Analysis of Longitudinal Binary Data from Behavioral Medicine

by

Shira Dunsiger

B.S., McMaster University, 2001
A.M., Brown University, 2007

Thesis

Submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy
in the Division of Biology and Medicine at Brown University

May 2009

© Copyright

by

Shira Dunsiger

2009

This dissertation by Shira Dunsiger is accepted in its present form by
the Division of Biology and Medicine as satisfying the
dissertation requirement for the degree of
Doctor of Philosophy

Date..... .....
Joseph Hogan, Advisor

Recommended to the Graduate Council

Date..... .....
Constantine Gatsonis, reader

Date..... .....
Bess Marcus, reader

Approved by the Graduate Council

Date..... .....
Sheila Bond
Dean of Graduate School

The Vita of Shira Dunsiger

Shira Dunsiger was born in the City of Toronto, Canada on April 20^{th} 1979.

She received a Bachelor of Science degree in Mathematics and Statistics from McMaster University in 2001. She obtained a Master's of Arts degree in Biostatistics from Brown University in 2007.

Since the year 2001, she has been attending the Ph.D. program in Biostatistics at Brown University.

Acknowledgments

First, I wish to express my deepest gratitude to my advisor and mentor Professor Joseph Hogan, for his patience, guidance and support on this journey. Without his advice, this thesis would not have been possible. I would also like to thank my committee members, Professor Constantine Gatsonis and Professor Bess Marcus for whom I am greatly indebted for all their guidance along the way. I have learned so much from both of them and am truly grateful for the opportunity I have been given to do so.

Many thanks to friends and colleagues: Jason Roy, Mike Daniels, Beth Lewis, David Williams, Mei Hsui Chen, Karen Scheider, Joo Yeon Lee and Hong Li who have provided much advice and support during this experience. The discussions and comments we have had has added much to this work.

Special thanks to my parents, Dave and Judy and my in-laws Joyce and Steve. This was truly a group effort. I am forever grateful for your love and support and truly would not be here today had it not been for each of you. Finally, to my husband Shawn and son Ethan, I am truly blessed to have you both on my side.

This work was supported by the following grants: R01-HL-79457, R01-DA021729, R01-HL-064342, R29CA59660 and K07CA01757.

Dedication

To my boys, Shawn and Ethan, this one is for you.

Table of Contents

List of Tables

List of Figures

Abstract of "Analysis of Longitudinal Binary Data from Behavioral Medicine," by Shira Dunsiger, Ph.D., Brown University, May 2009

This thesis work is motivated by the specific challenges in analyzing data from clinical trials in behavioral medicine. Specifically, smoking cessation and physical activity trials. Typically, the data are binary and longitudinal with an appreciable amount of missing responses. Here, we propose methodology to address questions of behavior change, treatment efficacy and mediation. Although these methods are applied to behavioral data, they are applicable in other clinical trial settings as well.

The first part of this thesis examines a method for identifying distinct patterns of behavior change amongst participants in a smoking cessation trial. A latent class model is proposed which allows for a degenerate class of responses. A mean model is used to identify behavior change based on both an initial intensity and slope function. Further subdivision based on first order correlation is also proposed. This model finds classes of responses that are both clinically meaningful and informative.

The second part of the thesis examines the challenges in estimating treatment efficacy in smoking cessation trials, which are often subject to non-compliance. The G-computation algorithm is used to estimate the effect of receiving treatment. This method is first demonstrated in a hypothetical scenario and later applied to real data. The G-computation algorithm is a tool that is both computationally efficient and particularly well-suited to smoking cessation trials.

The third part of the thesis examines methods for establishing mediation in a clinical trial from behavioral medicine. A critique of standard regression methods is provided as well as a summary of key statistical papers. A multiple imputation approach is used to estimate the descriptive direct effect of treatment and consequently, the mediated effect

of treatment on outcome. A case study is provided in which this estimator is applied to data from a physical activity trial with a binary mediator. Extensions of this method are highlighted as are the advantages of using a causal method instead of a regression-based approach.

Chapter 1

Introduction

In this thesis, we focus on methods for analyzing binary longitudinal data from behavioral medicine. We propose models for identifying patterns of behavior change and for estimating the efficacy of receiving treatment. Finally, we propose a method for establishing mediation based on a multiple imputation estimator of the descriptive direct effect of treatment.

1.1 Longitudinal Data from Behavioral Medicine

The main motivation for this work comes from Randomized Clinical Trials (RCT) in behavioral medicine. Smoking cessation and physical activity trials are two examples of behavioral RCT's and have a number of features in common. Participants are randomized to treatment at baseline and are typically followed for up to 12 months. Outcome data are collected at regular intervals during the treatment phase, along with other psychosocial measures suspected to be related to treatment and/or outcome. As such, the data for each participant consists of multiple longitudinal records. In smoking cessation trials for example, the primary outcome is a binary indicator of weekly quit status.

Clinical trials such as those described above are often subject to a high degree of missing

data. Participants are being followed for long periods of time and are expected to change their behavior to something more positive (e.g. quit smoking or become physically active). These are known to be difficult behaviors to adopt and maintain over time. As such, participants who relapse or never change behavior may be more likely to drop-out of the trial before end of treatment. Missing data is also possible for other reasons unrelated to outcome (e.g. moving out of the area, no longer able to meet time demands of the trial, etc.). Therefore, in addition to the challenges in analyzing binary longitudinal data, researchers must deal with the issue of incomplete records on participants.

One advantage of behavioral interventions is that they *do* collect large amounts of data over time. The challenge is how to make use of what is available instead of relying on ad-hoc summaries of the data (e.g. total weeks quit, time to first relapse, etc.). In this thesis, we focus on three different questions that typically arise in the analysis of this class of data: what types of behavior change are we observing, what is the effect of receiving treatment, and is there a mediator of treatment assignment on outcome.

1.2 Outline of the Thesis

In this thesis, we focus on data analysis methods specific to behavioral medicine. Chapter 2 describes a latent class model used to find distinct subgroups of the population with respect to behavior change. This model makes use of the weekly data that was collected instead of a summary measure at end of treatment. Furthermore, the model is specified to allow for a subgroup of participants who never change behavior. We further illustrate how to use the serial correlation in the data to further inform class structure. Chapter 3 demonstrates the use of the G-computation algorithm for estimating the effect of receiving treatment. A substantial review of the potential outcomes framework is provided in addition to two

detailed examples: a hypothetical scenario that mimics a typical behavioral trial and a smoking cessation trial from the literature. Chapter 4 reviews the standard approaches to assessing mediation and key statistical papers that describe total, direct and indirect effects. This chapter critiques the use of standard regression approaches to mediation and suggests an estimator for the case when the mediator is binary. This method is applied to data from a physical activity trial. We summarize the conclusions from this thesis in Chapter 5 and discuss issues that require further research and consideration.

Chapter 2

Latent Class Analysis of Incomplete Longitudinal Binary Data from Smoking Cessation Trials

Abstract of "Latent Class Analysis of Incomplete Longitudinal Binary Data from Smoking Cessation Trials," by Shira Dunsiger, Ph.D., Brown University, May 2009

Longitudinal studies in behavioral medicine are often concerned with studying patterns of behavior change. When multivariate responses are available, latent class modeling can be used to identify distinct components of a multivariate distribution, thereby providing a method for distinguishing between patterns of behavior change. Applied to longitudinal binary data, a standard assumption in latent class models is that class membership captures the correlation between responses on the same individual.

This chapter is motivated by clinical trials in smoking cessation, where binary cessation status is measured weekly for 3 months. We develop a latent class regression model that allows serial correlation within class, thereby using both the mean and serial correlation to inform class membership. The model allows a separate class for responses consisting entirely of zeros. Covariate effects operate on the latent class scale. The model is used to analyze data from a three-arm longitudinal clinical trial comparing different doses of fluoxetine (prozac) for smoking cessation. The model identifies distinct classes of behavior change and the treatment effects are summarized in terms of latent class membership distribution.

2.1 Introduction

Trials in behavioral medicine are often concerned with modifying a behavior or pattern of behaviors; interventions are designed to affect the participants' ability to change a negative behavior, and maintain a new more positive behavior. In many cases, longitudinal or multivariate data are collected and used to quantify behavior change ([1, 2, 3, 4]). For example, the Commit to Quit study ([2]) randomized 281 female, sedentary smokers to a cognitive-behavioral smoking cessation program plus either a vigorous exercise program or a contact control. The cessation status of participants was recorded weekly over a 12-week study period. Analysis of weekly abstinence rates suggested that on average, those randomized to exercise achieved higher rates of cessation compared to those in the contact control ([2]). Unfortunately, this sort of summary does not reveal any population structure, such as subclasses of individuals who have different patterns of cessation. Another example from smoking cessation is a recent study of the therapeutic effects of fluoxetine ([4]) for smoking cessation (FSC trial). The FSC trial was a multi-center, double blind study in which participants were randomized to receive either one of two doses of fluoxetine hydrochloride or a placebo. Participants were followed for a 12-week period and weekly cessation status was recorded for each participant. Analysis of cessation at end of treatment suggested a moderate effect of fluoxetine versus placebo.

In general, intervention studies concerned with promoting smoking cessation assess cessation status at regular intervals. The follow-up time is divided into two phases: the early part is the lead-in period, during which participants are urged to reduce the number of cigarettes they are smoking, but are not necessarily asked to quit. A target quit date is set for the end of the lead-in period, and outcomes recorded during the second part of follow-up, the weeks following the quit date, are used to assess treatment effects. The illustrative

examples introduced previously both share this common structure. Both of these trials ran for 12 weeks and the target quit date was set for week 4. Weekly indicators of cessation status were recorded for each participant, therefore, information available on each participant is a longitudinal vector of binary indicators of cessation.

In this chapter, we describe and illustrate latent class models for repeated binary outcomes of the type commonly encountered in smoking cessation trials and similar studies of behavior change. The motivation for our work is that many existing approaches do not fully exploit the information available from data of this type, particularly as it relates to identifying subtypes of behavior change. Of the several common approaches used for analysis, perhaps the simplest is to use derived variables such as an aggregated total of weeks quit, or time to relapse. These simple approaches do not capture the longitudinal features of behavior change and thus do not always make optimal use of the available data. Another common approach is to model weekly cessation probability as a function of time using regression for longitudinal data (e.g. GEE or random effects models). Procedures for characterizing the marginal mean (like GEE) within each treatment group cannot generally be used to capture between-subject heterogeneity. Random effects models *can* be used to model the between-individual heterogeneity, but for binary data, commonly used assumptions about the random effects distribution (e.g. normality) may not be appropriate. This is illustrated further in our analysis of the FSC trial in Section 2.1.1.

Another issue that requires attention in longitudinal studies is missingness and dropout. A common but ad-hoc approach is to assume dropouts are treatment failures (smokers). However, this assumption neglects uncertainty associated with the "filling in" process. These shortcomings also apply to other single imputation procedures, such as last observation carried forward, worst observation carried forward, etc ([5]). However, even GEE

and random effects models make simplifying assumptions with respect to the missing data mechanism. Specifically, GEE applied to the complete cases will yield consistent estimates if missingness is completely independent of participant characteristics (MCAR) or depends only on covariates (MAR-C). Random effects models assume that the probability of missingness depends only on observed variables (MAR).

When interest is in characterizing population heterogeneity and the longitudinal outcomes are binary, it is important to allow for a subpopulation in which behavior change will not occur. Models based on the binomial distribution generally will not be appropriate because of the constraint that success probability is bounded away from zero.

As an alternative to standard approaches, we propose using latent class models for longitudinal binary data; we illustrate the model and its advantages using data from the FSC trial. Examples of latent class models can be found in the literature ([6, 7, 8, 9, 10, 11]). For example, Reboussin and colleagues used latent class models to assess drug involvement amongst school-attending youth ([10]). The resulting three class model identified subgroups representing youths in different stages of drug use. Bandeen-Roche and colleagues fit a latent class model to data on self-report disability from the Women's Health and Aging Study ([6]). Different levels of disability were identified using a three class model.

Our model assumes that longitudinal outcomes arise from a mixture over a finite number of latent classes, which can be interpreted to represent distinct patterns of behavior change. In addition, the model allows for a subgroup with cessation probability identically equal to zero.

The latent classes can be viewed as random effects with a multinomial but unspecified distribution. Although we do not need to make distributional assumptions about the latent class distribution, we assume that the number of latent classes is fixed and known. In

Section 2.1.1 we will give suggestions for selecting the number of classes. Inference is based on a likelihood, so missing responses are assumed to be missing at random (MAR), which means that the probability of cessation only depends on observed variables (previous covariates or outcomes). An important feature of the model as it relates to treatment comparisons is that each distinct sequence of binary outcomes is mapped to a set of class membership probabilities.

This chapter is organized as follows: in Section 2.1.1 we provide a detailed description of the motivating data. Section 2.2.1 describes the general formulation of the latent class model and subsequently, in Section 2.2.2 we present the specific model for the FSC trial. The results are presented in Section 2.3 and conclusions are made in Section 2.4.

2.1.1 Motivating Study

The FSC Trial enrolled 989 smokers and randomized them to one of three treatment groups (60 mg of fluoxetine, 30 mg of fluoxetine, or placebo). All participants received concurrent cognitive-behavioral therapy ([4]). About 60% of participants were women and the average age was 41.7 years. The treatment phase ran for the last 10 weeks of a 12-week trial and quit dates for each individual were set for week 4. Figure 2.1 shows a schematic of the study design and key points in the trial period.

For each participant, weekly indicators of smoking cessation comprise a vector of binary responses that has length at most 12. Figure 2.2 shows a sample of individual-level cessation outcomes that are representative of different types of sequences.

Overall, 130 participants (13%) dropped out on or before week 4 (target quit date) and 419 participants (42%) dropped out by week 12 (end of treatment). A Kaplan Meier curve shows that the pattern of dropout is similar across intervention groups ([4]).

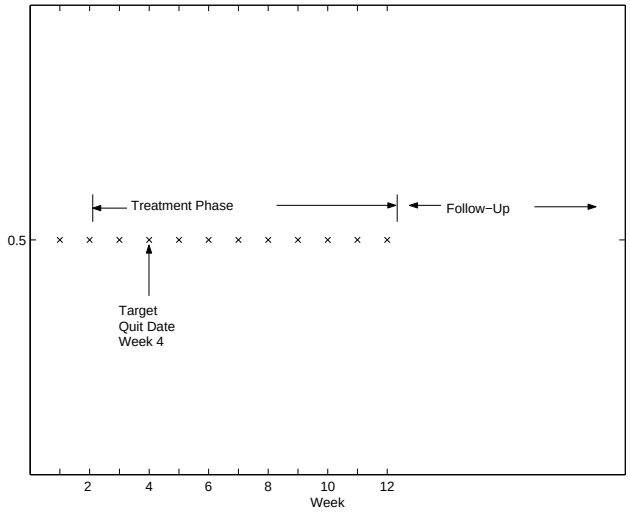

Figure 2.1: Schematic of the FSC Trial. Participants received treatment ("drug-on phase") during weeks 2-12 of the study and cessation outcomes were collected during this time. The target quit date was set for week 4 (two weeks into active treatment).

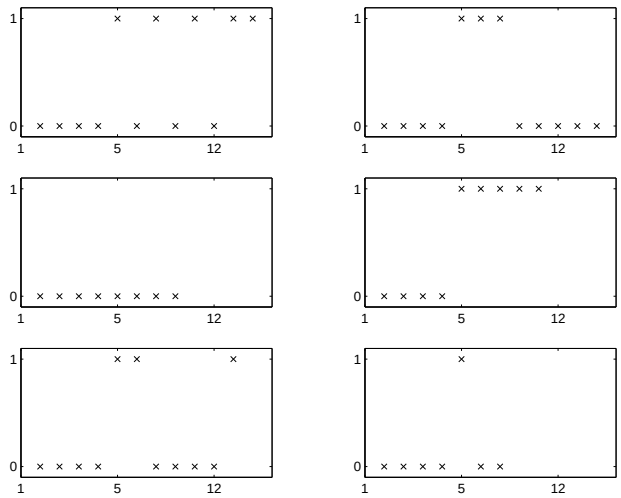

Figure 2.2: Sample of response vectors from 6 participants from the FSC Trial. The horizontal axis represents follow-up time in weeks and the vertical axis represents smoking cessation status (1 =yes, 0 =no). Week 5 is marked to indicate the first week beyond the target quit date.

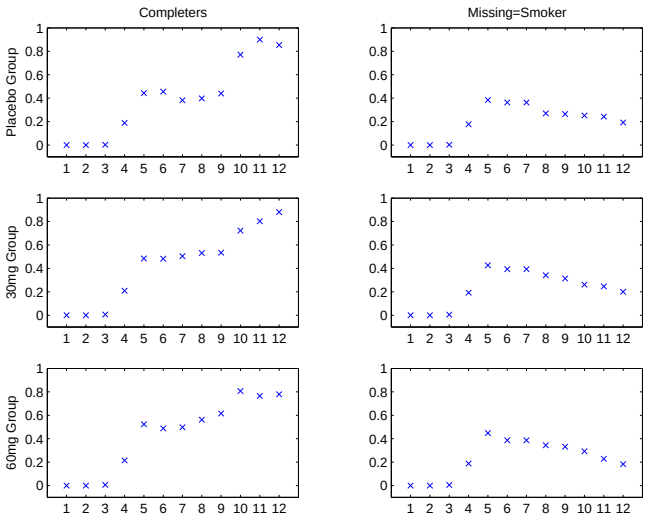

Figure 2.3: Proportion of participants quit at each week during post quit-date period. The right column assumes missing cases were treatment failures (smoking) and the left column considers only the complete cases.

Two simple ways to handle missing data are to treat all dropouts as smokers, or to evaluate quit rates based only on those still in follow-up. Plots of these are shown in Figure 2.3. We can see that these assumptions yield different estimates of quit rates over time. Under the assumption that missingness equals treatment failure, quit rates decrease over time in all 3 treatment groups. On the contrary, among complete cases, quit rates improve over time.

As indicated in Section 2.1, one approach to inferring treatment affects is to use a random effects model. One way to assess the assumption about the random effects distribution, is to graph the empirical cessation probabilities (as we have done in Figure 2.4). The U-shaped distribution suggests that a normality assumption on the random effects (from an intercept-only model) would not be reasonable in this case.

In order to understand whether population heterogeneity may be a relevant concern, we can look at a graph of the proportion of observed weeks quit across the study population.

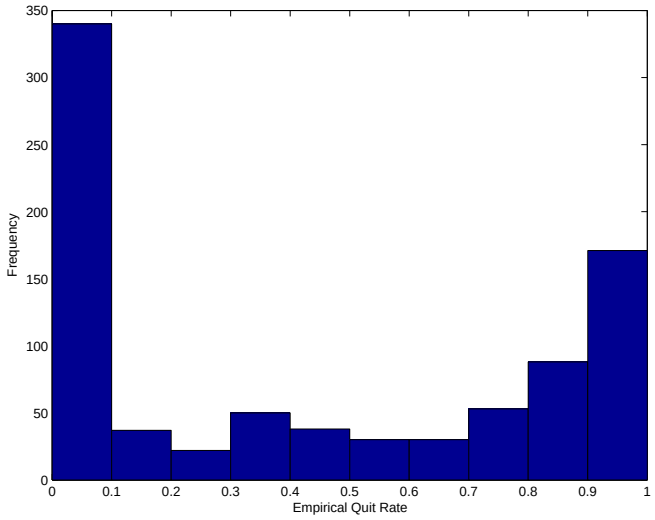

Figure 2.4: Distribution of observed quit rates during post quit-date period. The empirical quit rates were computed for each of the 989 participants in the FSC trial over weeks 5-12.

Here we define

$$
\begin{aligned}
Y_{it} &= \text{Indicator that participant } i \text{ is quit at week } t \\
R_{it} &= \text{Indicator that quit data is observed for participant } i \text{ at week } t \\
\hat{p_i} &= \frac{\sum_{t=5}^{12} Y_{it} R_{it}}{\sum_{t=5}^{12} R_{it}}
\end{aligned}
$$

For participant i, $\hat{p_i}$ is the proportion of observed weeks quit during post quit-date period (weeks 5-12). Figure 2.4 shows a histogram of these observed quit rates.

Figure 2.4 also suggests the existence of distinct classes of behavior (e.g. whether a behavior change took place). There is a noticeable mass at zero suggesting the existence of a class of participants with quit rates identically equal to zero.

The use of Latent Class Models for data reduction will allow us to subdivide the data set into more meaningful subgroups while still making use of the longitudinal structure of the data ([6, 7, 8, 10]).

2.2 Latent Class Model for Serial Binary Data

In this section, we review the notation for the general latent class model and the specifics pertaining to the FSC trial.

2.2.1 Notation for General LCM

Let $Y_i = (Y_{i1}, Y_{i2}, \cdots, Y_{iT})^T$ be a $T \times 1$ vector of binary responses for individual i, where $Y_{it} = 1$ indicates a positive response. In our application, $Y_{it} = 1$ denotes cessation at time t. Let η_i, a nominal, categorical variable, denote class membership for individual i, with $\eta_i \in \{1, 2, \cdots, J\}$. We assume that J is fixed in advance. In absence of covariates, the model factors the joint distribution $f(Y, \eta)$ as $f(Y \mid \eta)f(\eta)$, where for each j and for $t = 1, \cdots, T$,

$$\begin{aligned} Y_{it} \mid \eta_i = j &\sim \text{Bernoulli}(\pi_{it}^{(j)}) \\ \eta_i &\sim \text{Multinomial}(\psi_i) \end{aligned}$$

The joint distribution of $Y_1, Y_2, \cdots, Y_T$ is therefore a mixture over the latent classes. In particular, by defining $\psi_i^{(j)} = P(\eta_i = j)$, we have

$$P(Y_{i1} = y_1, Y_{i2} = y_2, \cdots, Y_{iT} = y_T) = \sum_{j=1}^{J} \psi_i^{(j)} \prod_{t=1}^{T} (\pi_{it}^{(j)})^{y_t} (1 - \pi_{it}^{(j)})^{1-y_t}$$

.

We will further assume $\pi_{it}^{(j)} = g(t, \beta^{(j)})$, a known function of t, indexed by unknown parameter $\beta^{(j)}$.

Time invariant covariates enter on the latent scale. Suppose for individual i we have a

$p \times 1$ vector of covariates X_i^T, and let $\psi_i^{(j)} = \psi^{(j)}(X_i)$. Then,

$$P(Y_{i1} = y_1, Y_{i2} = y_2, \cdots, Y_{iT} = y_T \mid X_i) = \sum_{j=1}^{J} \psi^{(j)}(X_i) \prod_{t=1}^{T} (\pi_{it}^{(j)})^{y_t} (1 - \pi_{it}^{(j)})^{1-y_t}$$

Thus $Y \perp X|\eta$.

2.2.2 Specification for the FSC Trial

First we describe the model assuming fully observed responses. Next we show how to fit the model from partially observed data. At Level I we assume that class $\eta = 1$ consists of individuals whose probability of cessation during the study is zero. Hence,

$$P(Y_{it} = 0|\eta_i = 1) = 1$$

for all t. For $j = 2, 3, \cdots, J$, we assume $(Y_{ij}|\eta_i = j) \sim \text{Bernoulli}(\pi_{it}^{(j)})$, where

$$\begin{aligned} \text{logit}(\pi_{it}^{(j)}) &= g(t, \beta^{(j)}) \\ &= \beta_0 I(t \leq q-2) + \beta_1 I(t = q-1) + \{\beta_2^{(j)} + (q+t)\beta_3^{(j)}\} I(t \geq q) \end{aligned}$$

We combined the data from weeks $q-2$ and $q-3$ into one indicator of cessation at at least one of these time points. This model implies that within class, the log odds of cessation over time follows a left continuous function with jump of size $\beta_2^{(j)}$ at q, and a linear trend thereafter with slope $\beta_3^{(j)}$ (see Figure 2.5).

Prior to the quit date, there is little variability in the responses. As such, β_0 and β_1 are assumed constant across class. We chose to separately include week $q-1$ since variability

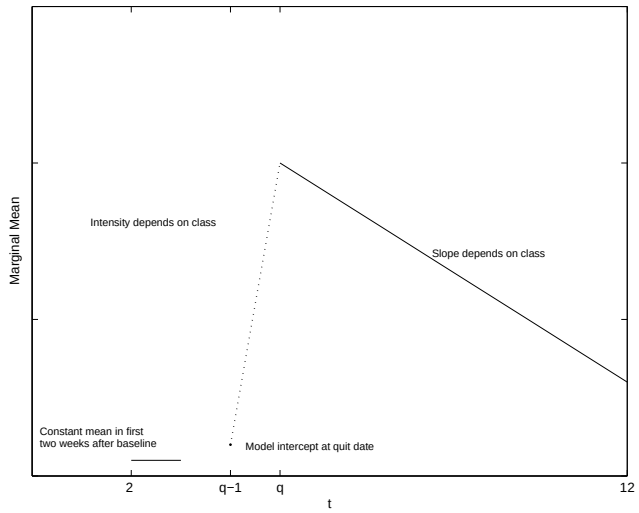

Figure 2.5: Model of marginal mean within each class. The horizontal axis denotes time (week in the study period) and q indicates first week post quit-date (week 5). The vertical axis is the marginal mean (on the logit scale).

in this particular week may be important for dropout. Latent classes are assumed to differ in both the magnitude of the jump at week q and in slope of the linear function after the quit date. Parameters characterizing both of these quantities are thus indexed by class.

At level II, we allow covariate effects to be parameterized on the latent scale using relative risk multinomial regression, e.g.,

$$\eta_i | X_i \sim \text{Multinomial}(\psi_i)$$

where

$$\psi_i = (\psi_i^{(1)}, \cdots, \psi_i^{(J)})^T, \delta = \{\delta^{(j)} : j = 1, 2, \cdots, J-1\}$$

and

$$\log \frac{\psi_i^{(j)}}{\psi_i^{(J)}} = X_i \delta^{(j)}$$

.

In this model, $\delta^{(j)}$ is the effect of X_i on the log relative risk for being in class j versus

the reference class J. Since η_i is assumed to follow a multinomial distribution, it is also necessary to impose the constraint that $\sum_{j=1}^{J} \psi^{(j)} = 1$ for all i.

Finally, to fully specify the Bayesian model, we assume that all regression parameters have independent priors following $N(0, 10^3)$ ([12]). Changes in the prior variance (e.g. 10^3 versus 10^4) did not affect inference. See Appendix A for details of the sampler.

2.2.3 Full Data Likelihood

Recall that the log odds of cessation can be written as a function of t, namely, $g(t, \beta^{(j)})$. The log relative risk of being in class j versus the reference class J is $\exp(X_i\delta^{(j)})$. The likelihood contribution for individual i is

$$L_i(\beta^{(j)}, \delta^{(j)}, \mid Y_{i1}, Y_{i2}, \cdots, Y_{iT}, X_i) \propto f_0(Y_{i1}, \cdots, Y_{iT} \mid \eta_i, \beta^{(j)}) f_1(\eta_i | X_i, \delta^{(j)})$$

where f_0 is the joint distribution of the responses conditional on latent class and f_1 is the distribution function of $(\eta_i | X_i)$. If we let $L_i^{(j)}$ be the likelihood contribution of individual i conditional on being in class j, then,

$$
\begin{aligned}
L_i^{(1)} &= \left\{ \frac{\exp(\delta_0^{(1)} + \delta_1^{(1)} X_i)}{1 + \sum_{j=1}^{J-1} \exp(\delta_0^{(j)} + \delta_1^{(j)} X_i)} \right\} \\
L_i^{(j)} &= \left\{ \prod_{t=1}^{T} (\pi_{it}^{(j)})^{Y_{it}} (1 - \pi_{it}^{(j)})^{(1-Y_{it})} \right\} \left\{ \frac{\exp(\delta_0^{(j)} + \delta_1^{(j)} X_i)}{1 + \sum_{j=1}^{J-1} \exp(\delta_0^{(j)} + \delta_1^{(j)} X_i)} \right\} \\
&= \left\{ \prod_{t=q-2}^{q-2} \frac{\exp(Y_{it}\beta_0)}{1 + \exp(\beta_0)} \right\} \left\{ \prod_{t=q-1}^{q-1} \frac{\exp(Y_{it}\beta_1)}{1 + \exp(\beta_1)} \right\} \left\{ \prod_{t=q}^{T} \frac{\exp(Y_{it}(\beta_2^{(j)} + (q+t)\beta_3^{(j)}))}{1 + \exp((\beta_2^{(j)} + (q+t)\beta_3^{(j)}))} \right\} \times \\
&\times \left\{ \frac{\exp(\delta_0^{(j)} + \delta_1^{(j)} X_i)}{1 + \sum_{j=1}^{J-1} \exp(\delta_0^{(j)} + \delta_1^{(j)} X_i)} \right\} \\
&= \left\{ L_i^{(j)}(\beta_0) \right\} \left\{ L_i^{(j)}(\beta_1) \right\} \left\{ L_i^{(j)}(\beta_2^{(j)}, \beta_3^{(j)}) \right\} \left\{ L_i^{(j)}(\delta_0^{(j)}, \delta_1^{(j)}) \right\}
\end{aligned}
$$

Thus, the likelihood for the full data model is

$$
\begin{aligned}
L(\beta, \delta) &= \prod_{i=1}^{N} L_i(\beta_0, \beta_1, \beta_2^{(j)}, \beta_3^{(j)}, \delta_0^{(j)}, \delta_1^{(j)} \mid Y_{i1}, Y_{i2}, \cdots, Y_{iT}, X_i) \\
&= \prod_{i=1}^{N} \left\{ \prod_{j=1}^{J} [L_i^{(j)}]^{I(\eta_i = j)} \right\} \\
&= \prod_{i=1}^{N} \left\{ \prod_{j=2}^{J} [L_i^{(j)}(\beta_0) L_i^{(j)}(\beta_1) L_i^{(j)}(\beta_2^{(j)}, \beta_3^{(j)})]^{I(\eta_i = j)} \prod_{j=1}^{J} [L_i^{(j)}(\delta_0^{(j)}, \delta_1^{(j)})]^{I(\eta_i = j)} \right\}
\end{aligned}
$$

To handle missing responses, we assume missingness at random and implement a data augmentation procedure based on the full data model. Specifically, if we let $Y_{i,t}^{mis}$ be a missing response for individual i, then

$$
(Y_{i,t}^{mis} | \eta_i, \beta_0, \beta_1, \beta_2^{(j)}, \beta_3^{(j)}) = \begin{cases} 0 & \text{if } \eta_i = 1; \\ \sim \text{Bernoulli}(\pi_{it}^{(j)}) & \text{if } \eta_i = j, j \geq 2. \end{cases}
$$

Specifics pertaining to the sampler can be found in Appendix A.

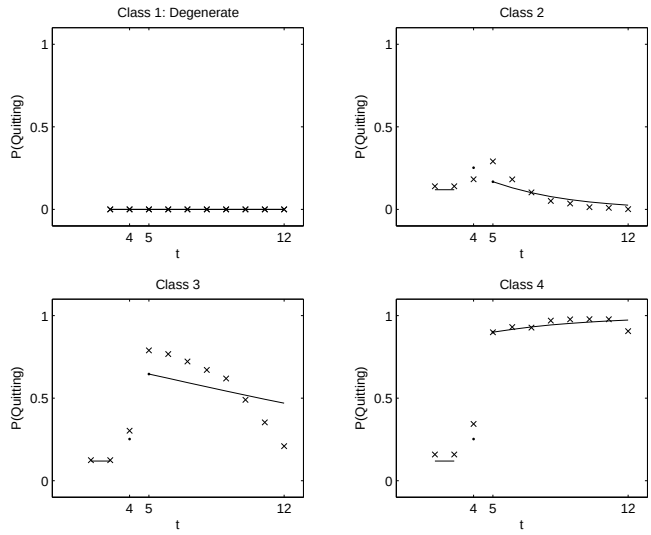

Figure 2.6: Estimated marginal mean over time within each latent class. The horizontal axis depicts time in the study period and the vertical axis represents probability of cessation. The x's denote the sample average of the mean response vectors (across iterations of the sampler).

Parameter	$\eta = 1$	$\eta = 2$	$\eta = 3$	$\eta = 4$
β_0	-	$-2.00(-2.09, -1.91)$	$-2.00(-2.09, -1.91)$	$-2.00(-2.09, -1.91)$
β_1	-	$-1.14(-1.15, -1.12)$	$-1.14(-1.15, -1.12)$	$-1.14(-1.15, -1.12)$
$\beta_2^{(j)}$	-	$1.07(1.00, 1.13)$	$1.76(1.73, 1.79)$	$0.21(0.16, 0.26)$
$\beta_3^{(j)}$	-	$-0.32(-0.43, -0.21)$	$-0.13(-0.25, -0.01)$	$0.20(0.05, 0.36)$

Table 2.1: Posterior Means and Intervals of Model Parameters

2.3 Results

A four class model was fit to data from the FSC Trial. Recall that $N = 989$ participants were followed for $T = 12$ weeks with quit date set at week 4. Define the treatment covariate as X_i =I(participant i received at least 30mg of fluoxetine) ([4]). Results are displayed in Figure 2.6 as plots of the marginal mean over time.

Figure 2.6 shows the four latent classes supported by the data. Lines represent the posterior mean of the probability of quitting as a function of time. Clinically speaking,

these are four very different subgroups of the population. Prior to the quit date, we see low cessation rates (approximately 0.12) which take a slight jump at the quit week (cessation around 0.24) suggesting that participants are perhaps "gearing up" to change behavior. Low cessation rates in the early study weeks can be attributed to the study design, as the participants were only instructed to reduce their cigarette use but not stop it all together. By definition, the mean takes a jump the week following the quit date. This can be likened to an intensity. The smallest jump was in class 2 (this is a subgroup with overall very low cessation rates). By contrast, the first week post-quit date in Class 4 was followed by quit rates greater than 80%. These are the highly successful quitters as not only is the *jump* positive but the slope of the mean after this time is positive as well. Class 3 represents an interesting subgroup of participants. The mean in this group jumps more than 40% the week after the quit date and then steadily decreases to around 50%. Posterior means of the model parameters are displayed in Table 2.1.

Figure 2.6 also addresses the model fit. The x's denote the mean response vectors averaged across iterations of the sampler. The degenerate class has a perfect model fit since the definition stipulated the response vectors consist entirely of zeros. Class 4 also shows a fairly tight fit particularly after the quit date. The most variation exists in the middle two classes perhaps suggesting that the mean model may not be entirely sufficient to describe the data. We will return to this issue of correlation in Section 2.4.

To better understand how the sampler works, we can look at a sample of observed responses and the corresponding latent class probabilities for those individuals (Figure 2.7). As indicated in Section 2.1.1, there is an appreciable amount of missing data. Figure 2.8 shows how the observed data were augmented by looking again at a sample of participants. The x's denote observed responses and the +'s denote distribution sampled responses with

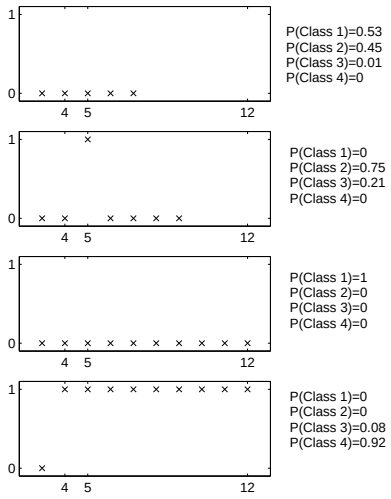

Figure 2.7: Observed response vectors and the estimated class probabilities from the model, for a sample of participants from the FSC trial.

Class	Probability of Being In Class j
Class 1	0.23 (0.20, 0.25)
Class 2	0.35 (0.32, 0.38)
Class 3	0.27 (0.24, 0.30)
Class 4	0.15 (0.13, 0.17)

Table 2.2: Distribution of Classes

standard error.

As suggested in Figure 2.7, for each of the N participants, there is a vector of probabilities corresponding to the probability of being assigned to each of the four latent classes. We can summarize these probabilities across the cohort of study participants and aggregate across treatment groups (Table 2.2).

From Table 2.2 we see the highest percentage of participants are assigned to Class 2 (low non-zero cessation rates). Table 2.3 summarizes the posterior means of the multinomial

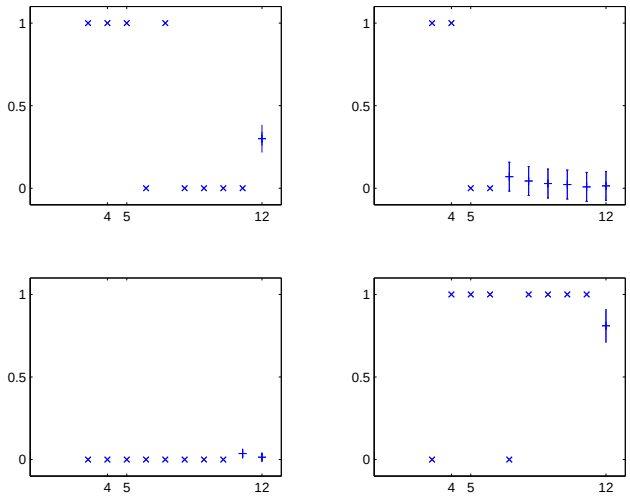

Figure 2.8: Sample of participants from the FSC Trial. The x's denote observed data. The means and error bars associated with imputed missing values are also represented. There is uncertainty in the imputed response value since the sampler augments observed responses at each iteration.

ψ_j	$X_i = 1$	$X_i = 0$	Risk Difference
ψ_1	0.23	0.23	-.007 (-0.06, 0.05)
ψ_2	0.36	0.35	0.01 (-0.05, 0.07)
ψ_3	0.28	0.25	0.03 (-0.03, 0.09)
ψ_4	0.14	0.17	-0.04 (-0.08, 0.01)

Table 2.3: Posterior Means of Class Probabilities Conditional on Treatment

parameters across levels of the binary treatment indicator.

Table 2.3 suggests that given treatment with at least a 30mg dose of fluoxetine, a participant is most likely to be in class 2 (low non-zero cessation rates). Intervals covering the risk difference between treatment groups does not suggest a significant treatment effect. This could be in part due to the significant drop-out among participants ([4]) and that fluoxetine appears to have early and modest effects on cessation which may be missed in these longitudinal patterns. Dose effect models can also be accommodated into this LCM

but again, in the case of this example, do not show differences across treatments.

2.4 Discussion

Our application of latent class modeling for smoking cessation trials has allowed us to find interpretable subgroups of this population of smokers and understand the covariate effects through the latent class distribution. Using this methodology we have avoided the use of ad-hoc summaries of the longitudinal data. Using latent class models we can understand the different types of quitters that make up this population of study participants without making distributional assumptions (on the random effects for example). We have easily accommodated for a class of never-quitters (which is a key subset of the population of smokers attempting to change their behavior) and handled the missing data in a systematic way (without assuming that they have necessarily dropped out to return to smoking).

The use of this particular model requires knowledge of the number of classes a priori. Preliminary analysis of the distribution of the observed quit rates showed a U-shaped curve that suggested the presence of distinct classes. We chose the preliminary number of classes based on this curve. We did vary the number of classes to see whether the data could support more or less classes (than the 4 presented here). Models were compared using a modified Bayesian Information Criterion (BIC). The model presented in Section 2.3 corresponds to the model with the lowest BIC value. Future work related to model selection is needed in order to use a more systematic approach to choosing the preliminary number of classes and for comparing models that differ only in the number of classes.

One of the main goals of this analysis was to identify the patterns of cessation among participants in the study. Understanding these patterns could be useful for designing future interventions. Referring to Figure 2.6, we can see that one distinguishing difference between

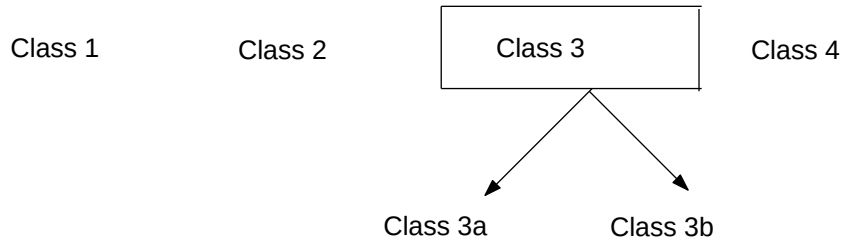

Figure 2.9: Model schematic. The first row describes the latent class model we have discussed up to the point. The second row suggests how further subdivision of the latent classes could be carried out by modeling first order correlation amongst one class of participants.

the classes is the magnitude of the jump one week post-quit date. Those taking the largest jump seem more likely to maintain cessation.

Resulting estimates of the marginal mean from a four class model suggest a class of participants with expected cessation around 50%. We could hypothesize that this class contains two different subgroups of the population: those who never make a permanent behavior change and instead alternate between quitting and smoking on a weekly basis and those who make a change for a few weeks at a time but ultimately cannot maintain that behavior. A model that assumes independence between responses within class is clearly not sufficient to capture the difference between these individuals. Since these cases represent two different challenges to clinicians, it is meaningful to further subdivide this group of people based on first order correlation structure. Figure 2.9 describes a future analysis plan.

The first level of Figure 2.9 represents the model outlined in the previous sections. Fitting this model results in estimates for the parameters characterizing the marginal mean and the distribution of the latent classes. The sampler terminates with a matrix of posterior probabilities of being in each of the J classes (as seen in Figure 2.7). Using these multinomial probabilities we could draw the class membership of each of the individuals. We restrict our model to those who are sub-sampled into class 3. Fitting a first order Markov

model to this subset of participants will force the marginal mean to be the same but allow us to find two more latent classes based on the correlation. Specifically, let

$$\begin{aligned} \mu &= P(Y_{it} = 1) \\ \phi_{it}(j) &= P(Y_{it} = 1 | Y_{i,t-1}, \eta_i^* = j) \end{aligned}$$

At Level I, we assume that conditional on latent class, current smoking cessation status depends only on status at the previous week. That is, for $j = 1, 2$,

$$(Y_{it} | Y_{i,t-1}, \eta_i^* = j) \sim \text{Bernoulli}(\phi_{it}(j))$$

where

$$\text{logit}(\phi_{it}(j)) = \tau_0^{(j)} + \tau_1^{(j)} Y_{i,t-1}$$

At Level II, we assume that latent class follows a multinomial distribution with parameters γ. That is,

$$\eta_i^* | \gamma \sim \text{Multinomial}(\gamma)$$

such that

$$\gamma_1 + \gamma_2 = 1, 0 < \gamma_j < 1$$

Finally, assume all regression parameters have independent priors following $N(0, 10^3)$, e.g., non-informative priors, and the multinomial parameters have a Dirichlet prior with parameters $(1, 1)$. We make the same assumptions for the missing data as before and modeling is again carried out using Gibbs sampling with multiple Metropolis Hastings steps. Prelimi-

Y_{t-1}	$\eta^* = 3a$	$\eta^* = 3b$
0	$P(Y_t = 1\|Y_{t-1}) = 0.41$	$P(Y_t = 1\|Y_{t-1}) = 0.65$
1	$P(Y_t = 1\|Y_{t-1}) = 0.49$	$P(Y_t = 1\|Y_{t-1}) = 0.33$
$P(\eta^* = j)$	0.81	0.19

Table 2.4: Posterior Means of Transition Probabilities from Correlation Model

nary results from fitting this model can be seen in Table 2.4.

From Table 2.4 we see that class *3a* consists of those participants with a higher probability of sustaining behavior. On the contrary, class *3b* consists of those with a higher probability of switching behavior. In other words, those in the class *3a* are about 63% as likely to go from smoking to quitting as those in the class *3b* and approximately 1.5 times more likely to maintain quitting once it has been achieved. As expected, these are two very different subgroups of smokers. Future work in this area could allow for one succinct model that captures both the marginal mean and first order correlation structure of the data.

Although the analysis shown in Section 2.3 has allowed for treatment comparisons, we have not assessed the individual level predictors of cessation pattern. Smoking cessation literature has documented the effects of mood and weight gain on achieving and maintaining quitting. Given the appropriate data set, these covariates could be incorporated through the latent class distribution.

Data in smoking cessation trials are challenging to work with. However, they lend themselves to studies of response profiles due to their longitudinal design and correlated outcomes. As nicotine dependence is a noted cause of more than one health problem ([13]), it is imperative to understand the ways in which people can be aided in smoking cessation. Understanding the patterns of quit behavior amongst participants who are motivated to

make a change may help us better understand the ways in which quitting happens.

Chapter 3

G-Computation Algorithm for Smoking Cessation Data

Abstract of "G-Computational Algorithm for Smoking Cessation Data," by Shira Dunsiger, Ph.D., Brown University, May 2009

Smoking cessation trials often collect longitudinal records on both cessation status and compliance with assigned treatment. Since many of these studies are subject to non-compliance, it may be of interest to estimate both the effectiveness and the efficacy of the intervention. The latter poses challenges to the researcher since compliance is a post-randomization variable.

This chapter is a translation research piece. The objectives are to: (1) describe the G-Computation Algorithm (GCA) for estimating the effect of receiving treatment and illustrate it with a detailed hypothetical example and (2) apply it to weekly data from a smoking cessation trial (Commit to Quit). In Commit to Quit, participants were randomized to a smoking cessation program plus either exercise or contact control (wellness). Results suggest that if participants had been fully compliant with assigned treatment, the

difference in weekly quit rates between the exercise and wellness arms would have ranged from 0.09-0.19. This chapter includes a comparison between these GCA estimates and those from the intent to treat analyses. Finally, the specific advantages of the GCA for use in smoking cessation trials are highlighted.

3.1 Introduction

The primary goal of most intervention trials is to characterize the effect of an intervention on an outcome. Many behavioral interventions are subject to non-compliance due to participant burden associated with the treatment. For this reason, the intention to treat effect alone may not be sufficient and estimating the effect of *receiving* treatment may be of particular interest.

Consider smoking cessation trials, where the goal is to compare the cessation rates between groups randomized to two or more interventions. Studies like these can have low rates of compliance. With non-compliance, a fundamental issue is how to define the treatment effect. Options include the Intent to Treat effect (ITT), the Local Average Treatment Effect (LATE) and the Average Treatment Effect (ATE) ([14, 15, 16]). To estimate the effectiveness of the intervention, we can use an ITT approach, in which cessation rates are compared across treatment groups, regardless of whether or not participants were compliant. In this case, we must make an assumption about the missing data since all participants randomized will be included in the analysis. Estimating the causal effect of *receiving* treatment is more challenging because compliance is a post-randomization variable. Available methods for estimation include propensity scores, instrumental variables, inverse weighting by the propensity score and the G-computation algorithm ([17, 15, 18, 19, 20]).

For this chapter, we use the Commit to Quit study ([21]) to illustrate the G-computation algorithm for estimating the ATE. Commit to Quit (CTQ) randomized participants to a smoking cessation program plus one of two active treatments: a vigorous exercise program or a contact control ("wellness" group). Requirements stipulated that participants attend three treatment sessions per week over the 12-week study period, in addition to their weekly smoking cessation session. In this study, compliance measures were available and observable,

unlike with pharmaceutical trials. Although quit status was recorded on a weekly basis, there were missing observations in some cases (when participants were lost to follow-up for example). In addition, covariate information was collected on participants at baseline, including both demographic and psychosocial variables.

The chapter will be organized as follows: in section 3.2.1, we will describe the potential outcomes framework ([16]) in clinical trials and how it can be used to define the causal effect of receiving treatment (ATE). Next, in Section 3.2.3 we will demonstrate how a causal model leads to observed data using a hypothetical example. In Section 3.3.1 we will review the assumptions needed to use the GCA and show how the ATE can be estimated using this method. We will return to our hypothetical example in order to demonstrate the GCA. In Section 3.3.4 we will describe the GCA implementation for longitudinal data and present the results of applying this method to data from the Commit to Quit trial (Section 3.4). In addition, we will discuss advantages of GCA for use with moderate-sized sequences of longitudinal binary data (and smoking cessation trials in particular) in Section 3.5.

3.2 Defining Treatment Effects with Potential Outcomes

3.2.1 Potential Outcomes

Potential outcomes are random variables that represent responses under potential exposures (only one exposure is actually observed). These hypothetical constructs provide a formal structure for defining the ATE ([16]). In CTQ, consider the simplified case when interest is in cessation status at end of treatment. Let Z denote treatment assignment (1 =wellness and 2 =exercise) and let D denote compliance pattern. For simplicity, we assume that the possible values of D are:

$$d = \begin{cases} 1 & \text{if complied with wellness and was expected to} \\ 2 & \text{if complied with exercise and was expected to} \\ 3 & \text{if did not comply with wellness but was expected to} \\ 4 & \text{if did not comply with exercise but was expected to} \end{cases}$$

Note that d values partition cases into mutually exclusive sets. Each participant is assumed to have a set of potential outcomes enumerated by z and d, and denoted by $Y(z, d)$. These represent the cessation status at end of treatment if the participant was randomized to treatment z and had compliance pattern d. For example, $Y(1, 1)$ represents the quit status of a participant who was randomized to wellness and fully complied whereas $Y(1, 2)$ is the quit status of a participant randomized to wellness who would have fully complied with exercise if expected to do so.

Although each potential outcome is well-defined, not all of them are observable. In CTQ, those randomized to one arm do not have access to the other. Hence, for those randomized to wellness, it is only possible to observe one of $Y(1, 1)$ and $Y(1, 3)$. Similarly, for those randomized to exercise, it is only possible to observe one of $Y(2, 2)$ or $Y(2, 4)$. When Z is randomly assigned as it is in a randomized trial, we can assume that $Y(z, d) = Y(d)$, meaning that once we know d, z does not contribute any more information. This assumption is known as an exclusion restriction ([15, 18]). Thus, we will observe $Y(1)$ or $Y(3)$ for those randomized to wellness, and $Y(2)$ or $Y(4)$ for those randomized to exercise.

Potential outcomes can be used to represent causal effects. The contrast $Y(2) - Y(1)$ is the causal effect of complying with exercise versus complying with wellness. This contrast is well defined (although not observed) on an individual level. In terms of potential outcomes, the ATE is $E\{Y(2) - Y(1)\}$, where $E\{X\}$ denotes the average (or expectation) of X.

Because only one of the potential outcomes is observable for each individual, the challenge is in estimating $E\{Y(2) - Y(1)\}$.

3.2.2 Relationship Between Observed and Potential Outcome

The potential outcomes framework can be used to characterize common approaches to analysis, like the intent to treat approach or compliers only analysis. Recall that D =observed compliance. The *observed* outcome Y is therefore $Y = Y(D)$, or the potential outcome corresponding to compliance pattern D.

The intent to treat estimate is the difference in cessation rates between those randomized to exercise and those randomized to wellness: $E(Y|Z = 2) - E(Y|Z = 1)$. If we estimate the effect of treatment among compliers only,

$$E(Y|Z = 2, D = 2) - E(Y|Z = 1, D = 1) = E\{Y(2)|Z = 2, D = 2\} - E\{Y(1)|Z = 1, D = 1\}$$

.

In general this is not equal to the causal effect $E\{Y(2) - Y(1)\}$. Therefore, in order to estimate the ATE, we cannot simply estimate the effect of treatment among compliers only (since they are not equivalent).

3.2.3 A Hypothetical Example

The goal of this chapter is to describe and illustrate the GCA for estimating the ATE. Before we describe the details of this method, we will present a hypothetical example to illustrate the key concepts from the previous section. Specifically, we will demonstrate how the potential outcomes are related to non-compliance and how to calculate the causal effect using potential outcomes.

$Y(0)$	$Y(1)$	$Y(2)$	$P\{Y(0), Y(1), Y(2)\}$
0	0	0	0.55
1	0	0	0.01
0	1	0	0.05
0	0	1	0.10
0	1	1	0.15
1	0	1	0.06
1	1	0	0.03
1	1	1	0.05

Table 3.1: Hypothetical distribution of potential outcomes within the target population.

Consider a study like CTQ, in which participants are randomized to exercise or wellness and smoking cessation and compliance are measured at end of treatment. To simplify, we will define possible values of observed compliance D as :

$$d = \begin{cases} 0 & \text{if not compliant with assigned treatment} \\ 1 & \text{if complied with wellness and was expected to do so} \\ 2 & \text{if complied with exercise and was expected to do so} \end{cases}$$

Each participant has a set of potential outcomes, $\{Y(0), Y(1), Y(2)\}$. Theoretically, there is a distribution of the set of potential outcomes within the target population. Table 3.1 gives a hypothetical population distribution for the potential outcomes.

From Table 3.1 we see all 8 realizations of potential outcomes as well as the probability associated with each. In our hypothetical example, 55% of participants have

$$\{Y(0), Y(1), Y(2)\} = (0, 0, 0)$$

, meaning they would not quit regardless of compliance status on either treatment. On the

other hand, 5% of participants have

$$\{Y(0), Y(1), Y(2)\} = (1, 1, 1)$$

meaning that they would quit regardless of whether they receive either exercise or wellness or whether they do not comply with either. Furthermore, there exist participants who will quit only if they comply with wellness,

$$\{Y(0), Y(1), Y(2)\} = (0, 1, 0)$$

and those who will quit only if they comply with exercise,

$$\{Y(0), Y(1), Y(2)\} = (0, 0, 1)$$

.

From Table 3.1, we can calculate the marginal probabilities of each potential outcome by summing the relevant entries from the last column. If everyone complied with exercise, the probability of quitting would be

$$P\{Y(2) = 1\} = 0.36$$

. If all participants complied with wellness, the probability of quitting would be $P\{Y(1) = 1\} = 0.28$. Finally, if everyone was non-compliant, the probability of quitting would be $P\{Y(0) = 1\} = 0.15$. Hence, the average causal effects are:

Z	$Y(0)$	$Y(1)$	$Y(2)$	$P\{Y(0), Y(1), Y(2)\|Z\}$
$Z = 1$	0	0	0	0.55
	1	0	0	0.01
	0	1	0	0.05
	0	0	1	0.10
	0	1	1	0.15
	1	0	1	0.06
	1	1	0	0.03
	1	1	1	0.05
$Z = 2$	0	0	0	0.55
	1	0	0	0.01
	0	1	0	0.05
	0	0	1	0.10
	0	1	1	0.15
	1	0	1	0.06
	1	1	0	0.03
	1	1	1	0.05

Table 3.2: Distribution of potential outcomes within levels of Z.

$$
\begin{aligned}
E\{Y(2) - Y(0)\} &= 0.36 - 0.15 = 0.21 \\
E\{Y(1) - Y(0)\} &= 0.28 - 0.15 = 0.13 \\
E\{Y(2) - Y(1)\} &= 0.36 - 0.28 = 0.08
\end{aligned}
$$

Relative to not complying, we see a beneficial effect of complying with either treatment. In addition, the causal effect of complying with exercise versus wellness is 0.08. The latter is the effect typically of interest (and what we refer to as the ATE in this chapter). We could also summarize the average causal effects as ratios but we will focus on risk differences.

When participants are randomized to treatment arms, the potential outcomes are jointly independent of treatment assigned (Z), and the distribution of potential outcomes is the same across levels of Z (Table 3.2).

So far we have specified the joint distribution of the potential outcomes and used these probabilities to compute the average causal effects. The next step in the example is to generate a random variable D=observed compliance. The following section describes how this is accomplished. For the purposes of our illustration, we assume that compliance is a function of quit status under receipt of no treatment, $Y(0)$, and of the benefit of treatment being offered relative to no treatment, $Y(Z) - Y(0)$, according to:

$$\begin{aligned} \log \frac{P\{D=1|Z=1,Y(0),Y(1),Y(2)\}}{P\{D=0|Z=1,Y(0),Y(1),Y(2)\}} &= \alpha_0 + \alpha_1 Y(0) + \alpha_2\{Y(1) - Y(0)\} \\ \log \frac{P\{D=2|Z=2,Y(0),Y(1),Y(2)\}}{P\{D=0|Z=2,Y(0),Y(1),Y(2)\}} &= \beta_0 + \beta_1 Y(0) + \beta_2\{Y(2) - Y(0)\} \end{aligned}$$

We can think of α_0 as a "baseline" odds of compliance among those with $Y(0) = Y(1) = 0$; α_1 is the association between compliance and quit status under no treatment whereas α_2 is the association between compliance and the benefit of receiving wellness. Setting $\alpha_1 > 0$ implies that those more likely to quit are more likely to comply. Setting $\alpha_2 > 0$ implies that compliance is higher among those who stand to benefit from treatment. Parameters $\beta_0, \beta_1, \beta_2$ are similarly defined for the exercise arm. In order to fully specify the distributions in this example, we fix the values of each of these parameters as:

$$(\alpha_0, \alpha_1, \alpha_2) = (-0.90, 0.10, 0.65)$$

and

$$(\beta_0, \beta_1, \beta_2) = (-2.00, 0.25, 0.85)$$

.

These values were chosen to mimic a real-world scenario. Specifically, by forcing $(\alpha_0, \alpha_1, \alpha_2) \neq (\beta_0, \beta_1, \beta_2)$, we are assuming that compliance behavior differs by Z, conditional on

$$\{Y(0), Y(1), Y(2)\}$$

. Setting $\alpha_0 = -0.90$ and $\beta_0 = -2.00$ means that participants with $Y(0) = Y(1) = 0$ are more likely to comply with wellness compared to exercise. The choice to set $\alpha_1 < \beta_1$ and $\alpha_2 < \beta_2$ implies that compliance with exercise is more tied to cessation than compliance with wellness.

Next, we need to fill in the probability of compliance conditional on the potential outcomes and treatment. We will need these values in order to calculate the distribution of compliance within each treatment group. For example, the probability of complying with wellness among those with $Y(0) = Y(1) = 0$ is

$$\begin{aligned} P\{D = 1|Y(0) = 0, Y(1) = 0, Y(2), Z = 1\} &= \frac{\exp[\alpha_0 + \alpha_1 Y(0) + \alpha_2\{Y(1) - Y(0)\}]}{1 + \exp[\alpha_0 + \alpha_1 Y(0) + \alpha_2\{Y(1) - Y(0)\}]} \\ &= 0.29 \end{aligned}$$

Similarly, $P\{D = 2|Y(0) = 0, Y(1), Y(2) = 0, Z = 2\} = 0.12$, suggesting that a person is even less likely to comply with exercise if they would not quit regardless of whether or not they received treatment.

To complete our calculation of $D|Z$, we use the data from Table 1 and the estimates above as follows:

$$P(D = d|Z = 1) = \sum_{\{Y(0),Y(1),Y(2)\}} P\{D = d|Y(0), Y(1), Y(2), Z\} P\{Y(0), Y(1), Y(2)\}$$

which is a weighted average over the joint distribution of the potential outcomes. To complete our table, we compute the joint distribution of the potential outcomes conditional on compliance and treatment assigned as follows:

$$P\{Y(0), Y(1), Y(2)|D, Z\} = \frac{P\{D|Y(0), Y(1), Y(2), Z\}P\{Y(0), Y(1), Y(2)\}}{P(D|Z)}$$

Both of these quantities can be found in Table 3.3.

Table 3.3 contains the distribution of the potential outcomes within levels of D and Z. The last column in Table 3 is the observed response. For example, for those randomized to wellness but non-compliant, $Y = Y(0)$. Using this data we can calculate the difference in cessation rates among compliers only.

$$\begin{aligned} E\{Y|D=2, Z=2\} - E\{Y|D=1, Z=1\} &= (0.16 + 0.24 + 0.06 + 0.05) \\ &- (0.07 + 0.21 + 0.03 + 0.05) \\ &= 0.15 \end{aligned}$$

Again, this is not equal to the ATE, $E\{Y(2) - Y(1)\} = 0.08$. In general, we cannot identify or estimate the ATE from observed data.

This example shows how we would use the potential outcomes to estimate the ATE if we could actually observe $\{Y(0), Y(1), Y(2), D, Z\}$ for each participant. However, with real data (such as CTQ), we only get to observe $\{Y, D, Z\}$ and thus we need other tools for estimation.

In the next section, we will describe and illustrate the GCA for estimating the ATE from only observed data.

Z	D	$Y(0)$	$Y(1)$	$Y(2)$	$P\{Y(0),Y(1),Y(2)\|D,Z\}$	Y
$Z=1$	$D=0$	0	0	0	0.57	0
		1	0	0	0.01	1
		0	1	0	0.04	0
	$P(D=0\|Z=1)=0.69$	0	0	1	0.10	0
		0	1	1	0.12	0
		1	0	1	0.07	1
		1	1	0	0.03	1
		1	1	1	0.05	1
$Z=1$	$D=1$	0	0	0	0.51	0
		1	0	0	0.01	0
		0	1	0	0.07	1
	$P(D=1\|Z=1)=0.31$	0	0	1	0.09	0
		0	1	1	0.21	1
		1	0	1	0.04	0
		1	1	0	0.03	1
		1	1	1	0.05	1
$Z=2$	$D=0$	0	0	0	0.57	0
		1	0	0	0.01	1
		0	1	0	0.05	0
	$P(D=0\|Z=2)=0.85$	0	0	1	0.09	0
		0	1	1	0.13	0
		1	0	1	0.06	1
		1	1	0	0.03	1
		1	1	1	0.05	1
$Z=2$	$D=2$	0	0	0	0.44	0
		1	0	0	0.00	0
		0	1	0	0.04	0
	$P(D=2\|Z=2)=0.15$	0	0	1	0.16	1
		0	1	1	0.24	1
		1	0	1	0.06	1
		1	1	0	0.01	0
		1	1	1	0.05	1

Table 3.3: Distribution of potential outcomes within levels of Z,D.

3.3 The G-Computation Algorithm

The GCA is a method that uses the observed data to estimate the causal effect $E\{Y(2) - Y(1)\}$. It requires three key assumptions that we list below. When the assumptions hold, a specific type of weighted average can be used to estimate the causal effect.

3.3.1 Assumptions

The GCA requires three assumptions with respect to the potential outcomes:

1. No interference between participants (SUTVA).

2. Exclusion restriction: $Y(z, d) = Y(d)$. It means that for a given participant, once we know d, z does not contribute any information.

3. Existence of a set of covariates X such that the potential outcomes are independent of D conditional on X and Z: $Y(d) \perp D|X, Z$. It means that X explains the non-compliance within levels of Z, and within levels of X, $Y \perp D$.

3.3.2 The GCA Estimator

The GCA estimates $E\{Y(d)\}$ using a weighted average of $E(Y|X, D = d, Z)$ over $X|Z$. The $E(Y|X, D = d, Z)$ is estimable across levels of X, within treatment arm Z and under the above assumptions. If X is binary, we can estimate the sample mean of Y among those with $D = d, Z = z, X = x$ and take a weighted average over all possible values of X (for those participants with $Z = z$). It can be shown that this weighted average yield $E\{Y(d)\}$ under assumptions.

$$
\begin{aligned}
E_{X|Z}\{E(Y|Z=d, D=d, X)|Z=d\} &= E_{X|Z}\{E(Y(d)|Z=d, D=d, X)|Z=d\} && (3.1)\\
&\quad \text{by the def'n of potential outcomes} \\
&= E_{X|Z}\{E(Y(d)|Z=d, X)|Z=d\} && (3.2)\\
&\quad \text{since } Y(d) \perp D|X, Z \\
&= E\{Y(d)|Z=d\} && (3.3)\\
&\quad \text{by the law of iterated expectations} \\
&= E\{Y(d)\} \\
&\quad \text{because Z is randomized}
\end{aligned}
$$

To go from (3.2) to (3.3), we are integrating over the distribution of $X|Z$. In the case when X is binary, this integration is simply a sum.

The practical concern in estimating $E\{Y(d)\}$ using the GCA is identifying a suitable subset of covariates X that (a) meet assumption 3 and (b) can be integrated out of expression (3.1). In smoking cessation trials, we need the set of covariates X to explain the systematic variation in compliance. Clearly, the choice of X is crucial. If X is multi-dimensional and continuous, this calculation can be quite difficult (since averaging over X would amount to taking integrals over a high dimensional space). Hence, the GCA is ideal for cases in which averaging over X is computationally feasible. One such instance is when X is finite and discrete.

X	$Y(0)$	$Y(1)$	$Y(2)$	$P\{Y(0), Y(1), Y(2)\|X\}$
$X = 0$	0	0	0	0.10
	1	0	0	0.025
	0	1	0	0.095
$P(X = 0) = 0.40$	0	0	1	0.205
	0	1	1	0.30
	1	0	1	0.12
	1	1	0	0.045
	1	1	1	0.11
$X = 1$	0	0	0	0.85
	1	0	0	0.00
	0	1	0	0.02
$P(X = 1) = 0.60$	0	0	1	0.03
	0	1	1	0.05
	1	0	1	0.02
	1	1	0	0.02
	1	1	1	0.01

Table 3.4: Distribution of potential outcomes within levels of X.

3.3.3 Hypothetical Example Revisited

In this section, we return to our hypothetical example in order to illustrate how to estimate the ATE using the GCA. Suppose there exists a covariate X that is related to both compliance and outcome. In smoking cessation, this could plausibly be nicotine dependence. Instead of just randomizing participants to exercise or wellness, we use stratified randomization. Within levels of X, we randomize participants to $Z = 1, 2$. We assume that within levels of X, there is a joint distribution of the potential outcomes. For those with high levels of nicotine dependence ($X = 1$), there is a distribution of $\{Y(0), Y(1), Y(2)\}$ (and similarly for those with low levels of nicotine dependence). These distributions are summarized in Table 3.4.

From Table 3.4 we see that among those with low nicotine dependence, 10% of participants would not quit regardless of whether or not they complied with treatment offered. However, among those with high nicotine dependence, this percentage jumps to 85%. In

addition, 30% of participants with low nicotine dependence would quit if they complied with exercise or wellness, as opposed to the 5% of participants with high nicotine dependence who would do the same. Clearly, the distribution of the potential outcomes differs by X.

From Table 3.4 we see that 40% of participants have low nicotine dependence and 60% have high nicotine dependence. A simple weighted average will show that the joint distribution of the potential outcomes unconditional on X is the same as in Table 3.1. For example,

$$\begin{aligned} P[\{Y(0),Y(1),Y(2)\}=(0,0,0)] &= P[\{Y(0),Y(1),Y(2)\}=(0,0,0)|X=0]P(X=0) \\ &+ P[\{Y(0),Y(1),Y(2)\}=(0,0,0)|X=1]P(X=1) \\ &= 0.10\times 0.40+0.85\times 0.60 \\ &= 0.55 \end{aligned}$$

The covariate X can be thought of loosely as a confounder, since it is related to both compliance and the potential outcomes (details have been provided in Appendix B).

Applying GCA to this example yields

$$
\begin{aligned}
E\{Y(2)\} &= P(Y=1|D=2,Z=2,X=1)P(X=1|Z=2) \\
&+ P(Y=1|D=2,Z=2,X=0)P(X=0|Z=2) \\
&= (.03+.05+.02+.01)\times .60 + (.205+.30+.12+.11)\times 0.40 \\
&= 0.36 \\
E\{Y(1)\} &= P(Y=1|D=1,Z=1,X=1)P(X=1|Z=1) \\
&+ P(Y=1|D=1,Z=1,X=0)P(X=0|Z=1) \\
&= (.02+.05+.02+.01)\times .60 + (.095+.30+.045+.11)\times .40 \\
&= 0.28 \\
E\{Y(2)-Y(1)\} &= .08
\end{aligned}
$$

This method of computing the causal effect of receiving exercise versus wellness is the GCA estimate. Notice that we used only observed data (that is, Y, X, D and Z) and that this calculation yields exactly the same value as was calculated in the previous section (from Table 3.1). Therefore, if we can find a covariate X such that all three of the necessary assumptions are met, then we are able to estimate the causal effect of interest from only observed data.

3.3.4 Implementation with Longitudinal Data

Although our example is for the cross-sectional case, many smoking cessation trials have the advantage of collecting longitudinal data on the response. The potential outcomes framework can be extended to the longitudinal case as follows. First, define D_t as an indicator of treatment received during the period between cessation assessments at weeks

$t-1$ and t. For CTQ, we define $D_t = 2$ if the participant randomized to exercise was compliant during week t. Compliance history is denoted as $\bar{D}_t$. For example, $\bar{D}_t = 2$ if $D_1 = \cdots = D_t = 2$, meaning that the participant is compliant at each week up to time t. Cases in which changes in D are observed, are not considered compliant. As before, we define Z =assigned treatment (1 =wellness and 2 =exercise). Thus, our potential outcomes are denoted by $Y_t(z, \bar{d}_t)$ (the response at week t for a participants randomized to $Z = z$ with compliance history $\bar{d}_t$). Again, the first two assumptions required by the GCA are that there is no interference between participants and the exclusion restriction, $Y_t(z, \bar{d}_t) = Y_t(\bar{d}_t)$. The third assumption is that there exists a set of covariates X such that $Y_t(\bar{d}_t) \perp \bar{D}_t | X, Z$. This is what Robins refers to as *explainable, non-random, non-compliance*. As in the cross sectional case, the GCA estimates $E\{Y_t(\bar{d}_t)\}$ using a weighted average.

In the Section 3.3.1, we saw an example in which a single binary covariate X met the third assumption of the GCA. Now, assume we measure the response Y (quit status) at time $t = 2$ and we have a time varying covariate X_t that is measured at times $t = 0, 1$. For example, X_t could be an indicator of depression at visit t. We assume that the non-compliance is explained by both X_0 and X_1. In this case, the GCA estimate is a weighted average of Y_2 within levels of $Z, \bar{D}_2$ across all possible values of $X_0, X_1 | Z$. For those in the wellness arm for example,

$$\begin{aligned} E\{Y_2(1)\} &= E_{X_0, X_1|Z}\{E(Y_2|Z=1, \bar{D}_2 = 1, X_0, X_1)|Z=1)\} \\ &= \sum_{x_0=0}^{1} \sum_{x_1=0}^{1} E(Y_2|Z=1, \bar{D}_2 = 1, X_0 = x_0, X_1 = x_1) \\ &\times \; P(X_1 = x_1|Z=1, \bar{D}_1 = 1, X_0 = x_0) \times P(X_0 = x_0|Z=1, D_0 = 1) \end{aligned}$$

$E\{Y_2(2)\}$ can be estimated using a similar formula. The difference $E\{Y_2(2)\} - E\{Y_2(1)\}$ is the causal effect of complying with exercise versus wellness.

In a trial like CTQ, quit status at end of treatment is of particular interest. In this case, some covariates were recorded weekly over 12 weeks. Data consists of $X_1, \cdots, X_{11}, Y_{12}$. Assuming that all three of the GCA assumptions are met, we can estimate $E\{Y_{12}(\bar{d}_{12})\}$ for $\bar{d}_{12} = 1, 2$ as:

$$\begin{aligned}
E\{Y_{12}(\bar{d}_{12})\} &= E_{X_1,\cdots,X_{11}|Z}\{E(Y_{12}|Z = \bar{d}_{12}, \bar{D}_{12} = \bar{d}_{12}, X_1, \cdots, X_{11})|Z = \bar{d}_{12})\} \\
&= \sum_{x_1=0}^{1} \cdots \sum_{x_{11}=0}^{1} E(Y_{12}|Z = \bar{d}_{12}, \bar{D}_{12} = \bar{d}_{12}, X_1 = x_1, \cdots, X_{11} = x_{11}) \\
&\times P(X_{11} = x_{11}|Z = \bar{d}_{11}, \bar{D}_{11} = \bar{d}_{11}, X_1 = x_1, \cdots, X_{10} = x_{10}) \\
&\times P(X_{10} = x_{10}|Z = \bar{d}_{10}, \bar{D}_{10} = \bar{d}_{10}, X_1 = x_1, \cdots, X_9 = x_9) \\
&\times \cdots \\
&\times P(X_1 = x_1|Z = \bar{d}_1, D_1 = \bar{d}_1)
\end{aligned}$$

In our analysis of the CTQ data we assume that $X_1, X_2, \cdots, X_{11}$ are the smoking outcomes leading up to week 12 (e.g. $Y_1, Y_2, \cdots, Y_{11}$). In addition, we assume that

$$E(Y_{12}|Y_1, \cdots, Y_{11}, \bar{D}_{12} = \bar{d}_{12}) = E(Y_{12}|Y_{11}, \bar{D}_{12} = \bar{d}_{12})$$

. That is, the mean of Y at end of treatment depends only on the previous weeks quit status and compliance with assigned treatment. The GCA is feasible in this case because we are adding up a reasonable number of terms. Under our assumptions, the GCA formula reduces to

$$
\begin{aligned}
E\{Y_{12}(\bar{d}_{12})\} &= \sum_{y_1=0}^{1} \cdots \sum_{y_{10}=0}^{1} \sum_{y_{11}=0}^{1} E(Y_{12}|Z=\bar{d}_{12}, \bar{D}_{12}=\bar{d}_{12}, Y_{11}=y_{11}) \\
&\times \; P(Y_{11}=y_{11}|Z=\bar{d}_{11}, \bar{D}_{11}=\bar{d}_{11}, Y_1=y_1, Y_2=y_2, \cdots Y_{10}=y_{10}) \cdots \\
&\times \; P(Y_2=y_2|Z=\bar{d}_2, \bar{D}_2=\bar{d}_2, Y_1=y_1) \times P(Y_1=y_1|Z=\bar{d}_1, D_1=\bar{d}_1)
\end{aligned}
$$

We can think of the quantity being estimated as how well participants would have done *if* they had complied with assigned treatment throughout the study. By stratifying on cessation history, we are essentially controlling for the confounding effects of past success. We estimate the weights (probabilities of cessation history conditional on compliance) non-parametrically, whereas the expected cessation at end of treatment is estimated from a logistic regression model run among those who were fully compliant (Y_{11} included as a covariate). One question that naturally arises is whether to include other covariates (besides cessation history) to explain the non-compliance. We tested this hypothesis by fitting models that included self-efficacy, stage of change, depression, income, years of education, age, body mass index, nicotine dependence and baseline weight, as well as previous smoking status. Models suggested no additional effect of these covariates over and above the effect of smoking history. Results are presented in Section 3.4.

3.4 Results

For the purposes of our analysis, we consider CTQ participants compliant at time t if they attended 2/3 of all treatment sessions prior to week t. At end of treatment, this translates into attending at least 23 treatment sessions (there were 33 treatment sessions in the first 11 weeks and 2 in the last week. The 36^{th} treatment session was reserved for filling out end

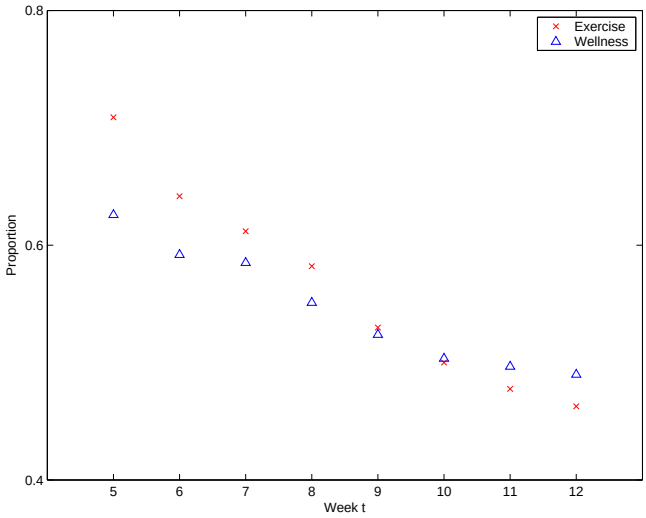

Figure 3.1: Proportion of participants fully compliant up to week t, for $t = [5, 12]$.

of treatment questionnaire packets and thus no dose of treatment was given). Figure 3.1 shows the proportion of participants who are fully compliant with assigned treatment up to week t (for $t = 5, \cdots, 12$). The pattern suggests that those on the exercise arm are more compliant early on in the trial and the wellness arm are more compliant during the later weeks. Note that by the end of treatment (week 12), only about 49% of all participants were fully compliant.

To understand how compliance was related to outcome, we modeled compliance as a function of previous weeks cessation status (separately within each treatment arm). Specifically within treatment arm $Z = d$, we fit the following model:

$$\log \frac{P(\bar{D}_t = d)}{P(\bar{D}_t \neq d)} \quad = \quad \alpha_0 + \alpha_1 Y_{t-1}$$

Using the parameter estimates from the above model, we calculate the probability of being fully compliant at each week (Table 3.5).

	$Z=1$		$Z=2$	
	Y_{t-1}		Y_{t-1}	
	0	1	0	1
$\bar{D}_5$	0.62	0.67	0.69	0.91
$\bar{D}_6$	0.53	0.81	0.51	0.95
$\bar{D}_7$	0.53	0.77	0.46	0.95
$\bar{D}_8$	0.48	0.81	0.41	0.93
$\bar{D}_9$	0.47	0.72	0.33	0.91
$\bar{D}_{10}$	0.43	0.75	0.34	0.85
$\bar{D}_{11}$	0.45	0.67	0.33	0.81
$\bar{D}_{12}$	0.42	0.74	0.28	0.82

Table 3.5: Estimated probabilities of compliance as a function of previous cessation status. Models are run separately within each treatment arm.

Table 3.5 suggests that at any given week, the probability of being fully compliant with assigned treatment is greater when the participant was quit during the previous week of the study. If we compare the exercise and wellness participants in which $Y_{t-1} = 1$, we see that at any given week, the probability of being fully compliant conditional on cessation past is greater among those randomized to exercise. However, during the majority of the study, the probability of being fully compliant among participants who report smoking during the previous week is greater among those randomized to wellness. Overall, Table 3.5 suggests the existence of a relationship between cessation and compliance. This is an assumption we are making in applying the GCA method to the CTQ data.

Using the GCA, we estimated the ATE at end of treatment. Specifically, $E\{Y_{12}(2) - Y_{12}(1)\} = 0.09$ (standard error$= 0.06$). Standard errors were estimated based on 1000 bootstrapped samples (with replacement). The GCA estimate suggests that *if* participants had been fully compliant with assigned treatment, the causal effect of receiving exercise versus wellness would have been 0.09.

We can compare the GCA estimate to the intent to treat estimate and the effect of exercise among compliers only. These three effects are summarized in Table 3.6.

Estimator	Estimate	Z-Score
Intent to Treat	0.09	1.80
$E(Y_{12}\|Z=2) - E(Y_{12}\|Z=1)$	(0.05)	
Compliers Only	0.19	2.33
$E(Y_{12}\|Z=2, \bar{D}_{12}=2) - E(Y_{12}\|Z=1, \bar{D}_{12}=1)$	(0.08)	
GCA	0.09	1.47
$E\{Y_{12}(2)\} - E\{Y_{12}(1)\}$	(0.06)	

Table 3.6: Intent to treat, compliers only and GCA estimates of the effect of exercise versus wellness. Note that the intent to treat estimator requires an assumption on the missing data. Here, we fill in missing data with treatment failures.

As mentioned previously, the GCA estimate of the ATE is 0.09 (standard error= .06). This estimate suggests that even if we could get all participants to fully comply, there would still be a subset of participants who would not successfully quit smoking. In addition, this estimate is nearly the same as the intent to treat estimate (which is averaged over compliance), suggesting that even if researchers could have gotten participants to comply, they may not have seen more substantial treatment differences than they did. Also, both the GCA estimate and the standard error are smaller than the estimate among compliers only. Estimating the effect among only those who complied with treatment is not the same as understanding what would have happened if all participants had complied. The former compares two potentially different subsets of participants: those who complied with exercise and those who complied with wellness. There is no way to know for certain whether those who complied with wellness would have complied with exercise had they been randomized to do so (and vice versa).

Using the GCA we can estimate the treatment effect at each week t, $t = 5, 6, \cdots, 12$. This may be of interest to understand whether there are weeks during which the exercise arm does significantly better than the wellness arm.

Table 3.7 summarizes the ATE at each week. The first entry in the second and third

Week t	Exercise	Wellness	Exercise-Wellness	$Z_{Estimator}$
5	GCA= 0.39(.01) ITT= 0.30(.04)	GCA= 0.30(.02) ITT= 0.22(.03)	GCA= 0.09(.02) ITT= 0.08(.05)	$Z = 4.07$ $Z = 1.60$
6	GCA= 0.39(.04) ITT= 0.31(0.04)	GCA= 0.27(.03) ITT= 0.21(.03)	GCA= 0.12(.05) ITT= 0.10(.05)	$Z = 2.52$ $Z = 2.00$
7	GCA= 0.42(.04) ITT= 0.31(.04)	GCA= 0.28(.04) ITT= 0.22(.03)	GCA= 0.14(.06) ITT= 0.20(.05)	$Z = 2.50$ $Z = 4.00$
8	GCA= 0.44(.05) ITT= 0.34(.04)	GCA= 0.26(.05) ITT= 0.22(.03)	GCA= 0.18(.07) ITT= 0.22(.05)	$Z = 2.82$ $Z = 4.40$
9	GCA= 0.38(.05) ITT= 0.31(.04)	GCA= 0.27(.04) ITT= 0.22(.03)	GCA= 0.12(.06) ITT= 0.16(.05)	$Z = 1.96$ $Z = 3.20$
10	GCA= 0.37(.05) ITT= 0.31(.04)	GCA= 0.18(.04) ITT= 0.20(.03)	GCA= 0.19(.06) ITT= 0.17(.05)	$Z = 3.05$ $Z = 3.40$
11	GCA= 0.39(.05) ITT= 0.34(.04)	GCA= 0.22(.04) ITT= 0.21(.03)	GCA= 0.17(.06) ITT= 0.18(.05)	$Z = 2.85$ $Z = 3.60$
12	GCA= 0.27(.05) ITT= 0.31(.04)	GCA= 0.17(.04) ITT= 0.22(.03)	GCA= 0.09(.06) ITT= 0.09(.05)	$Z = 1.47$ $Z = 1.80$

Table 3.7: GCA estimates at each week during the post-quit date period compared to ITT estimates at each week.

columns represents the GCA estimate of $E\{Y_t(\bar{(d)}_t = z)\}$. The value below represents the mean with treatment group, $E\{Y_t|Z = z\}$. The fourth column includes the GCA estimate of the ATE as well as the ITT estimator. We can use these to understand what the effect of treatment would have been if the study had ended early. For instance, if the study had ended at week 6, the effect of receiving exercise versus wellness would be 0.12. Focusing on column 2 (rates within the exercise arm), we will notice that the cessation rates under full compliance gradually increase until week 8, the point at which they peak. Rates more or less decrease after that point until the end of treatment. This is in contrast to the wellness arm where participants have more steady cessation rates under full compliance on a week-to-week basis. This result suggests that even if participants had been fully compliant with exercise all the way through the study, the quit rates decrease after week 8. This is an effect that would be missed by simply summarizing the data at the end of treatment or not reliably estimated if only the sub-sample of compliers had been analyzed. If we compare the two estimates (GCA and ITT) within treatment arm, we see that rates estimated using the GCA show more variability across weeks.

Table 3.7 suggests that cessation rates under full compliance with exercise drastically decrease between weeks 11 and 12. One potential explanation is that the threshold for compliance is lower during the last week of the trial. In addition, participants were given incentives to return during week 12 (compensating for the fact that attendance tends to be lower at the end of the study). For this reason, there could be an oversampling of smokers returning for their exercise sessions.

A more intuitive way to understand how the GCA works is to look at a plot of weekly cessation rates among compliers, among non-compliers and the GCA estimate of the ATE (Figure 3.2).

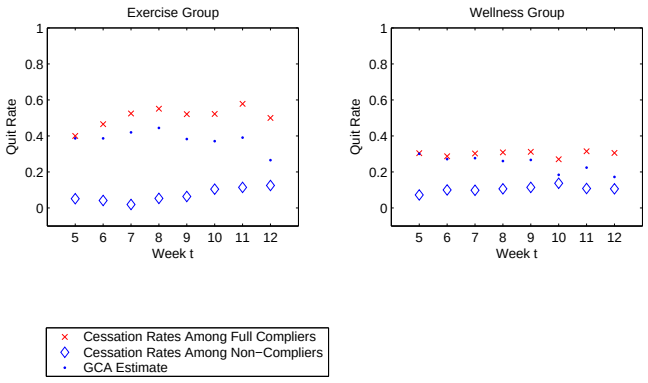

Figure 3.2: Weekly quit rates among compliers, those not fully compliant and GCA estimates under full compliance. GCA estimates pull up the cessation rates for non-compliers towards those of compliers.

Figure 3.2 is a demonstration of how the GCA method works. Cessation rates among non-compliers and full compliers are compared to the GCA estimate. Notice that the GCA estimates pull the quit rates among non-compliers up towards those of compliers. This indicates what would have happened if non-compliers had adhered to their assigned treatment.

3.5 Discussion

Researchers in smoking cessation are often interested in understanding the effects of complying with treatment (versus active control). Data are typically longitudinal, with records of both weekly quit status and attendance to treatment sessions. Although there are numerous methods for estimating the average treatment effect, the G-computation algorithm is particularly advantageous in smoking cessation trials for the following reasons. Typically, the GCA method is computationally intensive but the fact that we are summing over a binary covariate makes it feasible. In general, the GCA method is well suited to cases where there is one covariate that explains most of the non-compliance. In this case, it is smoking history. It is also particularly useful in cases where this one covariate is binary

since it allows for high dimensional summation. In addition, it allows us to compare two active treatments since the treatment contrast can be identified (so even in a trial with no pure control group, estimating the causal effect of compliance would still be possible). The GCA method works both quickly and efficiently from a computing standpoint and is easily programmed into a statistical software package (such as R or Matlab).

In estimating the effectiveness of an intervention an intent to treat approach can be used under an assumption on the missing data. One major advantage of the GCA is that compliance is directly observed. That is, if attendance is missing, it means that participants did not show up to receive treatment and hence were not compliant. Therefore, no further assumptions need to be made on the missing compliance data. Depending on how compliance is defined, there may or may not be missing data on quit status. In our analysis of the CTQ data, there were a few cases in which participants were compliant but still did not show up for their smoking measurements (and hence quit status was missing). The GCA method drops these cases but since the frequency of such cases was so low, risk of bias was limited. However, if compliance could not be directly observed or if drop-out under compliance was great, additional assumptions would be needed.

In Section 3.3.1, we outline the assumptions of the GCA. In CTQ, the assumption that units do not interact with each other may be violated. Study participants went through the trial in cohorts (exercise cohorts attended treatment sessions together and so it is unknown whether the support of fellow participants may have had an impact). As such, our results risk being biased should this assumption not hold. Robins et al. also make and justify this assumption in an analysis of mothers' stress and children's morbidity ([19]). Although we have shown that cessation history is the best predictor of future success, we are assuming that there were no other covariates that explain compliance that were unmeasured in this

trial. This is an assumption we can never test but must keep in mind while presenting these results.

The GCA method is not without its drawbacks. First, we cannot test the three assumptions on which the method hinges. As mentioned previously, the assumption of non-interference between participants may not be reasonable in group-based intervention trials. Furthermore, the crucial choice of covariate(s) X such that the potential outcomes are independent of compliance conditional on X, Z cannot be tested. Furthermore it is hard to build a sensitivity model testing this assumption. In addition to the model assumptions, the method can prove to be computationally challenging if X is a set of continuous covariates (high level integration is much more difficult than summation). Although these limitations exist, in our opinion, the advantages in analysis of data such as CTQ, are much more substantial.

The GCA is an effective method for estimating the ATE in smoking cessation trials in particular. It allows researchers to understand how well the treatment would have done if participants had been fully compliant. This not only validates the treatment options but also provides a sense of how much additional effort needs to be made in order to relieve the burden of participation. This method is recomended for smoking cessation trials when X is finite and binary. It is a useful tool in this scenario since it makes use of the fact that smoking history is a good predictor of future cessation status. Since the choice of X is crucial, it is not recommended for cases in which predictors of non-compliance are not known (or easily hypothesized). One avenue for future research would be to build in a sensitivity analysis when covariates are not sufficient to control for confounding. This could make the GCA tool more applicable in other types of behavioral trials (when less is known about X).

Chapter 4

A Multiple Imputation Approach to Mediation: Application in Behavioral Medicine

Abstract of "A Multiple Imputation Approach to Mediation: Application in Behavioral Medicine," by Shira Dunsiger, Ph.D., Brown University, May 2009

Behavioral researchers are often interested in finding mediators of treatment effect. Longitudinal clinical trials randomize participants to treatment at baseline and collect data on various measures (e.g. psychosocial constructs) at some point before end of treatment. Standard approaches for finding mediators rely on regression models which do not necessarily estimate causal effects without assumptions. In this chapter, we review these standard approaches and provide a critique using the potential outcomes framework. We summarize key statistical papers which motivate different ways in which mediation can be quantified. We focus on the descriptive direct effect of treatment and how an estimator of this effect can

be used to assess mediation in the context of a binary variable. We apply this method to data from Project STRIDE, a physical activity intervention which randomized participants to one of three treatment arms at baseline and compare results to the regression based method. The specific advantages of this method are highlighted as are future extensions.

4.1 Introduction

Mediation is a naturally arising question in the analysis of data from randomized clinical trials. Participants are randomized to the intervention at baseline and an outcome of interest is measured at some later time point. During the interim, various data are collected on participants, some of which are potential mediators of the treatment effect. The following examples are studies for which mediation can be illustrated.

Miller and colleagues ran a randomized trial aimed at evaluating the efficacy of two strategies for promoting physical activity adoption among women of pre-school aged children ([22]). At baseline, 554 women were randomized to one of three intervention groups: a print-based intervention, print materials and an invitation to attend a planning meeting (aimed at developing strategies to increase physical activity among mothers of young children), and a control group. At the end of 8 weeks, physical activity data were collected (using a modified version of the 7-day Physical Activity Recall questionnaire) on each of the women. Self-efficacy was one measure collected at both baseline and 8-week follow-up. In addition to asking whether the active intervention arms were more efficacious in promoting physical activity adoption, Miller and colleagues were interested in whether self-efficacy mediated the effects of the intervention on physical activity. That is, did the active intervention arms cause an increase in physical activity minutes (compared to a control group) by, in part, increasing the women's confidence.

Another example comes from the study Physically Active for Life (PAL) in which researchers looked at physician-based physical activity counseling amongst adults over the age of 50 ([23]). Physician practices were randomly assigned to one of two treatment groups: office-based activity counseling or control. Researchers followed 355 men and women for 8 months, at which time motivational readiness for physical activity was assessed. Pro-

cess variables, including decisional balance, were measured at 6 weeks. One question of interest was whether decisional balance mediated the effects of intervention on motivational readiness for physical activity.

A third example is Project STRIDE which was a randomized trial aimed at promoting physical activity adoption amongst previously sedentary adults ([3]). At baseline, 239 men and women were randomized to one of three treatment arms: tailored phone, tailored print, or control. The two active intervention arms (phone and print) provided two different channels through which the same materials to encourage the adoption of regular exercise were provided. The treatment phase ran for 12 months, and physical activity data were collected at 6 and 12 months. One of the primary outcomes was whether or not participants met the ACSM guideline for physical activity adoption (defined as reporting at least 150 minutes/week of at least moderate intensity activity). Primary analysis suggested that the odds of meeting criteria for physical activity adoption at 12 months was significantly greater for the print group compared to the control group. After 6 months, various psychosocial data, such as self-efficacy were collected. Whether or not self-efficacy mediated the effect of this intervention on physical activity at 12 months, is still unknown.

This chapter is organized as follows: in Section 4.2 we summarize the standard approaches for assessing mediation. After reviewing the potential outcomes framework in Section 4.3, we return to the standard methods and assess them from a causal standpoint in Section 4.4. Section 4.5 summarizes the statistical literature on direct and indirect effects. A method for estimating direct effects that would allow us to assess mediation is discussed in Section 4.6. Our analysis plan is outlined in Section 4.7 and illustrated using data from Project STRIDE in Section 4.8. Finally, the advantages of the method and possible extensions for future analyses are presented in Section 4.9.

4.2 Standard Approaches for Assessing Mediation

Baron and Kenny conceptualized mediation in a regression framework ([24]). To summarize, Baron and Kenny postulated that in order to be classified as a mediator, the variable M must meet the following three criteria:

1. Variations in Z significantly account for variations in Y; (condition 1)

2. Variations in Z significantly account for variations in M; (condition 2)

3. When M is controlled for, a previously significant relation between Z and Y is no longer significant. (condition 3)

Baron and Kenny's definition of mediation depends on how *significance* is defined and assessed. It can be thought of as an algorithm that classifies a variable M as a mediator based on p-values (for example) ([25]). This is different than using a method that employs causal thinking. We will return to this issue in the subsequent sections.

When the relationships are linear, we can use the following regression models to determine whether M meets Baron and Kenny's criteria for mediation.

1. $Y = \alpha_0 + \theta_{Y|Z} Z + e_{Y|Z}$

2. $M = \alpha_1 + \theta_{M|Z} Z + e_{M|Z}$

3. $Y = \alpha_2 + \theta_{Y|Z(M)} Z + \theta_{Y|M(Z)} M + e_{Y|MZ}$

Here we use the subscript to denote the model. That is, $\theta_{Y|Z}$ is the effect of Z on Y and $\theta_{Y|Z(M)}$ is the effect of Z on Y when controlling for M.

To establish mediation, Baron and Kenny suggest that the following conditions must be met:

i. The total effect $\theta_{Y|Z} \neq 0$

ii. The association between treatment assignment and the mediator is statistically different than zero, $\theta_{M|Z} \neq 0$

iii. The association between the mediator and the outcome is statistically different than zero, after controlling for treatment assignment, $\theta_{Y|M(Z)} \neq 0$

iv. After controlling for the mediator, the association between treatment assignment and outcome is not statistically different than zero, $\theta_{Y|Z(M)} = 0$

It has been shown that the three regression models of Baron and Kenny can be reduced as follows.

$$\begin{aligned} Y &= \alpha_2 + \theta_{Y|Z(M)}Z + \theta_{Y|M(Z)}M + e_{Y|MZ} \\ &\quad \text{From condition 3} \\ &= \alpha_2 + \theta_{Y|Z(M)}Z + \theta_{Y|M(Z)}(\alpha_1 + \theta_{M|Z}Z + e_{M|Z}) + e_{Y|MZ} \\ &\quad \text{From condition 2} \\ &= (\alpha_2 + \alpha_1\theta_{Y|MZ}) + (\theta_{Y|Z(M)} + \theta_{Y|M(Z)}\theta_{M|Z})Z + (e_{Y|MZ} + e_{M|Z}\theta_{Y|M(Z)}) \end{aligned}$$

By equating condition 1 to expression (1) above and gathering like terms, we see that,

$$\begin{aligned} \alpha_0 &= \alpha_2 + \alpha_1\theta_{Y|MZ} \\ e_{Y|Z} &= e_{Y|MZ} + e_{M|Z}\theta_{Y|MZ} \\ \theta_{Y|Z} - \theta_{Y|Z(M)} &= \theta_{Y|M(Z)}\theta_{M|Z} \end{aligned}$$

The first two expressions above are constraints on intercepts and errors respectively. The third expression gives way to other methods for establishing mediation (via the Baron

and Kenny criteria). That is, Baron and Kenny's criteria for mediation are met if $\theta_{Y|Z} - \theta_{Y|Z(M)} \neq 0$ or equivalently, $\theta_{Y|M(Z)}\theta_{M|Z} \neq 0$. The latter expression requires that $\theta_{Y|M(Z)} \neq 0$ and $\theta_{M|Z} \neq 0$. MacKinnon and colleagues ([26]) suggest that Baron and Kenny may be requiring more than is necessary to establish mediation. Specifically, mediation is possible even without a treatment effect (condition 1) because of the so-called suppressor effects. More discussion of this can be found in the literature ([27]).

Loosely speaking, Baron and Kenny quantified the percentage of variation in Y explained by M over and above Z. However, by definition, mediation is a causal process. That is, treatment assignment Z causes changes in the mediator M which in turn, cause the outcome Y. As such, causal models are required to establish mediation. There is a considerable statistical literature on this topic, specifically the literature on *direct* and *indirect* effects ([28, 29, 30]). We will return to these key papers in Section 4.5 after reviewing a potential outcomes framework for mediation.

4.3 Potential Outcomes

Potential outcomes are random variables that correspond to responses under different potential exposures ([16]). If we think of treatment assignment as the exposure in a randomized trial, then only one of these potential outcomes is observed. For example, in Project STRIDE, participants were randomized to one treatment group Z (active treatment or control). Let us assume that $Z = 1$ if participants were randomized to tailored phone or tailored print (active treatment) and $Z = 0$ if participants were randomized to contact control. Each participant has a set of potential outcomes: one outcome if they had been assigned to active treatment and one outcome if they had been assigned to control. We can think of randomization as a switch. If we turn treatment assignment to "active interven-

tion", we will see one outcome. If we switch it to "control", we will see another. We will denote these outcomes by $\{Y_1, Y_0\}$.

Although each participant has a set of potential outcomes, in actuality, only one is observed. Specifically, the response corresponding to the treatment the participant was assigned to receive. For those randomized to phone or print, the observed response Y is equal to Y_1. In general, $Y = Y_z$ where $Z = z$. Using the potential outcomes framework, we can define various contrasts that correspond to causal effects. That is, $Y_1 - Y_0$ is the causal effect of being randomized to active intervention versus control. Similarly, we can use the potential outcomes framework to define the causal effect of the intervention on the mediator as $M_1 - M_0$.

One way to conceptualize mediation is as a two-step problem. First, one must establish a causal link between the treatment assignment and the mediator. Next, a causal link between the mediator and the outcome must be established. As we will see in subsequent sections, the latter is both conceptually and operationally challenging.

4.4 Baron and Kenny Revisited

Now that we have reviewed the basic potential outcomes framework, we can return to Baron and Kenny's approach to establishing mediation. Figure 4.1 depicts the scenario in question.

That is, the total effect of Z on Y can be thought of as a sum of the direct and indirect effects. In the notation from Section 4.2 this is equivalent to

$$\theta_{Y|Z} = \theta_{Y|Z(M)} + \theta_{M|Z}\theta_{Y|M(Z)}$$

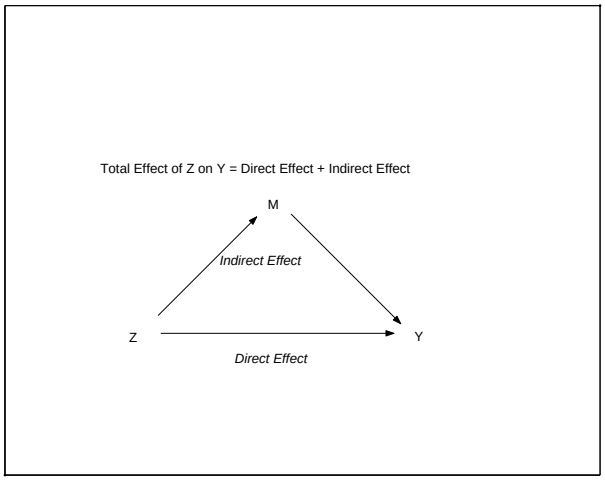

Figure 4.1: Total effect of Z on Y is the sum of the direct and indirect effects.

or

$$\theta_{Y|Z} - \theta_{Y|Z(M)} = \theta_{M|Z}\theta_{Y|M(Z)}$$

Since Z is randomization, both $\theta_{Y|Z}$ and $\theta_{M|Z}$ are causal effects. Using the potential outcomes framework, this can be shown as follows:

$$\begin{aligned}
\text{Causal effect of } Z \text{ on } Y &= E(Y_1) - E(Y_0) \\
&= E(Y_1|Z=1) - E(Y_1|Z=0) \\
&= E(Y|Z=1) - E(Y|Z=0) \\
&= \theta_{Y|Z} \\
&= \text{Total effect}
\end{aligned}$$

and

$$\begin{aligned}
\text{Causal effect of } Z \text{ on } M &= E(M_1) - E(M_0) \\
&= E(M_1|Z=1) - E(M_1|Z=0) \\
&= E(M|Z=1) - E(M|Z=0) \\
&= \theta_{M|Z}
\end{aligned}$$

Egleston et al. have shown that under certain assumptions, Baron and Kenny's direct effect ($\theta_{Y|Z(M)}$) is a causal effect itself ([29]). First we will need to introduce some new notation. In Section 4.3 we showed that potential outcomes can be indexed by Z (e.g. Y_Z, M_Z). We can also index a potential outcome by more than one variable. Specifically, Y_{Z,M_Z} which is the outcome under treatment Z and mediator value M_Z. Thus,

$$\begin{aligned}
E(Y_{z,m}) &= E(Y_{z,m}|Z=z) \\
&\quad \text{by randomization} \\
&= E(Y_{z,m}|Z=z, M_z=m) \\
&\quad \text{if } Y_{z,m} \perp M_z|Z=z \text{ for all } z,m \\
&= E(Y|Z=z, M_z=m) \\
&= \alpha_2 + \theta_{Y|Z(M)}z + \theta_{Y|M(Z)}m \\
&\quad \text{by Baron and Kenny's third regression model}
\end{aligned}$$

Therefore,

$$E(Y_{1,m}) - E(Y_{0,m}) = E(Y_{1,m} - Y_{0,m}) = \theta_{Y|Z(M)}$$

Since $\theta_{Y|Z(M)}$ does not depend on m, we must also assume that $E(Y_{1,m} - Y_{0,m})$ is the same for all m (this is Robins' no-interaction assumption) ([29]). In Section 4.5, we will see that Pearl refers to $E(Y_{1,m}) - E(Y_{0,m})$ as the Prescriptive Direct Effect ([28]), and thus, under the conditions above, Baron and Kenny's direct effect is a causal effect itself.

In Section 4.2 we saw that under Baron and Kenny's models,

$$\theta_{Y|Z} - \theta_{Y|Z(M)} = \theta_{M|Z}\theta_{Y|M(Z)}$$

. We have shown that $\theta_{Y|Z}, \theta_{Y|Z(M)}$ and $\theta_{M|Z}$ are causal effects (under certain key assumptions). However, $\theta_{Y|M(Z)}$ does not have such a clear causal interpretation since M is a post-randomization variable.

Although the randomization assumption is straightforward, it may not be reasonable to assume that $Y_{z,m} \perp M_z | Z = z$ for all z, m. If this assumption does not hold, then Baron and Kenny's model is no longer estimating a causal effect. This assumption states that within levels of Z, the potential outcomes are independent of M, a post-randomization variable. In data such as Project STRIDE, this is a strong assumption.

4.5 Conceptualizing Mediation

For the purpose of illustration, we will make use of two examples. Our first example comes from an HIV/AIDS study in which HIV infected participants were randomized to receive anti-retroviral therapy (HAART) in one of two ways: modified directly observed therapy (in which outreach workers delivered the medication and watched as participants took it) or standard of care (participants were given the medication to take on their own) ([31]). One of the primary outcomes was viral load suppression. In this case, the potential mediator

was compliance (a behavior). This is an example of complete mediation, as *taking* HAART is the only way in which treatment assignment effects viral load suppression. Here we define M as,

$$M = \begin{cases} 1 & \text{if participant took HAART} \\ 0 & \text{the participant did not take the medication} \end{cases}$$

For our second example, we return to the data from Project STRIDE. In this case, Z =treatment assignment ($Z = 1$ denotes active treatment, $Z = 0$ denotes control) and Y =minutes of at least moderate intensity physical activity (PA) at 12 months. Among other psychosocial variables, self-efficacy was measured at baseline and 6 months post-randomization. Let M be an indicator of change in self-efficacy. Specifically,

$$M = \begin{cases} 1 & \text{if self-efficacy increases from baseline} \\ 0 & \text{otherwise} \end{cases}$$

Psychosocial variables (such as self-efficacy, decisional balance) are a primary focus of psychologists. These measures are considered to be inherent characteristics/opinions of the individual, as opposed to compliance, for example, which is more of a behavior than a characteristic.

With these examples in mind, we turn to some key papers from the statistical literature on direct and indirect effects. Recall that the **direct effect** quantifies how changes in Z affect Y when M is held constant. By fixing M, the only path from Z to Y is the direct link between them. In contrast, the **indirect effect** is the effect of Z on Y that happens through M. The **total effect** of Z on Y is denoted by $Y_1 - Y_0$ and is a combination of the

direct and indirect effects of Z on Y. Since Z is randomization, methods for estimating the total effect are well established.

Assessing mediation involves estimating direct and indirect effects, which according to Pearl, can be further subdivided as: descriptive direct effects, prescriptive direct effects, descriptive indirect effects and prescriptive indirect effects ([28]).

The **prescriptive direct effect** is the effect of Z on Y when M is fixed at some predetermined (prescribed) level. Pearl denoted this effect as $Y_{1m} - Y_{0m}$ (recall that under certain assumptions this is what Baron and Kenny refer to as the direct effect. See Section 4.4). In the STRIDE example, the prescriptive direct effect measures the causal effect of being randomized to active treatment versus control on physical activity minutes when $M = m$. Since participants can only be randomized to one treatment arm, we cannot estimate this effect at the individual level. As such, Pearl defined the average prescriptive direct effect as $\triangle_m = E(Y_{1m} - Y_{0m})$. The two prescriptive direct effects in this example would be:

$\triangle_1$ = the average causal effect of treatment on physical activity minutes under the scenario that everyone's self-efficacy increases from baseline

$\triangle_0$ = the average causal effect of treatment on physical activity minutes under the scenario that no one's self-efficacy increases from baseline

However, the pair (Y_{1m}, Y_{0m}) may not be well defined for both $m = 0$ and $m = 1$ since it assumes that it is possible for everyone's mediator to take on the same value (e.g. every participant's self-efficacy will increase regardless of treatment assignment). It is conceivable

to think that there are participants enrolled in the study whose confidence for exercise will never increase even if they are randomized to active treatment. However, in the HIV example, when the mediator is a behavior such as taking medication, these effects are well defined. In fact, $E(Y_{10} - Y_{00}) = 0$, since this is an example of full mediation. That is, the only way in which treatment assignment changes the outcome is through treatment receipt. Hence, if no one took the assigned HAART, then the treatment effect would be identically zero. An advantage of using this estimator to establish mediation is that it has a clear interpretation. However, as we have seen in the case of Project STRIDE, (Y_{1m}, Y_{0m}) may not be well defined in all settings.

When prescriptive direct effects are well defined, it may be useful to use a linear structural model as follows. Suppose $E(Y_{zm}) = \alpha_0 + \alpha_1 z + \alpha_2 m + \alpha_3(zm)$; Y_{zm} can be described by an additive model of z, m and $z \times m$. Then,

$$\begin{aligned} E(Y_{11} - Y_{10}) &= \alpha_2 + \alpha_3; \\ E(Y_{01} - Y_{00}) &= \alpha_2 \end{aligned}$$

Establishing mediation would mean showing that $\alpha_2 \neq 0$ or $\alpha_3 \neq 0$. However, this is only in the case in which the potential outcome can be described by this simple additive model.

At the individual level, the **descriptive direct effect** is defined as $Y_{z,M_{z^*}} - Y_{z^*}$. When $z = 1$ and $z^* = 0$, this effect is $Y_{1,M_0} - Y_0$. In the STRIDE example, this is the difference in PA minutes between those randomized to active treatment and whose change in self-efficacy was set at the level observed for those randomized to control, versus those randomized to

control. In essence, it is what is left of the treatment effect after removing the influence M. Since Y_{1,M_0} cannot be observed at the individual level, we can estimate the average descriptive direct effect as $E(Y_{1,M_0}) - E(Y_0)$. The challenge here is in estimating $E(Y_{1,M_0})$ since M_0 is only observed for those randomized to control. We will return to this issue in Section 4.6. If we would estimate the descriptive direct effect, then we could use it to establish mediation. That is, the difference between the total effect and the descriptive direct effect, $E(Y_1 - Y_0) - E(Y_{1,M_0} - Y_0)$ represents the effect that must travel through M. If this difference is non-zero, then we have evidence for mediation.

One can think of the indirect effect as the effect of Z on Y that happens through M. Again, in accordance with Pearl, there are in theory, descriptive and prescriptive versions of this effect. In the project STRIDE example, suppose we consider participants randomized to active treatment, $Z = 1$. At the individual-level, to estimate the **descriptive indirect effect**, we must ask what would the PA minutes be for those randomized to active treatment if the change in self-efficacy were observed to be the value taken on by those randomized to control. In terms of potential outcomes, an example of the descriptive indirect effect is $Y_{1,M_0} - Y_1$. In other words, how do changes that happen to M through Z affect Y. Again, Y_{1,M_0} cannot be observed, so we can only estimate the average descriptive indirect effect, $E(Y_{1,M_0}) - E(Y_1)$. Like the descriptive direct effect, the challenge is in estimating $E(Y_{1,M_0})$.

In general, prescriptive effects are defined by fixing M at a prescribed level. In the case of direct effects, this makes sense as we are looking for the effect of Z on Y that does not happen through M. In the case of indirect effects, this is not easily interpreted. As Pearl discusses, there is no clear interpretation of the **prescriptive indirect effect** since one cannot fix M and still estimate the effect of Z on Y through M. A detailed discussion of identifiability issues can be found in ([28]).

Our two examples suggest different conceptualizations of the mediation problem. If M is a behavior (e.g. compliance) the prescriptive direct effect has a clear interpretation. When M is a characteristic or belief (e.g. self-efficacy), the descriptive direct effect is more well-suited to the problem. Deciding which conceptualization better suits the specific data under consideration, is a key step in choosing an analysis method for mediation.

Egleston et al.([29]) provide a detailed review of traditional methods in mediation (Baron and Kenny) as well as a worked example of a case in which all potential outcomes can be thought of as observable. They present an example in which UVB exposure mediates the relationship between eye glass use and risk of cataracts. The key is that there is a model that can be used to estimate the amount of UVB exposure a participant who wore sunglasses would have had, had they not worn them. That is, both potential outcomes, M_1 and M_0 can be thought of as observed. This alleviates the challenge of estimating descriptive effects since $E(Y_{1,M_0})$ can be estimated from "observed" data. However, this is not typically the case in behavioral studies. In the next section we will present an example in which we can estimate $E(Y_{1,M_0})$ without a known model of M_0 for $Z = 1$.

4.6 Descriptive Direct Effects

For this section, we will refer back to the data from Project STRIDE. Recall that $Z =$ treatment assignment, $M = I$(increase in self-efficacy from baseline to 6 months) and $Y =$ physical activity status at end of treatment. We will also define $X =$ set of baseline covariates related to M For this example, the potential mediator is an inherent characteristic of the participant (as opposed to a behavior). For this reason, we will use the *descriptive* direct effect to establish mediation (as opposed to the *prescriptive* direct effect). From Section 4.5 we know that the average descriptive direct effect can be written as $E(Y_{1,M_0}) - E(Y_0)$. The challenge is

in estimating $E(Y_{1,M_0})$ since M_0 is only observed when $Z = 1$. If we assume there exists a subset of participants for whom $M_1 = M_0$ then we could estimate $E(Y_{1,M_0})$ from observed data (the cases in which $M_1 = M_0$). In order to do so, we would need to specify the joint distribution of (M_1, M_0). Roy and colleagues have specified an association model between M_1 and M_0 conditional on X (up to a sensitivity parameter ϕ) ([32]). We will return to this shortly.

In order to estimate the descriptive direct effect, we need $E(Y_{1,M_0})$. Conditional on X,

$$
\begin{aligned}
E(Y_{1,M_0}|X) &= E_{(M_0,M_1|X)}[E(Y_{1,M_0})|M_1 = M_0, X\}] \\
&= E_{(M_0,M_1|X)}[E(Y_{1,M_1})|M_1 = M_0, X] \\
&= E(Y_{1,M_1}|M_1 = M_0 = 1, X)\frac{P(M_1 = M_0 = 1|X)}{P(M_1 = M_0|X)} + \\
&\quad E(Y_{1,M_1}|M_1 = M_0 = 0, X)\frac{P(M_1 = M_0 = 0|X)}{P(M_1 = M_0|X)} \\
&= E(Y_{1,M_1}|M_1 = M_0 = 1, X)\frac{P(M_1 = M_0 = 1|X)}{P(M_1 = M_0 = 1|X) + P(M_1 = M_0 = 0|X)} + \\
&\quad E(Y_{1,M_1}|M_1 = M_0 = 0, X)\frac{P(M_1 = M_0 = 0|X)}{P(M_1 = M_0 = 1|X) + P(M_1 = M_0 = 0|X)}
\end{aligned}
$$

Therefore, $E(Y_{1,M_0})$ is a weighted average over $(M_0, M_1|X)$. In order to estimate $E(Y_{1,M_0})$ from the observed data, we need to specify a model for $P(M_0 = M_1 = m|X)$. Roy and colleagues showed that if X is a set of baseline covariates associated with M in at least one of the treatment arms and ϕ is a sensitivity parameter ([32]), then

$$
\begin{aligned}
& P(M_1|M_0 = m, X = x) = m[P(M_1 = 1|X = x) + \phi\{U(x) - P(M_1 = 1|X = x)\}] \\
+ \quad & \frac{(1-m)}{1 - P(M_0 = 1|X = x)} \times [P(M_1 = 1|X = x) - [P(M_1 = 1|X = x) \\
+ \quad & \phi\{U(x) - P(M_1 = 1|X = x)\}]P(M_0 = 1|X = x)]
\end{aligned}
$$

Note that this association model depends on the marginal probabilities $P(M_0|X)$ and $P(M_1|X)$. Since M is binary in this case, we can estimate these marginal probabilities using logistic regression (here X are the set of predictors of M within treatment arm Z). For the STRIDE example, potential predictors of change in self-efficacy include baseline demographics (age, smoking status, race, ethnicity, marital status, education), baseline psychosocial variables (processes of change, decisional balance) and weight-related variables (BMI, percent body fat).

The estimation of the descriptive direct effect also depends on $E(Y_{1,M_1}|M_1 = M_0 = m, X)$. If both M_0, M_1 were observed, then $E(Y_{1,M_1}|M_1 = M_0 = m, X)$ is the expected Y among participants randomized to $Z = 1$ with $M_0 = M_1 = m$. By averaging over all participants, we are essentially averaging over X. Therefore, if we define:

$$
\begin{aligned}
\hat{\mu}_{1,M_1}(m) &= \hat{E}(Y_{1,M_1}|M_1 = M_0 = m) \\
&= \frac{\sum_{i=1}^{n} Z_i Y_{1,M_1,i} I(M_{1,i} = M_{0,i} = m)}{\sum_{i=1}^{n} Z_i I(M_{1,i} = M_{0.i} = m)} \\
\hat{\pi}(m) &= \frac{\sum_{i=1}^{n} Z_i I(M_{1,i} = M_{0,i} = m)}{\sum_{i=1}^{n} Z_i [I(M_{1,i} = M_{0,i} = 1) + I(M_{1,i} = M_{0,i} = 0)]}
\end{aligned}
$$

then,

$$
\hat{E}(Y_{1,M_0}) = \hat{\mu}_{1,M_1}(1)\hat{\pi}(1) + \hat{\mu}_{1,M_1}(0)\hat{\pi}(0)
$$

The estimator of the descriptive direct effect depends on the distribution of $M_0|M_1$. This distribution is outlined in detail in Roy et al., along with the assumptions made by the authors ([32]). There are five key assumptions:

1. Stable unit-treatment value assumption (SUTVA): no interference between participants.
2. Randomization holds: $Z \perp (Y_0, Y_1, M_0, M_1, X)$.
3. Exclusion restriction: $Y_{z,M_z} = Y_{M_z}$.
4. Treatment access restriction.
5. Monotonicity: $P(M_0 = 1|M_1 = 1, X = x) \geq P(M_0 = 1|M_1 = 0, X = x)$.

In the example from Project STRIDE, the first assumption states that whether or not a participants self-efficacy increased from baseline is independent of the changes in self-efficacy of other participants. This would seem reasonable since there was no interaction between participants within treatment arm. The randomization assumption is a standard assumption and is extended here to be conditional on X. The exclusion restriction says that for a given participant, once we know M_z, z does not contribute any additional information. Treatment access restriction holds for Project STRIDE since subjects in one arm were not provided materials in the same way as subjects in other arms of the study. Finally, monotonicity says that conditional on X, the probability of an increase in self-efficacy when assigned to $Z = 0$ is higher among those whose self-efficacy would increase if assigned to $Z = 1$ compared to those who would not. We will return to these assumptions in Section 4.9.

4.7 Analysis Plan

Using the data from Project STRIDE, we will estimate the descriptive direct effect of treatment assignment on physical activity behavior. Recall that the descriptive direct effect

is defined $E(Y_{1,M_0}) - E(Y_0)$. For each participant, the full data are

$$(Y_{1,M_0}, Y_{1,M_1}, Y_{0,M_0}, Y_{0,M_1}, M_1, M_0, X, Z)$$

corresponding to the set of potential outcomes for the response, the set of potential outcomes for the mediator, baseline covariates and treatment assignment respectively. However, the observed data for each participant consists only of (Y_{Z,M_Z}, M_Z, X, Z). The key missing information is $(Y_{Z,M_{1-Z}}, M_{1-Z})$. Therefore, data augmentation will be necessary in order to estimate $E(Y_{1,M_0})$. Using a multiple imputation scheme, we will estimate $E(Y_{1,M_0})$ as follows:

1. For each Z, estimate $\psi_Z(X) = P(M_Z = 1|X) = g_Z(X; \lambda_Z)$

   where g_Z is a user-defined function of X with parameters λ_Z

2. Draw$(M_Z|M_{1-Z} = m, X = x) \sim$ Bernoulli$(\pi_Z(X; \phi))$where

   $\pi_Z(X; \phi) = m[\psi_Z(x) + \phi\{U(x) - \psi_Z(x)\}]$

   $+ \quad (1-m)\dfrac{\psi_Z(x) - [\psi_Z(x) + \phi\{U(x) - \psi_Z(x)\}]\psi_{1-Z}(x)}{1 - \psi_{1-Z}(x)}$

3. Calculate I$(M_Z = M_{1-Z})$ for each participant

4. Calculate $\hat{E}(Y_{1,M_0}) = \hat{\mu}_{1,M_1}(1)\hat{\pi}(1) + \hat{\mu}_{1,M_1}(0)\hat{\pi}(0)$

5. Calculate the within sample descriptive direct effect $\hat{\theta}_{0(q)} = \hat{E}(Y_{1,M_0}) - \hat{E}(Y_0)$

6. Calculate the within sample mediated effect $\hat{\theta}_{1(q)} = \hat{E}(Y_1) - \hat{E}(Y_{1,M_0})$

7. Bootstrap the within sample standard errors of $\hat{\theta}_{0(q)}$ and $\hat{\theta}_{1(q)}$

The sampling scheme above will be repeated Q times, resulting in the following esti-

mates:

$$\begin{aligned}
\hat{\theta}_0 &= \frac{1}{Q}\sum_{q=1}^{Q}\hat{\theta}_{0(q)}, \text{with variance} \\
Var(\hat{\theta}_{0(q)}) &= E(\text{SE}(\hat{\theta}_{0(q)})^2) + (1+\frac{1}{Q})(\frac{1}{Q-1}\sum_{q=1}^{Q}(\hat{\theta}_{0(q)} - \bar{\hat{\theta}_0})^2) \\
\hat{\theta}_1 &= \frac{1}{Q}\sum_{q=1}^{Q}\hat{\theta}_{1(q)}, \text{with variance} \\
Var(\hat{\theta}_{1(q)}) &= E(\text{SE}(\hat{\theta}_{1(q)})^2) + (1+\frac{1}{Q})(\frac{1}{Q-1}\sum_{q=1}^{Q}(\hat{\theta}_{1(q)} - \bar{\hat{\theta}_1})^2)
\end{aligned}$$

4.8 Results

Analysis was conducted to assess whether an increase in self-efficacy from baseline to 6 months mediated the effect of treatment assignment on physical activity outcome at 12 months. We combined phone and print into an "active treatment" group and compared this to control. We considered both a continuous outcome

$$Y = \text{minutes of at least moderate intensity activity reported at 12 months}$$

and a binary outcome

$$Y = \text{I(meeting ACSM guidelines for physical activity at 12 months)}$$

as our estimator allows for both cases. The former was the primary outcome for Project STRIDE. In the event of missing data, we followed the strategy employed in the primary analyzes of these data and imputed missing responses (for the continuous outcome, this meant

ϕ	$E(Y_{1,M_0})$	$E(Y_1) - E(Y_0)$	$E(Y_{1,M_0}) - E(Y_0)$	$E(Y_1) - E(Y_{1,M_0})$
0	122.47 $\sigma = 19.01$	57.18 $\sigma = 20.45$ $Z = 2.80$	40.55 $\sigma = 18.94$ $Z = 2.14$	16.61 $\sigma = 12.00$ $Z = 1.38$
0.5	124.67 $\sigma = 16.54$	57.18 $\sigma = 20.45$ $Z = 2.80$	42.74 $\sigma = 20.95$ $Z = 2.04$	14.41 $\sigma = 10.30$ $Z = 1.40$
1.0	127.12 $\sigma = 15.41$	57.18 $\sigma = 20.45$ $Z = 2.80$	45.20 $\sigma = 18.08$ $Z = 2.50$	11.96 $\sigma = 7.47$ $Z = 1.60$

Table 4.1: Estimates of the total effect, descriptive direct effect and the mediated effect when comparing active treatment (phone or print) versus control. In this case $Y =$ physical activity minutes at 12 months.

carrying forward the last value; in the binary case we assumed that missing participants did not meet ACSM guidelines for physical activity). Results are provided in Table 4.1 and Table 4.2 (for various values of the sensitivity parameter ϕ).

Table 4.1 and Table 4.2 provide estimates of the descriptive direct effect based on the estimator described in Section 4.6. Recall that the descriptive direct effect can be thought of as the effect of Z on Y that *goes around* M. The difference between the total effect $E(Y_1) - E(Y_0)$ and the descriptive direct effect $E(Y_{1,M_0}) - E(Y_0)$ is $E(Y_1) - E(Y_{1,M_0})$, which represents the *mediated effect*. A significant difference between $E(Y_1)$ and $E(Y_{1,M_0})$ would provide evidence for mediation. Tables include estimates, posterior standard errors and Z scores.

Table 4.1 suggests a significant total effect of being randomized to active treatment versus control on physical activity minutes at 12 months ($57.18, \sigma = 20.45$). This is equivalent to $\theta_{Y|Z}$ from Baron and Kenny's model (see Section 4.2). The descriptive direct effect of randomization (column 6) ranges from $40.55 - 45.20$ ($\sigma = 18.94 - 20.95$), suggesting signif-

ϕ	$E(Y_{1,M_0})$	$E(Y_1) - E(Y_0)$	$E(Y_{1,M_0}) - E(Y_0)$	$E(Y_1) - E(Y_{1,M_0})$
0	0.32 $\sigma = 0.07$	0.20 $\sigma = 0.06$ $Z = 3.33$	0.16 $\sigma = 0.06$ $Z = 2.67$	0.04 $\sigma = 0.04$ $Z = 1.00$
0.5	0.32 $\sigma = 0.06$	0.20 $\sigma = 0.06$ $Z = 3.33$	0.17 $\sigma = 0.06$ $Z = 2.83$	0.03 $\sigma = 0.02$ $Z = 1.50$
1.0	0.32 $\sigma = 0.05$	0.20 $\sigma = 0.06$ $Z = 3.33$	0.17 $\sigma = 0.05$ $Z = 3.40$	0.03 $(\sigma = 0.02)$ $Z = 1.50$

Table 4.2: Estimates of the total effect, descriptive direct effect and the mediated effect when comparing active treatment versus control. In this case Y = I(meeting ACSM guidelines for physical activity at 12 months).

icant effects of Z on Y that go around M. Baron and Kenny's estimate of the direct effect, $\theta_{Y|Z(M)} = 39.20(\sigma = 21.00)$, meets the criteria that after controlling for M, the association between Z and Y is no longer significant. However, recall that $\theta_{Y|Z(M)}$ is only a causal effect when certain assumptions are met (Section 4.4) which may not be reasonable for this data set. Our estimate of the mediated effect ranges from $11.96 - 16.61$ ($\sigma = 7.47 - 12.00$) whereas Baron and Kenny's indirect effect $\theta_{Y|Z} - \theta_{Y|Z(M)} = 17.26$ ($\sigma_{bootstrap} = 7.11$). We cannot conclude (based on our estimate of the descriptive direct effect) that an increase in self-efficacy from baseline to 6 months mediates the effect of treatment assignment on physical activity outcome. Although Baron and Kenny's model suggest mediation, the estimates are only causal effects if the key assumptions outlined in Section 4.4 are met. Namely, $Y_{z,m} \perp M_z | Z = z$ for all z, m. If the assumptions hold, $\theta_{Y|Z(M)}$ = Prescriptive Direct Effect which we have suggested in Section 4.5 is more suited to a case in which M is a behavior instead of an inherent characteristic of the participant (as it may not be well-defined for the latter case). If the assumptions in Section 4.4 do not hold, then what is being estimated

from the regression approach is an association, not a causal effect.

Our multiple imputation estimator of the descriptive direct effect allows both continuous and binary outcomes Y (since we are averaging Y over a discrete set of points). For comparison, we have included estimates of the total effect, descriptive direct effect and mediated effect when $Y = \text{I(meeting ACSM guidelines for physical activity at 12 months)}$. This was a secondary outcome in Project STRIDE and is easily accommodated by our estimator. Results can be found in Table 4.2; results do not suggest a significant mediator in this case.

The method of Roy et al., allows us to estimate the joint distribution of $M_0, M_1|X$ up to a sensitivity parameter ϕ. Thus, we have provided estimates of the effects for a range of values of this parameter, $0 \leq \phi \leq 1$. We see more of an effect of ϕ for the continuous case (Table 4.1), although the conclusions regarding mediation do not differ depending on this value.

4.9 Discussion

Behavioral researchers are often interested in finding potential mediators of treatment assignment on outcome. Treatment is assigned to each participant at baseline and various psychosocial measures are collected at some point between baseline and end of treatment. Researchers have often relied on the regression methods first proposed by Baron and Kenny, in order to assess mediation. Although mediation is itself a causal process, the standard methods only estimate causal effects under certain assumptions. We have proposed an estimator of what Pearl refers to as the descriptive direct effect. That is, the effect of treatment assignment on outcome that goes around the potential mediator M. This estimator uses the joint distribution of the binary mediators M_0 and M_1 described by Roy

et al. ([32]). In conjunction with the total effect of treatment assignment on outcome, we can estimate the mediated effect, and thus provide a method for assessing mediation. We suggest that this method is particularly well-suited to the conceptualization in which M is a characteristic/belief of the participant and not a behavior.

We have demonstrated this method using data from a physical activity trial with binary outcome Y and potential mediator M. This method is well suited for the case when M is binary. However, it can easily be used when Y is either continuous or binary (since the estimator is a sum of Y over the distribution of $(M_0, M_1|X)$. This method is not computationally intensive and easily programmed into existing software (R or MATLAB).

As mentioned previously, Egleston et al. also discussed estimators of the descriptive direct effect ([29]). However, their estimator was for a specific scenario in which there exists a known model for $M_{1-Z}|Z$. Our estimator takes this one step further and describes a method for when no such model is known. In behavioral research in particular, this is a more likely scenario.

In Section 4.6, we outline the assumptions of this method. In Project STRIDE, the exclusion restriction may be violated. That is, although participants in each treatment arm were provided materials through different channels, those in the active treatment arms were given the same materials (just in two different ways). If this assumption does not hold, our results might be biased.

This method for assessing mediation is not without its drawbacks. Our paper focuses on the case when M is binary. We recognize that many of the psychosocial constructs of interest are continuous and hence we will need to focus on adapting this method for continuous data. This is one potential drawback of our data example (in which a continuous variable, self-efficacy, was transformed into a binary variable). Also, for the moment, we

assume that there is only one potential mediator M in the path between Z and Y. Much has been published with respect to multiple mediation and thus we will need to extend this estimator to the case of multiple mediators. However, even with these limitations, this method provides a practical way for establishing mediation using a causal framework.

This multiple imputation method for estimating the descriptive direct effect and consequently, establishing mediation is a needed addition to the mediation literature. It allows researchers to use a causal framework for identifying potential mediators without making the strong assumptions that the regression approach relies on.

Chapter 5

Conclusions

5.1 Discussion

This thesis develops methods for analyzing data from behavioral studies. We first developed a latent class model for finding distinct classes of behavior change amongst participants in a longitudinal smoking cessation trial. The strategy was to fit a model that forced participants who reported no behavior change into a separate class. Participants were further subdivided based on a mean model of responses over time. Treatment effects entered through the distribution of latent classes. Furthermore, we proposed a method for subdividing one class of participants based on the first-order serial correlation of the data. Second, we proposed a method for estimating the effect of receiving treatment in a longitudinal smoking cessation trial. We used the G-computation algorithm to estimate the ATE and highlight the advantages this tool has for use with smoking cessation data in particular. Third, we reviewed approaches for establishing mediation in behavioral trials and proposed a method for estimation specific to this type of data. We made use of the potential outcomes framework, as mediation is a causal process. The method was demonstrated using data from a

physical activity trial.

5.2 Direction For Future Research

Several topics warrant further research. First, a natural next step in our latent class model is to allow correlation to inform the behavior patterns in more than one class. In addition, to allow for a second degenerate class of participants, which represent those who will always succeed at changing behavior (the highly-motivated class). Furthermore, to use individual-level predictors of smoking cessation (such as weight gain and mood) to inform cessation pattern. Given the appropriate data set, all of these features could be incorporated into a single model. In addition, a more systematic approach for selecting the number of classes is warranted.

A natural extension of the G-computation approach to estimating the ATE is to allow for a set of binary and/or continuous variables to explain non-compliance. Higher level integration will be necessary for estimation as we would no longer be summing over a binary covariate. However, this addition would allow us to use this method in other scenarios (besides smoking cessation trials) wherein non-compliance is associated with other variables besides history.

The subject of mediation has been an ongoing debate for sometime. Although our proposed method takes the work of Egleston et al. ([29]) one step further, there are certainly many issues that warrant further work. First, our proposed method is specific to one binary potential mediator. Since there is a known correlation between a number of psychosocial variables (e.g. self-efficacy, behavioral processes, cognitive processes and decisional balance), a natural extension to our method would be to allow for multiple binary and/or continuous mediators. This would also address the issue of whether there are sup-

pressor effects of treatment on outcome. Our first step would be to develop a model for the joint distribution of the potential mediators conditional on a set of baseline covariates and sensitivity parameter. This distribution would allow us to draw the unobserved potential outcomes ($M_{1-Z}|Z = z$). The next step would be to define an estimator of the descriptive direct effect.

In many RCT's from behavioral medicine, potential mediators are measured over time. Hence, there are longitudinal records available for each participant. Another extension of our mediation approach would be to make use of this data in order to better understand the mediation pathway.

Appendix A

Sampling from the Posterior

A.1 Sampling Scheme: Sampling from the Posterior for LCM from Chapter 2

Sampling was carried out using the following algorithm.

1. Initial values for the level I parameters were set by fitting a longitudinal regression model to the marginal meaning using the observed data. Since β_0 and β_1 did not depend on class, the estimates from the regression model were used as initial values. For $\beta_2^{(2)}$ and $\beta_3^{(2)}$, we again used the regression parameters. $\beta_2^{(3)}$ was initialized as one standard deviation below the regression estimate and $\beta_2^{(4)}$ was one standard deviation above the parameter estimate from the model (similarly for $\beta_3^{(3)}$ and $\beta_3^{(4)}$). Initial values for the δ parameters were determined using posterior means of a 4-class LCM in which the marginal mean was assumed constant within class. We will call these β', δ'. Also, set prior distributions to be non-informative normals with large variance (10^3).

2. Initial latent classes, η', were determined for each individual. First, we calculated the empirical quit rates, $\hat{p}$, for each individual. Those with $\hat{p} = 0$ were assigned to Class 1. The remaining 3 classes were assigned based on quantiles of the distribution of $\hat{p}$.

3. Draw missing Y values (Y^{mis}). That is, conditional on initial values of the level I parameters and latent class, we sample the missing responses from the corresponding Bernoulli distribution. Specifically,

$$(Y_{i,t}^{mis}|\eta_i^{'},\beta^{'}) = \begin{cases} 0 & \text{if } \eta_i^{'}=1; \\ \sim \text{Bernoulli}(g(t,\beta^{(j)'})) & \text{if } \eta_i^{'}=j, j\geq 2. \end{cases}$$

4. Start the sampler. First, we draw β conditional on $(Y^{obs},Y^{mis}),\eta^{'}$. Specifically, we draw β from

$$\left\{\prod_{i=1}^{N} \quad \left\{\prod_{j=2}^{J}[L_i^{(j)}(\beta_0)L_i^{(j)}(\beta_1)L_i^{(j)}(\beta_2^{(j)},\beta_3^{(j)})]^{I(\eta_i=j)}\right\}\right\}\left\{\exp(\frac{-(\beta)^2}{2\times 10^3})\right\}$$

using a Normal Approximation (with mean equal to the mode of the above distribution and the variance equal to the negative inverse information).

5. Draw δ conditional on $(Y^{obs},Y^{mis}),\eta^{'}$ from

$$\left\{\prod_{i=1}^{N} \quad \left\{\prod_{j=1}^{J}[L_i^{(j)}(\delta_0^{(j)},\delta_1^{(j)})]^{I(\eta_i=j)}\right\}\right\}\left\{\exp(\frac{-(\delta)^2}{2\times 10^3})\right\}$$

using a Normal Approximation (with mean equal to the mode of the above distribution and the variance equal to the negative inverse information).

6. Draw η conditional on $(Y^{obs},Y^{mis}),\beta$ (from step 4), δ(from step 5). For example, For each i

$$P(\eta_i=2|\mid\beta,\delta)=\frac{\psi_2(X_i)\quad\prod_{t=q-2}^{T}(\pi_{it}^{(2)})^{Y_{it}}(1-\pi_{it}^{(2)})^{(1-Y_{it})}}{\psi_2(X_i)I(\quad\sum_{t=q-2}^{T}Y_{it}=0)+\quad\sum_{j=2}^{4}\psi_j(X_i)\quad\prod_{t=q-2}^{T}(\pi_{it}^{(j)})^{Y_{it}}(1-\pi_{it}^{(j)})^{(1-Y_{it})}}$$

7. Update missing data (as in step 3) using the newly sampled β,δ,η.

8. Return to step 4 and continue until convergence. This is determined by monitoring

trace plots of the draws of the sampler.

Appendix B

Relationship Between X and Compliance

B.1 Details on How X is Related to Compliance for Hypothetical Example from Chapter 3

To see how X is related to compliance, we assume

$$\begin{aligned}
\log \frac{P\{D=1|X,Z=1,Y(0),Y(1),Y(2)\}}{P\{D=0|X,Z=1,Y(0),Y(1),Y(2)\}} &= \gamma_0+\gamma_1 X+g\{Y(0),Y(1),Y(2)\} \\
\log \frac{P\{D=2|X,Z=2,Y(0),Y(1),Y(2)\}}{P\{D=0|X,Z=2,Y(0),Y(1),Y(2)\}} &= \delta_0+\delta_1 X+h\{Y(0),Y(1),Y(2)\}
\end{aligned}$$

The assumption that $Y(0),Y(1),Y(2) \perp D|X,Z$ means that $g\{Y(0),Y(1),Y(2)\} = h\{Y(0),Y(1),Y(2)\} = 0$. That is, D does not depend on $\{Y(0),Y(1),Y(2)\}$ once we have conditioned on X,Z.

We can interpret γ_0 as the odds of compliance among those randomized to wellness who have low levels of nicotine dependence($X=0$) and γ_1 as the association between compliance and nicotine dependence among those randomized to wellness. Parameters δ_0 and δ_1 are defined similarly for the exercise arm.

The values of $(\gamma_0,\gamma_1,\delta_0,\delta_1)$ are constrained by the fact that $P(D|Z)$ has already been

determined (see Section 3.2.1). Values of $(\gamma_0, \gamma_1, \delta_0, \delta_1)$ therefore must satisfy

$$\begin{aligned}
0.31 &= E_{Y(0),Y(1),Y(2)|Z=1}\{E(D|Z=1,Y(0),Y(1),Y(2)\} \\
&= E_{X|Z=1}\{E(D|Z=1,X)\} \\
&= P(D=1|Z=1) \\
&= P\{D=1|X=0,Z=1,Y(0),Y(1),Y(2)\}P(X=0|Z=1) \\
&+ P\{D=1|X=1,Z=1,Y(0),Y(1),Y(2)\}P(X=1|Z=1) \\
&= \frac{\exp(\gamma_0)}{1+\exp(\gamma_0)} \times 0.40 + \frac{\exp(\gamma_0+\gamma_1)}{1+\exp(\gamma_0+\gamma_1)} \times 0.60 \\
0.15 &= P(D=2|Z=2) \\
&= P\{D=2|X=0,Z=2,Y(0),Y(1),Y(2)\}P(X=0|Z=2) \\
&+ P\{D=2|X=1,Z=2,Y(0),Y(1),Y(2)\}P(X=1|Z=2) \\
&= \frac{\exp(\delta_0)}{1+\exp(\delta_0)} \times 0.40 + \frac{\exp(\delta_0+\delta_1)}{1+\exp(\delta_0+\delta_1)} \times 0.60
\end{aligned}$$

One set of possible solutions is $(\gamma_0, \gamma_1, \delta_0, \delta_1) = (-0.41, -1.10, -1.39, -0.60)$, corresponding to

$$P(D=1|Z=1,X=0) = 0.40$$

$$P(D=1|Z=1,X=1) = 0.25$$

$$P(D=2|Z=2,X=0) = 0.20$$

and

$$P(D=2|Z=2,X=1) = 0.12$$

Note that these parameter values also satisfy

$$\begin{aligned}
P(D=0|Z=1) &= P\{D=0|X=0,Z=1,Y(0),Y(1),Y(2)\}P(X=0|Z=1) \\
&+ P\{D=0|X=1,Z=1,Y(0),Y(1),Y(2)\}P(X=1|Z=1) \\
P(D=0|Z=2) &= P\{D=0|X=0,Z=2,Y(0),Y(1),Y(2)\}P(X=0|Z=2) \\
&+ P\{D=0|X=1,Z=2,Y(0),Y(1),Y(2)\}P(X=1|Z=2)
\end{aligned}$$

Appendix C

Sample Programs

C.1 Sample Programs

C.1.1 Sample Programs for GCA Estimator of ATE

```
#MATLAB code for setting up data

%read in data through "import data"

%specify total number participants
n=281;

%set weekly response vectors
for i=1:n
    Y1(i,1)=quit(12*(i-1)+1);
    Y2(i,1)=quit(12*(i-1)+2);
    Y3(i,1)=quit(12*(i-1)+3);
    Y4(i,1)=quit(12*(i-1)+4);
    Y5(i,1)=quit(12*(i-1)+5);
    Y6(i,1)=quit(12*(i-1)+6);
    Y7(i,1)=quit(12*(i-1)+7);
    Y8(i,1)=quit(12*(i-1)+8);
    Y9(i,1)=quit(12*(i-1)+9);
    Y10(i,1)=quit(12*(i-1)+10);
    Y11(i,1)=quit(12*(i-1)+11);
    Y12(i,1)=quit(12*(i-1)+12);
end;

%code treatment assignment
for i=1:n,
    if (trtgroup(12*(i-1)+1)==1) Group(i,1)=1;
else if (trtgroup(12*(i-1)+1)==2) Group(i,1)=2;
end;
    end;
end;
```

```
%define compliance as attending at least 2/3 of all previous
%treatment sessions

for i=1:n
    if (trt(12*(i-1)+1)>=2) comply1_atleast2thds(i,1)=1;
else comply1_atleast2thds(i,1)=0;
    end;
end;

%set up response matrices for definition of compliance by
%function "Response_Matrix"

%separate by assigned treatment

y1_atleast2thds_well=y1_atleast2thds(Group==1);

y1_atleast2thds_ex=y1_atleast2thds(Group==2);

%model for Y_t given smoking history (Y_(t-1)) and compliance
%history
b_ex=glmfit(smoking_history_ex, X_ex , 'binomial')

b_well=glmfit(smoking_history_well, X_well , 'binomial')

%model parameters yield estimates of P(Y_t=1|Y_(t-1)=y, Compliance History)
%call these prob_quit_ex and prob_smoke_ex (for the exercise arm)
%and prob_quit_well, prob_smoke_well (for the wellness arm)


%run estimator separately by treatment group
%standard errors are bootstrapped

ex_2=gcomp2(y1_atleast2thds_ex,prob_quit_ex,prob_smoke_ex)
B=bootstrp(1000,'gcomp2',y1_atleast2thds_ex,prob_quit_ex,prob_smoke_ex);
SE_E_Y2_ex=std(B)

well_2=gcomp2(y1_atleast2thds_well,prob_quit_well,prob_smoke_well)
B=bootstrp(1000,'gcomp2',y1_atleast2thds_well,prob_quit_well,
prob_smoke_well);
SE_E_Y2_well=std(B)
```

C.1.2 Sample Programs for Multiple Imputation Estimator of Descriptive Direct Effect

```
%read in data through "import data"
```

```
%specify sensitivity parameter
phi=0;

%for those in active treatment, we observe Y_(1,M_1), M_1, X. So we need to
%draw M_0. Note that N_1 are the number of participants in active treatment
%and N_0 are the number of participants in control. N=N_0+N_1.

%fit logistic regression model to estimate P(M_0|X) among those in Z=0.
beta0=logist(M_0,N_0,X_0);

%estimate P(M_0=1|X)

%similarly for P(M_1|X)

%for all participants randomized to Z=1, need to draw M_0 (since cannot be
%observed). Estimate P(M_0=1|M_1=m, X) through function "association_model".

%multiple imputation estimator calls the function "effects" which estimates
%the E(Y_(1,M_0)), DDE and ME
for q=1:1000
    [ey1m0(q,1, Descriptive_Direct_Effect(q,1), Mediated_Effect(q,1)]=...
    effects(phi, y_0, y_1,pa1a0,prob0,M, Y_Z);
end;

#"effects" function
function [ey1m0, dde, me]=effects(phi, y_0, y_1,pa1a0,prob0,M, Y_Z);

for i=1:N_1

m0(i,1)=binornd(1,prob0(i,1));
sum_m0a1(i,1)=m0(i,1)+a1(i,1);

if sum_m0a1(i,1)==0 im0m1_0(i,1)=1;
else im0m1_0(i,1)=0;
end;

if sum_m0a1(i,1)==2 im0m1_1(i,1)=1;
else im0m1_1(i,1)=0;
end;

end;



%calculate the E(Y_(1,M_0)), descriptive direct effect
%and the mediated effect
```

```
ey1m0(1,1)=(sum(Y_Z.*im0m1_1(:,1))/sum(im0m1_1(:,1)))*
(sum(im0m1_1(:,1))/sum(im0m1_1(:,1)+im0m1_0(:,1)))+...
           (sum(Y_Z.*im0m1_0(:,1))/sum(im0m1_0(:,1)))*
(sum(im0m1_0(:,1))/sum(im0m1_1(:,1)+im0m1_0(:,1)));

dde(1,1)=ey1m0(1,1)-y_0;
me(1,1)=(y_1-y_0)-dde;
```

Bibliography

[1] J. Killen, T. Robinson, C. Ammerman, C. Hayward, J. Rogers, C. Stone, D. Samuels, S. Levin, S. Green, and A. Schatzberg. Randomized clinical trial of the efficacy of bupropion combined with nicotine patch in the treatment of adolescent smokers. *Journal of Consulting and Clinical Psychology*, 72(4):729–735, 2004.

[2] B. Marcus, T. King, A. Albrecht, A. Parisi, and D. Abrams. ationale, design, and baseline data for commit to quit: An exercise efficacy trial for smoking cessation among women. *Preventative Medicine*, 26:586–597, 1997.

[3] B. Marcus, M. Napolitano, B. Lewis, A. King, A. Albrecht, A. Parisi, and et al. Examination of print and telephone channels for physical activity promotion: Rationale, design and baseline data for project stride. *Contemporary Clinical Trials*, 28:90–104, 2007.

[4] R. Niaura, B. Borelli, M. Goldstein, J. DePue, J. Ockene, J. Chiles, B. Spring, D. Hedeker, N. Keuthen, J. Kristeller, A. Prochazka, and D. Abrams. Multicenter trial of fluoxetine as an adjunct to behavioral smoking cessation treatment. *Journal of Consulting and Clinical Psychology*, 70:887–896, 2002.

[5] G. Molenberghs, H. Thijs, I. Jansen, C. Beunckens, MG. Kenward, C. Mallinckrodt,

and RJ. Carroll. nalyzing incomplete longitudinal clinical trial data. *Biostatistics*, 5(3):445–464, 2004.

[6] K. Bandeen-Roche, D. Miglioretti, S. Zeger, and P. Rathouz. Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, 440:1375–1386, 1997.

[7] D. Bartholomew and M. Knott. *Latent Variable Models and Factor Analysis, 2nd Edition*. Arnold, London, 1999.

[8] C. Clogg and L. Goodman. Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association*, 79(388):762–771, 1984.

[9] C. Dayton and G. Macready. oncomitant-variable latent-class models. *Journal of the American Statistical Association*, 83(401):173–178, 1988.

[10] B. Reboussin and J. Anthony. Latent class marginal regression models for modelling youthful drug involvement and its suspected influences. *Statistics in Medicine*, 20:623–639, 2001.

[11] B. Reboussin, M. Miller, K. Lohman, and T. Have. Latent class models for longitudinal studies of the elderly with data missing at random. *Applied Statisticsy*, 51(1):69–90, 2002.

[12] A. Gelman, A. Carlin, H. Stern, and D. Rubin. *Bayesian Data Analysis*. Chapman and Hall, Florida, 2000.

[13] Centers for Disease Control and Prevention. Cigarette smoking among adults - united states, 1998. *MMWR*, 49:881–884, 2000.

[14] G. Imbens, , and J. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994.

[15] J. Angrist, G. Imbens, and D. Rubin. Identification of causal effects using instrumental variable. *Journal of the American Statistical Association*, 91:444–455, 1996.

[16] D. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:688–701, 1974.

[17] J. Robins. A new approach to causal inference in mortality studies with sustained exposure periods - application to control of the healthy worker survivor effect. *Mathematical Modeling*, 7:1393–1512, 1986.

[18] J. Robins. Correction for non-compliance in equivalence trials. *Statistics in Medicine*, 17:269–302, 1998.

[19] J. Robins, S. Greenland, and F. Hu. Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *Journal of the American Statistical Association*, 94:687–700, 1999.

[20] J. Lunceford and M. Davidian. Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine*, 23(19):2937–2960, 2004.

[21] B. Marcus, A. Albrecht, T. King, A. Parisi, B. Pinto, M. Roberts, R. Niaura, and D. Abrams. The efficacy of exercise as an aid for smoking cessation in women. *Archives of Internal Medicine*, 159:1229–1234, 1999.

[22] Y.D. Miller, S.G. Trost, and W.J. Brown. Mediators of physical activity behavior

change among women with young children. *American Journal of Preventative Medicine*, 23(2S):98–103, 2002.

[23] B.M. Pinto, H. Lynn, B. Marcus, J. DePue, and M.G. Goldstein. Physician-based activity counseling: Intervention effects on mediators of motivational readiness for physical activity. *Annals of Behavioral Medicine*, 23(1):2–10, 2001.

[24] R.M. Baron and D.A. Kenny. The moderator-mediator variable distinction in social psychological research conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6):1173–1182, 1986.

[25] H.C. Kraemer, G.T. Wilson, C.G. Fairburn, and W.S. Agras. Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry*, 59:877–883, 2002.

[26] D.P. MacKinnon, C.M. Lockwood, J.M. Hoffman, S.G. West, and V. Sheets. A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7(1):83–104, 2002.

[27] M.A. Napolitano, G.D. Papandonatos, B.A. Lewis, J. A. Whiteley, D.M. Williams, A.C. King, B.C. Bock, B. Pinto, and B. Marcus. Mediators of physical activity behavior change: A multivariate approach. *Health Psychology*, 27(4):409–418, 2008.

[28] Pearl J. Direct and indirect effects. *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (San Fransisco, CA: Morgan Kaufmann)*, pages 411–420, 2001.

[29] B. Egleston, D.O. Scharfstein, B. Munoz, and S. West. Investigating mediation when counterfactuals are not metaphysical: Does sunlight uvb exposure mediate the effect

of eyeglasses on cataracts? *Johns Hopkins University, Department of Biostatistics working paper*, 2006.

[30] S. Goetgeluk, S. Vansteelandt, and E. Goetghebeur. Estimation of controlled direct effects. *Harvard University Biostatistics Working Paper Series*, 2008.

[31] G.E. Macalino, J.W. Hogan, J.A. Mitty, L.B. Bazerman, A.K. DeLong, H. Loewenthal, A.M. Caliendo, and T.P. Flanigan. A randomized clinical trial of community-based directly observed therapy as as adherence intervention for haart among substance users. *AIDS*, 21:1473–1477, 2007.

[32] J. Roy, J. Hogan, and B. Marcus. Principal stratification with predictors of compliance for randomized trials with 2 active treatments. *Biostatistics*, 9(2):277–289, 2007.

[33] A. Albrecht, B. Marcus, M. Roberts, D. Forman, and A. Parisi. Effect of smoking cessation on exercise performance in female smokers participating in exercise training. *American Journal of Cardiology*, 82:950–955, 1998.

[34] C. Bock, B. Marcus, T. King, B. Borrelli, and M. Roberts. Exercise effects on withdrawal and mood among women attempting smoking cessation. *Addictive Behaviors*, 24(3):399–410, 1999.

[35] B. Borrelli, B. Marcus, M. Clark, B. Bock, T. King, and M. Roberts. History of depression and subsyndromal depression in women smokers. *Addictive Behaviors*, 24(6):781–794, 1999.

[36] V. Dukic and J. Hogan. A hierarchical bayesian approach to modeling embryo implantation following in vitro fertilization. *Biostatistics*, 3:361–377, 2002.

[37] P. Congdon. *Applied Bayesian Modelling*. Wiley, London, 2003.

[38] D. Hall. Zero-inflated poisson and binomial regression with random effects: A case study. *Biometrics*, 56:1030–1039, 2000.

[39] T. King, M. Matacin, B. Marcus, B. Bock, and J. Tripolone. Body image evaluations in women smokers. *Addictive Behaviors*, 25(4):613–618, 2000.

[40] D. Lambert. Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34:1–14, 1992.

[41] P. Lazarsfeld and N. Henry. *Latent Structure Analysis*. Houghton-Mifflin, New York, 1968.

[42] W. Martinez and A. Martinez. *Computational Statistics Handbook with Matlab*. Chapman and Hall, Florida, 2002.

[43] C. McCulloch, H. Lin, E. Slate, and B. Turnbull. Discovering subpopulation structure with latent class mixed models. *Statistics in Medicine*, 21:417–429, 2002.

[44] B. Reboussin, D. Reboussin, K. Liang, and J. Anthony. A latent transition approach to modeling health-risk behavior. *Multivariate Behavioral Research*, 33(4):457–478, 1998.

[45] B. Reboussin, K. Liang, and D. Reboussin. Estimating equations fir a latent transition approach with multiple discrete indicators. *Biometrics*, 55:839–845, 1999.

[46] M. Sammel and L. Ryan. Latent variable models with fixed effects. *Biometrics*, 52(2):650–663, 2004.

[47] M. Tanner. *Tools for Statistical Inference, 3rd Edition*. Springer, New York, 1996.

[48] M. Tanner and Wing Hung. Wong. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540, 1997.

[49] J. Vermunt and J. Magidson. Latent class models for classification. *Computational Statistics and Data Analysis*, 41:531–537, 2003.

[50] J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2000.

[51] J. Vermunt and J. Magidson. Appendum to "a new approach to causal inference in mortality studies with sustained exposure periods - application to control of the healthy worker survivor effect". *Mathematical with Applications*, 14:923–945, 1987.

[52] J. Robins, M. Hernan, and B. Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11:550–560, 2000.

[53] J. Robins. Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics*, 23:2379–2412, 1994.

[54] C.E. Frangakis and D.B. Rubin. Principal stratification in causal inference. *Biometrics*, 58(1):21–29, 2002.

[55] Morgan and Winship. *Counterfactuals and Causal Inference*. Cambridge Univesrity Press, 2007.