RWC) and leaf water potential. Lower photosynthesis rate is a usual effect of water stress in plants and has been attributed primarily to stomatal limitation and secondarily to metabolic impairment.

Most of the studies of mediation have presupposed the joint measurement $((Z^{(t)}, Y^{(t)}(Z^{(t)}))$ of mediator and outcome for all subjects, apart from limited missing data. As we discussed above some mediating variables cannot be assessed while the subject is alive. In human studies, the assessment of such a terminal measure requires waiting until the death of the patient. Further, human studies preclude the possibility of obtaining the terminating data at the time that is most suitable for the scientific question. Animal studies have the potential to better answer mediating questions by allowing measurement of the mediator at a targeted time.

The challenge raised by a design involving terminal measures is that each animal can be measured on either the mediator or the outcome, but not both. Figure 6.1 shows the potential outcomes for $N$ subjects in a study the usual observed data and then terminal measures data. For convenience, we use $Y^{(t)}(Z^{(t)}) = Y^{(t)}$ for $t = 0,1$.

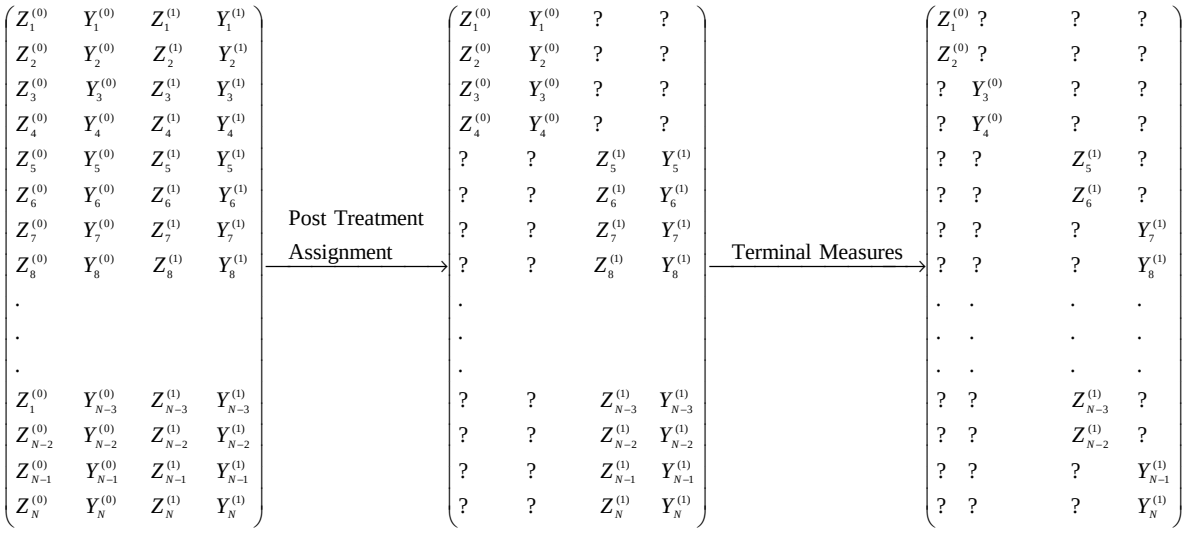



**Figure 6.1 Observable potential outcomes and the observed outcomes (terminal measures) for $N$ subjects for two sample completely randomized design.**

The set of *N* observable potential outcomes has the form given in Figure 6.1(left) that, after treatment assignment produces observed outcomes (centre) and then terminal measures of the form shown (right), and where the "?" represents an unobserved potential outcomes for a particular subject at the particular time under the treatment or control condition.

## 6.1 Data

We did not have a readily accessible terminal measures data set on which to base a simulation. Thus, a mice data set from two experiments, one where Z was measured and another where Y was measured were used to choose the parameters for the simulation study. Even though the diet, weight loss and the lifespan are observable outputs, we prepared this data set as a terminal measure data set. The description of the mouse data set is given below:

These mouse data sets were provided by Thomas E. Johnson & Brad A. Rikke, Institute for Behavioral Genetics, University of Colorado at Boulder Boulder, CO. One set of data was from an experiment to evaluate the effect of dietary restriction on lifespan, and a second data set evaluated the effect of dietary restriction on weight (as well as other variables). Two levels of diet were used for both experiments, DR(T=1)=dietary restriction and AL(T=0) = *ad libitum*, and

all mice from both experiments were female. After some data cleaning to eliminate documented atypical cases, the lifespan data set had a total of 881 mice from 42 different strains in two cohorts. The weight data set had 473 mice from 51 different strains in three cohorts. There were 33 strains that overlapped between the two data sets. In the lifespan data set, we define the primary outcome variable, Y = Lifespan in days. In the weight data set, we defined a candidate mediator variable to be Z = weight in grams between 7 and 16 weeks. It may be noted that the effect of the diet variable on this body weight measurement was already highly significant at this time point. Note that for individual mice in both studies, either a value of Z or Y is observed but not both. If the interest is to determine the mediating effect of Z on Y, then missing values of Z or Y must be imputed using a valid imputation model to use individual level models for analyzing the indirect effect of dietary restriction on lifespan through the variable Z. The strain information for the 33 strains that overlap between the two data sets was evaluated as a grouping variable, analogous to how one could use a litter variable in study. Of the 33 strains that overlapped, there were 315 mice in the weight data set and 692 mice in the lifespan data set that were from strains common to both experiments.

Within the weight data set, there was an average of 4.7 mice from each strain in each of the two diet groups. Within the lifespan data set, there was an average of 9.6 and 11.4 mice from each strain in the AL and DR diet groups, respectively. Figure 6.1 shows the distributions of the candidate mediator, weight, and the outcome variable, Lifespan, for each strain within the diet groups. One can see that the weight variable is less variable within strain than the lifespan variable. However, there is visual evidence that both variables differ across strains. To test this, analysis of variance (ANOVA) was conducted for each of the four data subsets shown in the four panels of Figure 6.1. These data were used as a framework for assessing parameter values in a simulation illustration to be described next.

The percentage of the total sums of squares of the weight variable within the AL and DR diets that is explained by the strain variable is 68% and 79% , respectively. The percentage of the total sums of squares of the lifespan variable within the AL and DR diets that is explained by the strain variable is 38% and 45%, respectively. The analyses show that within each diet group, the

strain variable is a highly statistically significant blocking variable for both of the outcomes, weight and lifespan, with a p-value near zero.

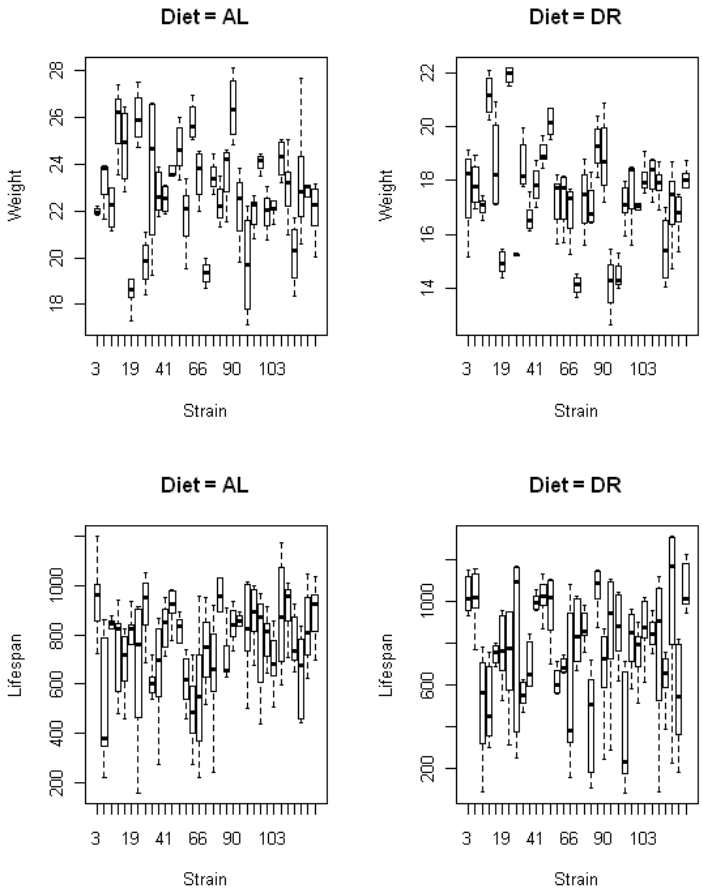



**Figure 6.2 Boxplots showing the distributions of the variables weight and lifespan for Each strain within the diet groups, AL and DR.**

## 6.2 Designs of the Study

A key challenge is to find methods suitable for a study design in which the mediating variables are terminal measures. In this study, we consider two experimental deigns; (1). strain as a blocking variable and (2) strain as an experimental unit. Here we discuss these two designs separately and estimate the mediating effects in a simulation.

### *6.2.1 Design 1: Mice as Experimental Units and Strains as Blocking Variable (Individual-level assessment of mediation).*

In the simulation, the diet level (DR or AL) was assigned to mice within strain so that each strain contains *n* mice. Strains act as a blocking variable and mice act as experimental units (Completely Randomized Block Design). This design has *n* mice with *s* strains. We assign two diet levels for each strain so that the design is balanced. Each strain contains *n*/2 replicates for each diet level and there are *s*- strains. The total number of mice would be *ns*.

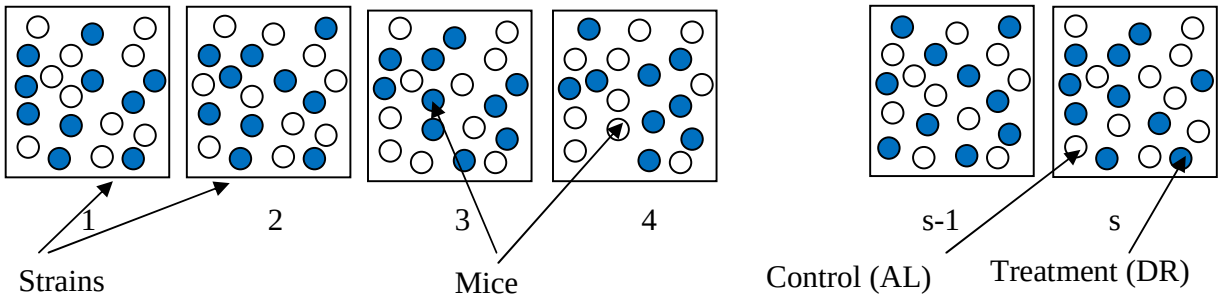



**Figure 6.3 Design 1: Treatment Assignment, open circles are AL (Control) and solid circles are DR (Treatment).**

### *6.2.2 Design 2: Strains as Experimental Units (Strain-level assessment of mediation)*

Strains were assigned for either treatment (DR) or control (AL) so that each strain contains *n*/2 mice, *s*-strains for the treatment & *s*-strains for the control. 2*s* strains act as experimental units. The total number of mice would be *ns*. This design is called a group randomized design. Group randomized trials are experiments in which the intervention occurs at the level of the group but observations are made on individuals within the groups. Group randomization is particularly useful when there is a high risk for contamination if group members are randomized as individuals. A group randomized trial is based on a multi-stage sampling technique.

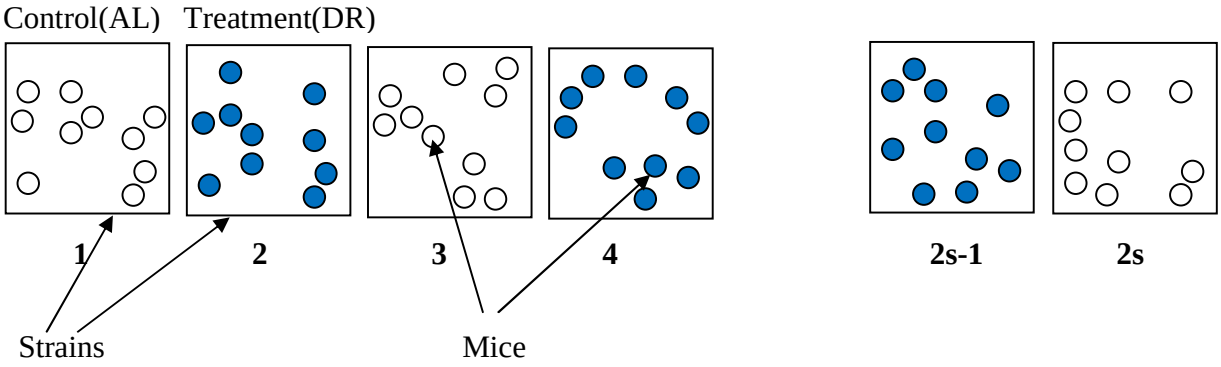



**Figure 6.4 Design 2: Treatment Assignment, open circles are AL and solid circles are DR.**

Two-stage sampling is a common practice in many disciplines. In two-stage sampling, the first stage refers to the primary sampling unit which is a cluster of objects, followed by a second stage where individual objects, or sub-units, are sub sampled from the cluster . A cluster, or primary sampling unit, is a natural grouping of objects that may have similar attributes. This study includes mice (sub-units) in a strain (cluster). This sampling scheme is a form of multilevel sampling and is also referred to as hierarchical or nested sampling. This study of a completely randomized design has *n/2* pseudoreplications (mice) per experimental unit (strain) with *s* replications. The term “pseudoreplication” was coined by Hurlbert (1984, 190-191) to refer to "the use of inferential statistics to test for treatment effects with data from experiments where either treatments are not replicated (though samples may be) or replicates are not statistically independent." The context of his paper was ecological field experiments, but pseudoreplication can occur in other contexts as well.

Replications are having more than one experimental unit with the same treatment. That is, each unit with the same treatment is called a replicate. True replication permits the estimation of variability within a treatment. Comparing two treatments, we assign each treatment on several units, then we can obtain some information about the variability of each treatment.

True replicates are often confused with repeated measures or with pseudoreplicates. The following illustrate some of the ways this can occur. A consequence of doing statistical inference using pseudoreplicates rather than true replicates is that variability will probably be underestimated.

The mice in the same strain are not independent observations, but rather are correlated. For example, in measuring mice weight, mice life time tends to aggregate with other mice of a similar age, so measuring one mouse from a strain provides information on the weights & life time of the other mice in the same strain. The error of ignoring that objects (mice) were sampled in clusters (strains) and treating the mice as a simple random sample is a form of pseudoreplication, specifically sacrificial pseudoreplication, which has been well documented, and has been a common mistake made by researchers. According to Hurlbert (1984), the major problem of pseudoreplication is that individual units (the mice) are treated as independent objects, when in fact, they are not (Susan, J., 2011).

In design 1, mice were randomly assigned to the two treatments in the strain, and will be considered as a design with replicates. In design 2, the strains were randomly assigned to the two treatments and mice within the strain can be considered as pseudoreplicates as they are inter-correlated.

## 6.3 Methodology

We compare analytical methods for estimating true mediating effect for design 1 & design 2. The design 1 includes completely randomized balanced design within strain. The complexity of design 2 is the lack of independence among the data. We assume that the strains contain the same numbers of mice and the number of strains per treatment level is also the same. The number of treatment levels (Treatment vs. Control) is fixed at 2.

### *6.3.1 Simulation Study*

Under each design, the potential outcomes, the observed data & the terminal measures are generated. Then the estimated mediating effects & the estimated standard error of estimate are calculated for several cases. For each design, 1000 samples were created.

For designs 1 & 2, the potential outcomes were generated for $s$-strains with $n$-individuals in each strain and $2s$-strains with $n/2$ individuals in each strain, respectively, so that the total number of mice was $ns$. For each design, the observed data were generated by removing counterfactuals and then from the observed data the terminal measures were generated by removing $Y$ & $Z$ randomly within strains. Here, the missing values were replaced by observed

averages of $Z$ & $Y$ within each strain for each treatment assignment and then the missing values were imputed using parametric bootstrapping. Initially set the true mediating effect as $\beta(\mu_{Z^{(1)}} - \mu_{Z^{(0)}})$ = (-1)(-5) = 5.

### *6.3.2 Steps of Generating Potential Outcomes, Observed Data and Terminal Measures*

**Design 1:** To generate potential outcomes of $Z$'s, we use the following steps:

1. Specify $\mu_{Z^{(0)}} = 22$, overall mean of $Z$ when $T = 0$ and $\mu_{Z^{(1)}} = 17$,overall mean of $Z$ when $T = 1$.
2. Specify $\sigma^2_{Z_\mu}$ and $\sigma^2_{Z_e}$ so that $\sigma^2_{Z_0} = \sigma^2_{Z_\mu} + \sigma^2_{Z_e}$ where $\sigma^2_{Z_\mu} = 4.5$, the variance of strain means of $Z$ when $T = 0$ and $\sigma^2_{Z_e} = 0.5$, the variance of $Z$ within the strain when $T = 0$.
3. Generate $a_1, a_2, ..., a_s \sim N(0, \sigma^2_{Z_\mu})$.
4. Generate $\mu_{Z_i^{(0)}} = \mu_{Z^{(0)}} + a_i$, the mean of $Z$ for the $i^{\text{th}}$ strain.
5. Generate $Z_{ij}^{(0)} = \mu_{Z_i^{(0)}} + \varepsilon_{ij}$ , the $j^{\text{th}}$ observation of $Z$ for the $i^{\text{th}}$ strain when $T = 0$, where $\varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$. Note that $Z_{ij}^{(0)} = \mu_{Z^{(0)}} + a_i + \varepsilon_{ij}$.
6. Generate $Z_{ij}^{(1)} = Z_{ij}^{(0)} + \tau$, the $j^{\text{th}}$ observation of $Z$ for the $i^{\text{th}}$ strain when $T = 1$, where $\tau = \mu_{Z^{(1)}} - \mu_{Z^{(0)}}$. Note that the effect of $T$ on $Z$ confirms to an unit-treatment additive model.

To generate potential outcomes of $Y$'s given $Z$'s we use the following steps:

7. Specify $\mu_{Y^{(0)}} = 25$, overall mean of $Y$ when $T = 0$ and $\mu_{Y^{(1)}} = 27$,overall mean of $Y$ when $T = 1$.
8. Specify $\sigma^2_{Y_\mu}$ & $\sigma^2_{Y_e}$ so that $\sigma^2_{Y_0} = \sigma^2_{Y_\mu} + \sigma^2_{Y_e}$ where $\sigma^2_{Y_\mu} = 40$, the variance of strain means of $Y$ when $T = 0$ and $\sigma^2_{Y_e} = 4$, the variances of $Y$ within the strain when $T = 0$.
9. Generate $b_1, b_2, ..., b_s \sim N(0, \sigma^2_{Y_\mu})$ and $c_1, c_2, ..., c_s \sim N(0, \sigma^2_{Y_\mu})$.
10. Generate $\mu_{Y_i^{(0)}} = \mu_{Y^{(0)}} + b_i$, the mean of $Y$ for the $i^{\text{th}}$ strain when $T = 0$.
    Generate $\mu_{Y_i^{(1)}} = \mu_{Y^{(1)}} + c_i$, the mean of $Y$ for the $i^{\text{th}}$ strain when $T = 1$.
11. Generate the error terms using multivariate normal distribution with a covariance matrix.
12. Then generate the potential outcomes of $Y$'s so that

$$Y^{(1)}(Z^{(1)}) = \mu_{Y_i^{(1)}} + \beta(Z_i^{(1)} - \mu_{Z_i^{(1)}}) + e_{11}$$
$$Y^{(1)}(Z^{(0)}) = \mu_{Y_i^{(1)}} + \beta(Z_i^{(0)} - \mu_{Z_i^{(1)}}) + e_{10}$$
$$Y^{(0)}(Z^{(1)}) = \mu_{Y_i^{(0)}} + \beta(Z_i^{(1)} - \mu_{Z_i^{(0)}}) + e_{01}$$
$$Y^{(0)}(Z^{(0)}) = \mu_{Y_i^{(0)}} + \beta(Z_i^{(0)} - \mu_{Z_i^{(0)}}) + e_{00}$$

where $\beta$ is fixed. Since there is no interaction between *T* and *Z*, the first and fourth lines have the same slope but different intercepts. The second line has the same slope and intercept as the first while the third line has the same features as the fourth.

## 6.4 Results and Discussion

In design 1 for each simulation data set, five cases were applied to the data. Cases 1 and 2 use data that are typically observed in a study of mediation effects, and cases 3, 4, and 5 use terminal measures data. The five cases are: (1) apply Baron & Kenny to the observed data without strain in the model (design1.obs), (2) apply Baron & Kenny to the observed data with strain in the model (design1.obs.s), (3) apply Baron & Kenny to the observed average data (that is, the terminal measures are estimated using observed averages of *Y* & *Z* ) (design1.avg), (4) apply the Baron & Kenny to the imputed data without strains (design1.imp), and (5) apply the Baron & Kenny to the imputed data with strains (design1.imp.s). Figure 6.5 shows the boxplots for all five cases for design 1.

The second box plot shows that the mediating effect can be estimated when the Baron & Kenny approach was applied for the observed data in a model that included strains. This is, in fact, the correct model. The first boxplot cannot estimate the mediating effect, eventhough there are no terminal measures data. This is because the strain variable is a highly significant blocking variable and, when it is not included in the model, the model cannot detect a mediating effect. Boxplots 3, 4, and 5 reflect estimated mediating effects using terminal measures data (i.e., cases 3, 4, and 5 above). As can be seen, the mediating effect cannot be estimated in any case because the mice are experimental units and there is no information in terminal measures data regarding the partial correlation between Y and Z, given strain. One can see from case 5 that the variability in estimates is much less using imputed data with strain in the model, again indicating strains usefulness as a blocking variable even though the mediating effect cannot be identified.

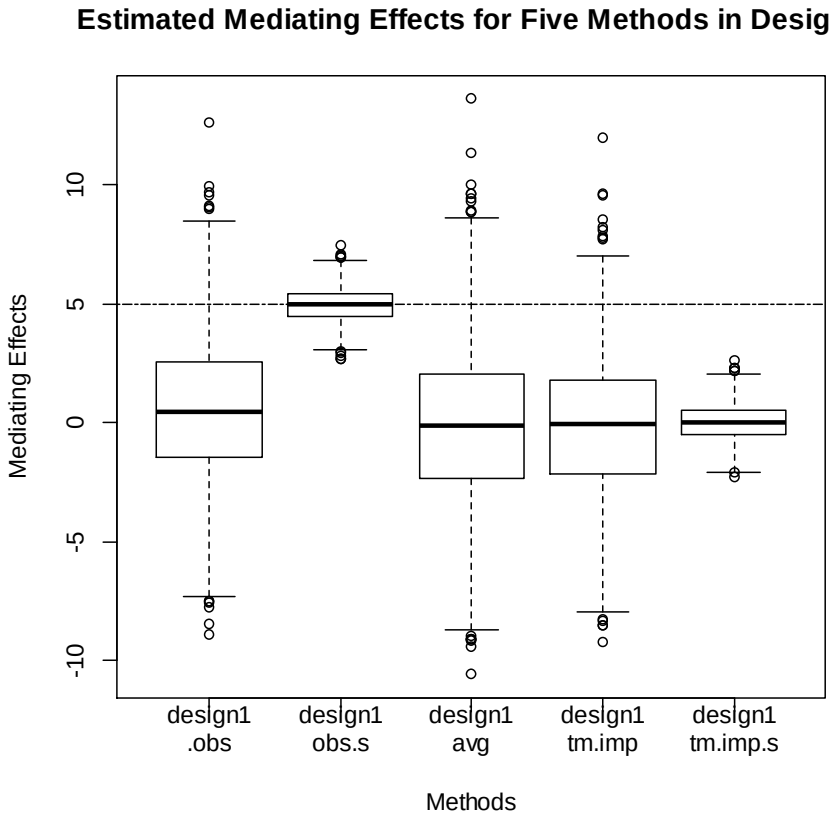



**Figure 6.5 Plots for Design 1 for Cases 1 – 5. The true mediating effect is equal to 5.**

**Design 2:** We use the similar steps as above with some modifications to generate the potential outcomes, observed data & terminal measures for the design 2. In design 2 for each simulated data set, five analyses were applied to the data. Again the first two cases used data that are typically observed in a mediation analysis, and cases 3 – 5 use terminal measures data. The five cases are: (1) apply Baron & Kenny to the observed data without strains (design2.obs), (2) apply Baron & Kenny to the observed data with strains (design2.obs.s), (3) apply Baron & Kenny to the strain level observed average data (that is, the terminal measures are estimated using observed averages of $Y$ & $Z$ ) (design2.tm.avg), (4) apply the Baron & Kenny to the imputed data without strains(design2.tm.imp), and (5) apply the Baron & Kenny to the imputed data with strains (design2.tm.imp.s). Figure 6.6 shows the boxplots for all five cases in design 2.

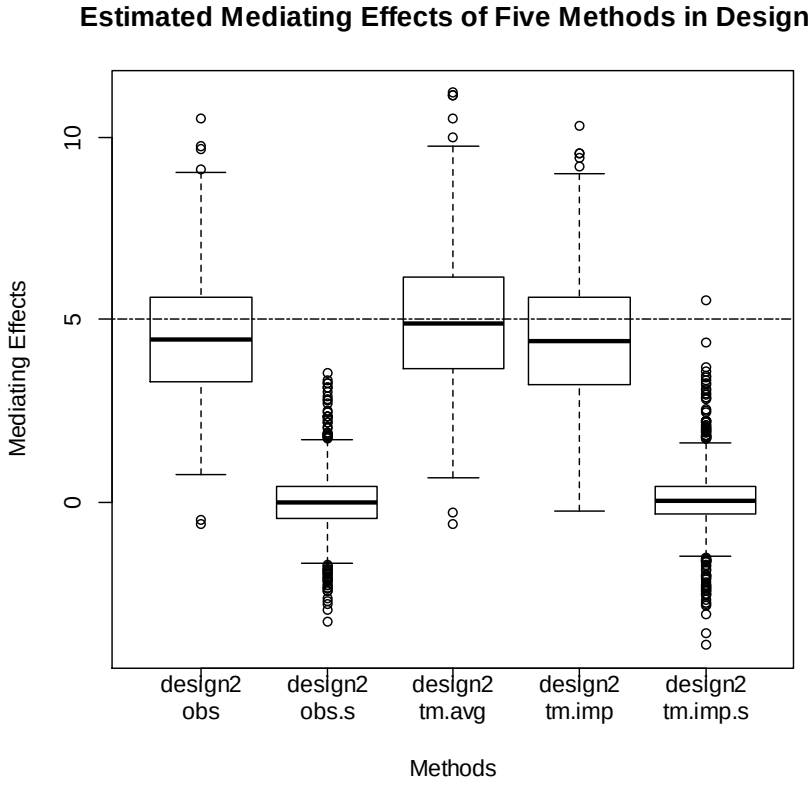



**Figure 6.6 Plots for Design 2 for cases 1 – 5. The true mediating effect is equal to 5.**

In design 2, strain is the experimental unit and the mice are pseudoreplicates. This allows for estimation of the mediating effect with or without terminal measures as seen in the boxplots for cases 1, 3, and 4. The assumption needed is that the treatment is acting at the level of the group, here strain. If strain is put into the model as was done in cases 2 and 5, then the mediating effect cannot be estimated. In other words, the mediating effect is being explained by strain, and the model cannot distinguish a mediating effect from a strain effect. Still, if one can assume that a treatment acts on the level of a group of units, then one can proceed with mediation analysis using terminal measures data. This assumption may be more plausible in certain settings, such as applying a treatment to a pot of multiple plants, or a litter of mice.

In brief, for design 1 the estimation of a mediating effect does not work for terminal measures data regardless of whether strain is included in the model since there was no information regarding partial correlation between $Y$ & $Z$, given strain. But in design 2, with strains as experimental units, the estimation procedure works for terminal measures data.

# Chapter 7 - Future Work

## 7.1 Summary of Dissertation

Some background of mediation analysis was introduced in chapter 1, and some hypothetical examples and the motivation of the study were discussed. In chapter 2, under the literature review, the past work done by many researchers was discussed and more importantly, the most popular approach of estimating average mediating effect, that is Baron & Kenny approach was introduced. The main assumption made was no interaction between treatment and the mediator. The potential outcomes framework for the causal effects was extended to the mediating effects. The total, direct and indirect effects were introduced and the relationship among them was established, based on the potential outcomes framework. The concept and the past work done on treatment heterogeneity were introduced. The mediation plot for the average mediating effect was described for a dichotomous treatment variable.

In chapter 3, the completed work done for the individual mediating effects for the preliminary exam presented. The mediation plot was extended to the individual mediating plot to describe the relationship among the individual total, direct and indirect effects. Using the individual mediation plots, sources of variability of the total effect across individuals was demonstrated. The variability of individual mediating effects, $\sigma^2_{IIE} = \mu^2_{D_Z}\sigma^2_B + (\sigma^2_B + \mu^2_B)\left[(\sigma_{Z^{(1)}} - \sigma_{Z^{(0)}})^2 + 2\sigma_{Z^{(1)}}\sigma_{Z^{(0)}}(1-\rho_{Z^{(1)}Z^{(0)}})\right]$ was formulated based on the estimable and nonestimable quantities. Upper and lower bounds of the variability of individual mediating effects were shown for three different cases. Then the mediating heterogeneity or subject-mediator (treatment) interaction was defined. In chapter 4, the completed work for heterogeneity of individual indirect effects was discussed and in chapter 5, testing for qualitative interactions was illustrated using a simulated data set. In chapter 6, a new concept which was called terminal measures data was introduced and a simulation study was conducted. Two designs were used to conduct simulation studies. Steps of generating potential outcomes, observed data and terminal measures data were described in detail and finally the results and discussion were presented.

## 7.2 Open Questions and Future Directions

1. Throughout this study no interaction between $T$ and $Z$ for an individual was assumed. This cannot be directly tested at the level of the individual. However, such an interaction can be investigated at the population level. Is an estimated $T \times Z$ interaction that is statistically significant evidence that this interaction must occur at the level of some individuals?

2. Are there improved imputation methods for terminal measures, perhaps in cases where $Y$ and $Z$ could be observed for at least a few units? Or if one had some prior knowledge regarding the partial correlation between $Y$ and $Z$ given a covariate, could this knowledge then be incorporated into the imputation procedure thereby allowing mediation analysis at the level of an individual?

3. Illustrate the size and power of Union-Intersection test, particularly, relaxing the assumptions that $\hat{\beta}_j$ and $\hat{\mu}_{D_{Z_j}}$ are independent.

4. Study the distributional properties of estimated $IE_j$ across subsets.

# References


Albert, J. M. (2008). Mediation analysis via potential outcomes models. Statistics in Medicine 27, 1282–1304.

Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology 1986; 51:1173–1182.

Berger, R.L.(1989). Uniformly more powerful tests for hypotheses concerning linear inequalities and normal means. Journal of the American Statistical Association, 84, pp. 192–199.

Booil, J.(2008) Causal Inference in Randomized Experiments With Mediational Processes. Psychological Methods, Vol. 13, No. 4, pp. 314–336.

Casella, G., and Berger, R. L. (2002). Statistical inference. Duxbury Press.

Ciminera,J.Heyse, J,Nguyen, H.Tukey, J (1993). Evaluation of Multicentre clinical trial data sing adoptions of the Mosteller-Tukey procedure. Statistics in Medicine12(11): 1047–61.

Dawid, AP (2000) Causal inference without counterfactuals. J Am Statist Assoc 95:407–424.

Frazier, P.A., Tix, A.P., Barron, K.E.(2004). Testing moderator and mediator effects in counseling psychology research. Journal of Counseling Psychology 51, 115–134.

Fritz M.S., McKinnon D. P. (2008) A graphical representation of the mediated effect. Behavior Research Methods 2008, 40 (1), 55-60.

Gadbury G. L. (2010). Subject – Treatment Interaction. In *Encyclopedia of Biopharmaceutical Statistics*, Third Edition, Revised and Expanded. Edited by Shein-Chung Chow. Informa Healthcare, London. p. 1316 – 1321.

Gadbury, G. L., Iyer, H. K., Allison, D.(2001) Evaluating subject-treatment interaction when comparing two treatments. J. Biopharm. Stat., 11, 313-333.

Gail, M. and Simon, R. Testing for qualitative interactions between treatment effects and patient subsets. Biometrics 1985, *41*, pp. 361-372.

Gallop R, Small D, Lin J, Elliott M, Joffe M and Ten Have T. Mediation analysis with principal stratification. Stat Med 2009; 28: 1108–1130.

Holland, P. W. (1986), Statistics and Causal Inference. Journal of the American Statistical Association, 81, 945–960.

Hurlbert S.H., (1984), Pseudoreplication and the Design of Ecological Field Experiments Ecological Monographs, Vol. 54, No. 2 , pp. 187-211.

Imai, Kosuke, Keele, Luke, & Tingley, Dustin (2010). A general approach to causal mediation analysis. Psychological Methods, 15(4), 309–334.

Lois A. Gelfand, Janell L.M., Thomas T.(2009).Mediation Analysis: A Restropective Snapshot of Practice and More Recent Directions, The Journal of General Psychology, 135:2,153-178.

MacKinnon D.P., Lockwood C.M., Hoffman J.M., West S.G., Sheets V. (2002). A comparison of methods to test mediation and other intervening variable effects. Psychological Methods, 7,83–104.

Masoro, E.J, Role of sirtuin proteins in life extension by caloric restriction. Mech Ageing Dev. 2004 Sep;125(9):591-4.

McKinnon, D. P. (2008). An introduction to mediation analysis. Mahwah, NJ: Erlbaum

Muller, D., Judd, C.M., Yzerbyt, V.Y.(2005). When moderation is mediated and mediation is moderated. Journal of Personality and Social Psychology, 89, 852-863.

Neyman, J. (1990). On the application of probability theory to agricultural experiments. Essay on principles Section 9 translated inStatistical Science, 5, 465–480.

Pearl, J. (2001). Direct and indirect effects. In Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, 411–420.

Piantadosi,S., Gai, M.H. (1993). A comparison of the power of two tests for qualitative interactions. Statistics in Medicine12(13): 1239–1248.

Peto R.(1982). Statistical aspects of cancer trials; in Halnan KE, (ed): The Treatmnent of Cancer. London, Chapman & Hall, pp 867-871.

Robert M, Beasley T.M, Gadbury G.L, Albert J.M, Kennedy R, Allison D.B. (2011). The validity and power of extreme sampling schemes for mediation analysis. Frontiers in Genetics: Behavioral and Psychiatric Genetics, 2:75.

Robert S. Poulson, Gary L. Gadbury, David B. Allison (2012). Treatment heterogeneity and individual qualitative interaction. American Statistician, 66, 16 - 24.

Roy, S.N. (1953). On a heuristic method of test construction and its use in multivariate analysis. Annals of Mathematical Statistics, 24, 220-238.

Rubin, D (1974) Estimating causal effects of treatments in randomized and nonrandomized

studies. J Educ Psychol 66:689.

Rubin, D (2005). Causal Inference Using Potential Outcomes: Design, Modeling, Decisions Journal of the American Statistical Association, 100, 469; ABI/INFORM Global, pg. 322.

Stephen, D., Jackie, A., David, B.,Susan, N., Carl, B.(2003) Caloric Restriction, Aging and Cancer Prevention: Mechanisms of Action and Applicability to Humans, Annu. Rev. Med. 2003. 54:131–52.

Steyer,R.(2005). Analyzing Individual and Average Causal Effects via Structural Equation Models. Methodology ; Vol. 1(1),39-54.

Susan J. Picquelle, Kathryn L. Mier(2011), A practical guide to statistical methods for comparing means from two-stage sampling.

Wright, S.(1934). The method of path coefficients. Annals of Mathematical Statistics, 5,161-215.

Wu, A. D., & Zumbo, B. D. (2008). Understanding and using mediators and moderators. Social Indicators Research, 87, 367–392.

Yusuf, S., J. Wittles, J. Probstfield, and H. A. Tyroler.(1991). ‘‘Analysis and Interpretation of Treatment Effects in Subgroups of Patients in Randomized Clinical Trials.’’ Journal of the American Medical Association 266 (1): 93–8.

Zelterman (1990). Tests for Qualitative Interactions. Statistics and Probability letters, 10: 59-63.