4-12-2010

Extending the Principal Stratification Method To Multi-Level Randomized Trials

Jing Guo
University of South Florida

Extending the Principal Stratification Method

To Multi-Level Randomized Trials

by

Jing Guo

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Epidemiology and Biostatistics
College of Public Health
University of South Florida

Co-Major Professor: Hendricks Brown, Ph.D.
Co-Major Professor: Yiliang Zhu, Ph.D.
Mary E. Evans, Ph.D.
Wei Wang, Ph.D.

Date of Approval:
April 12, 2010

Keywords: Causal Effect, CACE, Multi-Level Randomized Trials, Noncompliance,
Rubin Causal Model

Dedication

This work is dedicated to my husband and my son, Zhaohui and Avery Wang, who have been giving me their selfless support for my study and research all these years; to my parents for their unfailing support.

Acknowledgments

First of all, I would like to express sincere appreciation to my major advisor, Dr. C. Hendricks Brown, who has devoted himself on mentoring and nourishing me throughout my study and research. I could not have gone so far without his guidance and support. What I have learned from him and his research enables me to overcome many hurdles and accomplish what I have dreamed of. I believe his profound positive impact will continue to benefit my future professional work.

I would also like to thank Dr. Yiliang Zhu, my co-major advisor, whose unselfish contribution and diligent work has leaded me to the finish line. Dr. Zhu has been a mentor and a friend to me, and most of all, a great scientist.

Special thanks are given to Drs. Mary E. Evans, Wei Wang who have served as my committee members, and provided critical inputs to my study along the way.

Recognition is given to Dr. Peter A Wyman, who is co-PI of this NIMH supported project. He partially supported my study by allowing me to access the preventive trial data.

Finally, I would like to thank all faculties and staff members at the Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida for all assistances they have provided. I am very proud of being part of this community.

Table of Contents

List of Tables

List of Figures

Extending the Principal Stratification Method for Multi-Level Randomized Trials

Jing Guo

Abstract

The Principal Stratification method estimates a causal intervention effect by taking account of subjects' differences in participation, adherence or compliance. The current Principal Stratification method has been mostly used in randomized intervention trials with randomization at a single (individual) level with subjects who were randomly assigned to either intervention or control condition. However, randomized intervention trials have been conducted at group level instead of individual level in many scientific fields. This is so called "two-level randomization", where randomization is conducted at a group (second) level, above an individual level but outcome is often observed at individual level within each group. The incorrect inferences may result from the causal modeling if one only considers the compliance from individual level, but ignores it or be determine it from group level for a two-level randomized trial. The Principal Stratification method thus needs to be further developed to address this issue.

To extend application of the Principal Stratification method, this research developed a new methodology for causal inferences in two-level intervention trials which principal stratification can be formed by both group level and individual level compliance. Built on the original Principal Stratification method, the new method incorporates a range of alternative methods to assess causal effects on a population when data on exposure at the group level are incomplete or limited, and are data at individual level. We use the Gatekeeper Training Trial, as a motivating example as well as for illustration. This study is focused on how to examine the intervention causal effect for schools that varied by level of adoption of the intervention program (Early-adopter vs. Later-adopter). In our case, the traditional Exclusion Restriction Assumption for Principal Stratification method is no longer hold. The results show that the intervention had a stronger impact on Later-Adopter group than Early-Adopter group for all participated schools. These impacts were larger for later trained schools than earlier trained schools. The study also shows that the intervention has a different impact on middle and high schools.

Chapter One
Introduction

1.1 Causal Effects in Intervention Trials

In medical research, clinical trials are conducted to collect data on the safety and efficacy of new drugs or devices. Random assignment of patients is required in phase III clinical trials to produce intervention and control conditions under which patients are similar in characteristics at baseline. In theory, randomization is a powerful tool to yield statistically unbiased causal estimates of the intervention effects when clinical outcomes are compared between the intervention and control conditions.

In practice, however, evaluations could be subject to a number of problems even under randomization (Bloom, 1984). One of these problems is called “participation bias” (Brown, et al. 1999), a reference to the potential difference in outcomes due to a lack of full participation in intervention by subjects when assigned. In order to objectively evaluate the causal effect of intervention and exam how levels of participation influence the intervention effects, a key issue is to classify study subjects according to randomized assignment as well as level of participation, adherence, or compliance. Principal stratification (Frangakis, et al. 2002) is such a method, which aims to address some of the limitations of the Intent-to-Treat (ITT) method, a traditional approach in randomized clinical trials. Intent-to-Treat method evaluates the intervention effects based on intervention status originally assigned to subjects, and it ignores subjects’ compliance to the study protocol. Therefore, intervention effects estimated under ITT could be biased in the presence of subjects' non-compliance.

In order to estimate intervention effects while taking account for non-compliance, Bloom (1984) applied an instrumental variable (IV) approach in which the estimates of intervention effects are adjusted for subjects' rate of compliance. More recently, the IV approach has been refined with a more rigorous framework (Angrist, Imbens, & Rubin, 1996; Imbens & Angrist, 1994). The refined approach resulted in the estimation of complier average causal effect (CACE) which is a causal effect of intervention on the subjects who would comply with any assignment and is a likelihood-based method. CACE has led to significant improvement over the IV approach. Imbens and Rubin (1997) demonstrated CACE estimation through the maximum-likelihood (ML) estimation method using the EM algorithm as well as a Bayesian data augmentation algorithm. However, a major difficulty involved in CACE estimation is how to deal with missing data on compliance among study participants because a subject's participation or compliance is often unobservable. As the issue of non-compliance receives more attention (Frangakis, el at 2002), more and more researchers have started applying the Principal Stratification method in randomized intervention trials to evaluate intervention effect for compliers – individuals who would receive the intervention if offered. However, all previous analyses only focused on single level trials which subjects who were randomly assigned to either intervention or control condition. This study will extend Principal Stratification method on two-level trials to discuss how to define the principal strata membership, how to extend assumptions from single-level on two-level trials, and how to estimate appropriate causal effect. To illustrate the Principal Stratification method, we give two examples in the next section.

1.2. Motivating Examples

The examples given in this section utilized the Principal Stratification method on individual level to evaluate intervention effects. In the first example the control group had a zero probability to receive the intervention and in the second example the control group had same probability to receive the intervention as the intervention group. Different numbers of strata were formed due to these different situations.

1.2.1 Example One: A Case of Two Principal Strata

Conducted by the University of Michigan, the Job Search Intervention Study (JOBS- II) is a randomized field experiment intended to prevent poor mental health and to promote high-quality reemployment of unemployed workers (Vinokur et al 1995). Previous studies indicated a strong positive impact of job search intervention on high-risk workers. Those with high risk were the targets of JOBS-II. All study participants were randomly assigned into either an intervention condition or a control condition.

The intervention condition consisted of five half-day training sessions. Participants who completed at least one training session were categorized as compliers. Those who completed none were categorized as never-takers. Based on the definition, the compliance rate was 55% in this study. The participants who were assigned to the control condition did not have a chance to attend any session of the study. Therefore, their compliance status was not observable, the case of missing data. As a result, the compliance rate in the control group was assumed to be the same as that in the intervention group on the basis that the two groups were similar under randomization. A total sample size of 486 was included for analyses after removing individuals with missing records. One of the outcome variables was a baseline risk score, which could be computed based on risk variables at baseline screening and also at follow-ups to predict depressive symptoms (depression, financial strain, and assertiveness).

To apply principal stratification to evaluate the intervention effects, we first identify two principal strata. Let Z_i denote the intervention status assigned to subject i

$$Z_i = \begin{cases} = 1 & \text{Intervention} \\ = 0 & \text{Control} \end{cases}$$

and S_i the participation status of subject i

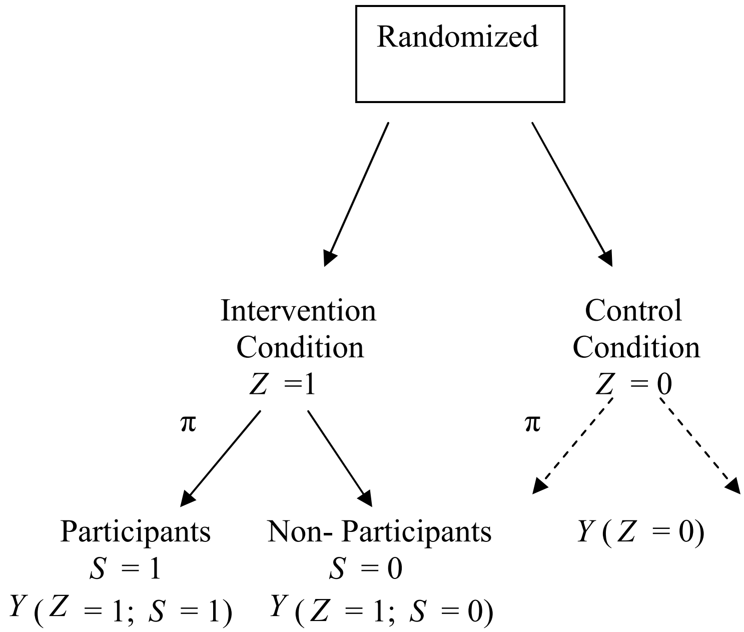
$$S_i = \begin{cases} = 1 & \text{Participants} \\ = 0 & \text{Non - Participants} \end{cases}$$

Here S is called a post-treatment variable, which is measured after treatment assignment and before the assessment of primary outcome. It is used in conjunction with Z_i to define two principal strata, compliers (or participants) and never-taker (or non-participants).

Figure 1 depicts the design where subjects were randomly assigned to either the intervention ($Z = 1$) or control ($Z = 0$) condition. The compliance rate π represents the proportion of subjects in the intervention condition who are actually compliers ($S_i = 1 | Z_i = 1$) and the rest in the group are never-takers ($S_i = 0 | Z_i = 1$). In other words, compliers are those who participate if the intervention is offered, and “never-takers” are those who would not participate when the intervention is offered. The response variable (Y) can be measured for compliers ($S = 1 | Z = 1$) and never-takers ($S = 0 | Z = 1$) in interventional condition as well as those receiving the control condition ($Z = 0$). In principle, our goal is to infer on the causal effects among "Participants", and among "Non-Participants". That is, to compare $Y(Z = 1; S = 1)$ with $Y(Z = 0; S = 1)$ and to compare $Y(Z = 1; S = 0)$ with $Y(Z = 0; S = 0)$. However, the difficulty is that the true compliance status S is unobservable in the control condition ($Z = 0$) because subjects assigned to the control condition had no chance to participate. As a result, it is not possible to distinguish compliers ($Z = 0; S = 1$) from never-takers ($Z = 0; S = 0$) within the control condition.

The main interest of JOBS-II was to estimate the causal effects of intervention for compliers, i.e., the difference between $Y(Z = 1; S = 1)$ and $Y(Z = 0; S = 1)$. This is a case of single (individual) level intervention. Principal stratum membership is more complex with a two-level study when individual participation status is combined with group level participation. The next example is a two-level intervention study.

Figure 1. Consort Diagram for JOB II Study



1.2.2 Example Two: A Case of Four Principal Strata

Many randomized experiments suffer from non-compliance (Imbens, el at., 2000). To address this problem some studies adopt an "encouragement" design that encourages subjects to change their behavior. Because the level of participation is determined by many self-selection factors, an "encouragement" design may mediate these factors, thereby encouraging a higher level of compliance or participation. We then evaluate intervention effects while adjusting for the impact of non-compliance or participation level. The study of inoculation against influenza or flu shot described below is an example of "encouragement" design study (McDonald et al, 1992). In this study, physicians were randomly selected to receive a letter that encouraged them to inoculate their patients at risk for flu. The research interest was to see if the encouragement letter (first level intervention) would lead to more patients receiving the flu shot (second level intervention). The outcome is whether the intervention would reduce flu-related hospitalization. Let Z be an indicator of the physicians' receiving the encouragement letter. If the physician of patient i received a letter, $Z_i = 1$, otherwise $Z_i = 0$. The post-treatment variable S is whether the patient received a flu shot. Let S_i be the indicator for receiving a flu shot given assignment Z_i, then $S_i = 1$ indicates the i^{th} patient received a flu shot, otherwise $S_i = 0$. Let $S_i\ (Z)$ be an indicator for the receipt of flu shot given assignment Z. If the physician of patient i did not receive an encouragement letter ($Z_i = 0$) and the patient received a flu shot $S_i\ (0) = 1$, otherwise $S_i\ (0) = 0$. Likewise, if the physician did receive an encouragement letter ($Z_i = 1$) and the patient received a flu shot $S_i\ (1) = 1$, otherwise $S_i\ (1) = 0$. The combination of these two assignments, receiving encouragement intervention (Z) and receiving flu shot (S), forms four compliance/behavior types or principal strata (C):

$$\text{Compliance/Behavior Types}\,(C_i) = \begin{cases} n \text{ (never - taker)} & \text{if } S_i\,(Z) = 0, \text{ for } Z = 0,1 \\ c \text{ (complier)} & \text{if } S_i\,(Z) = Z, \text{ for } Z = 0,1 \\ d \text{ (defier)} & \text{if } S_i\,(Z) = 1 - Z, \text{ for } Z = 0,1 \\ a \text{ (always - taker)} & \text{if } S_i\,(Z) = 1, \text{ for } Z = 0,1 \end{cases}$$

As presented in Table 1, this is an example of principal stratification with 4 strata. Membership in each stratum takes into consideration of all possible results of random assignment Z. Note that the four strata cannot be distinguished through cross-tabulation by Z and S, and is only partly observable. For example, among those patients who did not take a flu shot and whose physicians did not receive the encouragement letter ($Z = 0$; $S = 0$), some would be never takers and others would be compliers; among those patients who did not take a flu shot but their physicians received the letter ($Z = 1$; $S = 0$), some would be never takers and others would be defiers. Therefore, the observation ($Z = 0$; $S = 0$) or ($Z = 1$, $S = 0$) alone cannot determine whether a patient was a never-taker or a defier. Similarly, among those patients who took a flu shot but whose physician did not receive the letter ($Z = 0$, $S = 1$) some would be always takers and others defiers; and among those who took a flu shot but whose physicians received the letter ($Z = 1$, $S = 1$), some would be always-takers and others would be compliers. Thus again, observation ($Z = 0$, $S = 1$) or ($Z = 1$, $S = 1$) is insufficient to determine always takers. For the same reason, observed data cannot be used to distinguish compliers from always takers. In this study a set of population probabilities π is assumed under which patients may fall into one of these four principle strata. The probabilities can be dependent on patient's level covariates such as demographics and disease history (Little & Yau 1998).

Table 1: Principal Strata of the Flu Shot study Presented in 2 by 2 Table

		Control (Z = 0)	
		No Flu Shot (S = 0)	Flu Shot (S = 1)
Intervention (Z =1)	No Flu Shot (S = 0)	Never – Taker (n)	Defier (d)
	Flu Shot (S = 1)	Complier (c)	Always- Taker (a)

The response variable (Y) is a binary outcome which indicated whether a patient subsequently experienced a flu-related hospitalization during the study period. While the primary interest was to estimate intervention effects in each of the four principle strata defined by the compliance/behavior type, the four principle strata cannot be identified solely based on observed patients' flu shot status (S = 0 or 1) and their physician intervention status (receiving the encouragement letter) (Z = 0 or 1). Thus the principal stratification method aimed to estimate the causal effect in the absence of compliance/behavior classification. It further involved making an unverifiable assumption, namely that there are no defiers, in order to identify all remaining parameters. The method allowed the researchers to go beyond a standard intent-to-treat analysis and to adjust for the impact of compliance/behavior in each stratum. The researchers found strong evidence that the encouragement letter had a beneficial effect on the compliers, which is similar to that of the always-taker (Imbens el at., 2000). Based on this finding, they concluded that the flu shot had little beneficial effects on reducing flu-related hospitalization.

This second example is inherently a two-level intervention but has only been analyzed to date in a one dimensional framework by assuming that all physicians were fully participants. In other words, the post-treatment variable is at individual level, and principal stratification is thus also defined at this level – the state-of-the-art of principal stratification.

1.3 Proposed New Methodology

The preceding examples illustrate the idea and rational of principal stratification (PS) method and how it overcomes the limitations of the method of intent-to-treat in interpreting intervention effects. Previous applications of principal stratification (PS) are focused on studies principal stratification is defined using a single level post-treatment variable, often at the individual level. This dissertation extends the principal stratification method to the case where principal stratification may be defined by post-treatment variables at two levels, both group and individual level, and then develops causal inferences accordingly. Intervention trials with randomization at group level occur, yet compliance status can be determined at group level as well as at individual level. For example, we may assess an intervention that is delivered at the school level, A main challenge in such trials is that intervention exposure (participation level) can differ at the group level as well as at individual level; when this happens it is important, difficult but not impossible to distinguish the role that individual level self-selection factors and the role that the group level factors play in these groups in their impact on the effects of intervention. Our new method incorporates a range of models to evaluate causal effects on a population when intervention exposure (participation or compliance level) is determined at both group and individual levels, and moreover may not be completely observable at either level. To illustrate our proposed method, we use as a motivating example the Gatekeeper Training Trial - a multi-level randomized trial to improve services to middle and high school children with suicidal ideation in Cobb County, Georgia, USA. This Gatekeeper Training Trial is delivered to school teachers, who in turn provide QPR (Question, Persuade, and Refer) service to the students. Evaluation is focused on the school-level service of QPR.

One technical aspect of the evaluation of QPR training is to examine the intervention effect in association with the timing of program adoption (early-adopter vs. later-adopter). The varying time of program adoption implies the intervention status changes over time, i.e. exposure level potentially varies between early-adopters and later-adopters. Moreover, participation level varies across school staffs. In this context, PS is determined at multi-level, where randomization was applied at school level with intervention occurring at school staff level, and outcomes were collected at student level. As a result, some assumptions associated with Principal Stratification method of single level may no longer hold, and need to be relaxed. We also consider a random-effects component in the models under PS method. Our approach thus differs from modeling associated with single level PS method.

This dissertation is organized as follows. Basic concepts related to Principal Stratification method are introduced and reviewed first in Chapter Two, followed by a discussion of an extension of the PS method to multi-level trials. As an illustration for our multi-level PS method, the randomized QPR Gatekeeper Training trial is discussed in detail in Chapter Three. The Gatekeeper study adopted a randomized crossover design where schools were randomly assigned to a time at which to change from control to intervention. The two-stage design allowed us to look at where there is group level variation in participation, and examine whether intervention differences between early and late trained schools in earlier phase would continue over time at individual level. The multi-level PS analysis of QPR for the Gatekeeper program is conducted and compared with two traditional methods, Intent-to-Treat (ITT) and As-Treated (AT) analyses. Discussion and conclusion are presented in Chapter Four, including limitations of the PS method.

Chapter Two
Methodology for Two-Level Principal Stratification

Principal Stratification method has been mostly used in randomized intervention trials which a single level post-treatment variable determined principal stratification, often at an individual level. This chapter examines and discusses a new methodology that extends existing Principal Stratification method to randomized trials with multiple level post-treatment variables. The discussion is focused on how participation varies at either individual level or group level in a two-level randomized trial and how principal stratification can be used to develop causal inferences in such studies. We begin with an introduction of Principal Stratification method and its underlying assumptions for a single level trial design. We revisit the two examples given in Chapter One. For historical reasons, we call the two-stratum model in the first example the Bloom's model which we believe is the earliest application of the general principal stratification approach to correct for participation bias in randomized trials (Bloom, 1984). The model in the second example with four strata is referred to as the Angrist-Imbens-Rubin (AIR) model, which is the first to formalize participation/compliance involving concepts such as monotonicity and exclusion restriction assumptions. Both Bloom's model and AIR model were developed for cases where the post-treatment variable is at the individual level (first level). We then discuss how to extend the Bloom's model and AIR model to cases where participation may be determined at both group (second) and individual level and analyze intervention causal effects using mixtures and marginal maximum likelihood as in Jo

(2002). The method is capable of considering covariate effects at individual level in making inference of causal effects in all principal strata. Finally, we demonstrate how time-effect can be incorporated in the analysis in a randomization trial with longitudinal data.

2.1 Principal Stratification Method in Single Level Trials

In a standard single level (individual level), two-arm randomized trial that tests one active intervention against a control condition, a set of individuals are randomly assigned to either intervention condition ($Z_i = 1$) or control condition ($Z_i = 0$) ($i = 1, \ldots, n$). However, among those assigned to the intervention condition, some subjects "participate" in or "comply" with the intervention, i.e., take prescribed medication or attend a designed program, while others do not participate. Intervention impact under the method of intent-to-treat (ITT) is the difference in the mean outcome between all those assigned to intervention and those assigned the control. The ITT approach ignores individuals' participation completely. An alternative to ITT is to compare the mean of those participants in the intervention arm with that of the entire control condition. Because participants may differ from nonparticipants, this estimate can still be biased and cannot be interpreted as pure causal effects. Ideally we would estimate the causal effects based on those who participate in the intervention with those controls who would participate in intervention had they been offered the intervention.

Assuming there are intervention benefits among those participating in the intervention condition, it is natural to assume that the non-participants would not receive the benefit because they have no (or minimal) exposure to the intervention. Because those who participate can differ from those who do not want to participate, any assessment of the causal effect of intervention on the participants must account for potential participation bias. The following are notations that lay out a general method for adjusting for participation bias.

Let Z represent intervention status, with $Z = 0$ and $Z = 1$ being the control and intervention condition, respectively. This notation carries over to experiments where there are two different active interventions instead of a control and an intervention condition. Let S be a post-treatment variable which is measured after the treatment assignment but before assessment of the final outcome of interest. In the present context, S is a binary indicator of participation in the intervention, with S =1 representing participation and S = 0 non-participation. We use S_i^p to represent participation status at individual level for subject i . Here the superscript " p " refers to person-level, in distinction from the " g " or "group" level that we introduce later for two-level designs. In general, participation status of an individual depends on the random assignment Z_i . Thus, $S_i^p(Z)$ is the indicator of participation of subject i given the randomly assigned condition Z_i . For example, $S_i^p(1)$ and $S_i^p(0)$ correspond to the participation status when subject i is assigned to intervention or control, respectively. We use the term potential mediator outcome $S_i^p(Z)$ to represent participation status across the two intervention conditions. Note that an individual can be assigned to either $Z = 0$ or $Z = 1$, and can be either a participant or non-participant. Thus $S_i^p(0)$ and $S_i^p(1)$ are the potential mediator outcomes. In the most general case, both can take value of 1 or 0. The potential mediator outcomes of $S_i^p(0)$ and $S_i^p(1)$ define all possible subsets of individuals, corresponding to all possible combinations of ($S_i^p(0)$, $S_i^p(1)$): (0, 0), (0, 1), (1, 0), (1, 1) (Table 2). Such a classification is called principal stratification and the resultant classes are called principal strata.

Table 2: Principal Stratification Defined on the Basis of Potential Mediator Outcome in Single-level Randomized Trials

Principal stratum membership C^p	Potential mediator outcome		Potential outcome		Individual causal effect given C^p
	$S^p(0)$	$S^p(1)$	$Y_i(0 \mid C^p)$	$Y_i(1 \mid C^p)$	
Never takers (n^p)	0	0	$Y_i(0 \mid n^p)$	$Y_i(1 \mid n^p)$	$Y_i(1 \mid n^p)$ - $Y_i(0 \mid n^p)$
Compliers (c^p)	0	1	$Y_i(0 \mid c^p)$	$Y_i(1 \mid c^p)$	$Y_i(1 \mid c^p)$ - $Y_i(0 \mid c^p)$
Always takers (a^p)	1	1	$Y_i(0 \mid a^p)$	$Y_i(1 \mid a^p)$	$Y_i(1 \mid a^p)$ - $Y_i(0 \mid a^p)$
Defiers (d^p)	1	0	$Y_i(0 \mid d^p)$	$Y_i(1 \mid d^p)$	$Y_i(1 \mid d^p)$ - $Y_i(0 \mid d^p)$

The preceding discussion illustrates that principal stratification is a cross-classification of study subjects under the potential outcome of mediator variable(s) (Frangakis, et al., 2002). Let C_i^p = ($S_i^p(0)$, $S_i^p(1)$) = (0, 0), (0, 1), (1, 0), or (1, 1) be the indicator of possible principal stratum membership of subject i. Each and every subject is then randomly assigned intervention Z, so that we observe partial information on the potential outcome of the mediator. For example, if a subject is assigned to intervention and participates, then the two possibilities are C_i^p = (0, 1) or (1, 1). That is, the participant could be either a complier or an always-taker.

Let Y represents a response variable of interest, which can be measured post-intervention on each subject. Further let $Y_i(1)$ denote the potential outcome for individual i if the intervention condition is assigned and $Y_i(0)$ the potential outcome if the control condition is assigned. Depending on the participation status S_i^p, principal effects are defined as difference between potential outcomes under intervention and under control within a principal stratum. Consider examining the individual level causal effects on each subject whose participation status is given by $C^p = c^p$, c^p = (0, 0), (0, 1), (1, 0), or (1, 1). These individual level causal effects are defined as the difference between $Y_i(1) - Y_i(0)$ for each individual i conditional on C^p given above:

$$\{Y_i(1) - Y_i(0),\ i \in c_i^p,\ \ c^p = (0, 0), (0, 1), (1, 0), \text{or } (1, 1)\} \tag{2.1}$$

The average casual effect, T, of intervention over all individuals is the expected value of the difference $Y_i(1) - Y_i(0)$ conditional on $C^p = c^p$, c^p = (0, 0), (0, 1), (1, 0), or (1, 1) :

$$\begin{aligned} T &= E(Y(1) - Y(0)) \\ &= E(Y(1)) - E(Y(0)) \end{aligned} \tag{2.2}$$

Remark. For randomized trials that we focus here, we assume the assignment to intervention condition is independent of all baseline level characteristics (Holland, 1986). There is a fundamental property we impose on the potential mediator outcome $S_i^p(Z)$. That is $S_i^p(0)$ and $S_i^p(1)$ are unrelated to the assigned intervention condition, or equivalently, an individual's participation status in the active intervention condition is assumed not to change as a function of actual assignment to the intervention condition.

Property 1: The random indicator of principal strata C_i^p, to which the i^{th} subject belongs, is independent of the actual assignment of intervention condition. That is, C_i^p is independent of the treatment assignment Z_i (Frangakis & Rubin, 2002).

Property 2 is a consequence of Property 1.

Property 2: The expected principal effect within any principal stratum, as defined in Equation (2.2), is a causal effect (Frangakis & Rubin, 2002).

There are two fundamental assumptions that underline Principal Stratification method.

Assumption 2.1. Randomized Intervention Assignment.

Here we assume that the intervention assignment Z_i is exchangeable. That is, all individuals have the same probability of assignment to intervention. Additionally, Z is independent of all other random variables at the baseline. This assumption ensures comparability of subjects between the intervention condition and the control condition prior to the delivery of the intervention (Holland, 1986; Rubin, 1974, 1978, 1980).

Assumption 2.2: Stable unit treatment value assumption (SUTVA)

Potential outcome for any individual is unrelated to the intervention assignment of other individuals in the sample (Cox 1956, Rubin 1978).

Extending the notation of potential outcome $Y(Z_i)$ to $Y_i(Z_1 = z_1, Z_2 = z_2, \ldots, Z_i = z_i, \ldots Z_n = z_n)$ as the response of subject i when all subjects in the sample are assigned to their respective intervention condition $z_1, z_2, \ldots, z_n$, SUTVA is expressed by

$$Y_i(Z_1 = z_1, Z_2 = z_2, \ldots, Z_i = z_i, \ldots Z_n = z_n) = Y_i(Z_i = z_i).$$

This also implies that stratum membership of an individual is unchanged regardless of the assignment to others. That is,

$$S_i(Z_1 = z_1, Z_2 = z_2, \ldots, Z_i = z_i, \ldots Z_n = z_n) = S_i(Z_i = z_i).$$

Put it simply, SUTVA dictates that the potential outcomes are independent among individuals in the sample. It is conceivable, however, that SUTVA could be violated in some experiments. For example, if individuals assigned to the control condition have contact with individuals assigned to the intervention condition, the controls may become disappointed for not receiving intervention when the intervention condition is perceived to be more attractive. They could respond differently if their friends were in the same control condition.

We have laid out a general framework for single level experiments with self-selection into participation status. We permit subjects in both intervention conditions to have the chance of participating or not participating. However, for some experiments individuals assigned to one condition, i.e., control, may not have the possibility of participating. In the next section we review how Bloom applied Principal Stratification method to a standard single level two-arm randomized trial where individuals assigned to control have a zero probability to be exposed to active intervention.

2.2 Bloom's Model for a Single-Level Randomized Trial with Active Intervention versus Control

As the earliest application of principal stratification method to correct for participation bias in randomized trials, the Bloom model deals with the situation where individuals assigned to the control condition have no chance to be exposed to intervention. In contrast, those assigned to the intervention condition can decide to participate or not to participate. That is, $S^p(0)$ cannot be equal to 1. Framed in statistical terms used in the preceding section to define principal strata, we have

$$\Pr(S^p(0) = 1) = 0, \Pr(S^p(0) = 0) = 1, \Pr(S^p(1) = 0) > 0, \text{ and } \Pr(S^p(1) = 1) > 0.$$

When this situation arises in an experiment, data associated with $S^p(0) = 1$ are always missing, i.e., part of the data in the strata of always-takers or defiers are not observable (Tables 3). However, intervention participation rate among individuals assigned to control condition could be borrowed from that in the intervention group as assured by randomization. Table 3 shows that data are only available in two strata instead of four as the result of $S^p(0) = 1$ not being possible. The two principal strata are compliers ($S^p(0) = 0$, $S^p(1) = 1$) and never-takers ($S^p(0) = 0$, $S^p(1) = 0$).

Table 3: Potential Values of Mediator Outcome for Bloom's Model

Principal stratum membership	Potential mediator outcome		Potential outcome		Individual causal effect given C^p
C^p	$S^p(0)$	$S^p(1)$	$Y_i(0 \mid C^p)$	$Y_i(1 \mid C^p)$	
Never takers (Non-Participation) (n^p)	0	0	$Y_i(0 \mid n^p)$	$Y_i(1 \mid n^p)$	$Y_i(1 \mid n^p)$ - $Y_i(0 \mid n^p)$
Compliers (Participation) (c^p)	0	1	$Y_i(0 \mid c^p)$	$Y_i(1 \mid c^p)$	$Y_i(1 \mid c^p)$ - $Y_i(0 \mid c^p)$
Always takers (a^p)	Not Possible*	Not Possible*			
Defiers (d^p)	Not Possible*	Not Possible*			

Note: * these strata do not exist because subjects in the control condition can never get access to the intervention.

Bloom was interested in determining how effective an intervention was for those who participated in the intervention. He introduced a method of moment estimator that produces an unbiased estimate of the causal effect of the intervention. In addition to assumptions 2.1 (Randomized Intervention Assignment) and 2.2 (SUTVA), the Bloom's model requires the following assumptions.

- *Monotonicity Assumption* (Angrist et al. 1996). Pr ($S_i^p(0) = 1$) = 0. That is, no one in the control group can receive the active intervention.
- *Strongly Ignorable Treatment Assignment* (Rubin 1978). Under the assumption *Randomization Intervention Assignment*, the potential outcomes ($Y_i(1)$, $Y_i(0)$) are independent of intervention assignment Z_i. It also implies that stratum membership is independent of intervention assignment Z_i.

The assumptions of *Randomized Assignment* and *Monotonicity* together lead to

Pr (C_i^p = "complier" | $Z_i = 1$) = Pr (C_i^p = " complier" | $Z_i = 0$), $i = 1, \ldots, n$

Define $\pi = \Pr(S^p = 1 | Z = 1)$ to be the population level participation rate under intervention condition. It represents the proportion of subjects in the entire population who would participate if assigned to the intervention condition. We consider the case where the outcome variable Y is continuous. Recall that $Y_i(0)$ and $Y_i(1)$ are the potential outcomes for subject i if assigned to control or intervention, respectively. From the definition of causal effects given in Table 2 and equation 2.2, Bloom's Average Causal Effect is in fact defined among the compliers:

$$E(Y(Z=1|c^p)) - E(Y(Z=0|c^p)) = E(Y_{c^p}(Z=1, S^p=1)) - E(Y_{c^p}(Z=0, S^p=0))$$

To estimate the causal effect above, we need to know $E(Y_{c^p}(Z=0,S^p=0))$ which is not directly available. Bloom used an additional

- *Exclusion restriction*. The Average Causal Effect among nonparticipants is zero:

  $E(Y_{n^p}(Z=1,S^p=0)) - E(Y_{n^p}(Z=0,S^p=0)) = 0$

With this exclusion restriction, we can see

$$\begin{aligned} EY(Z=0) &= E\,Y_{n^p}(Z=0)\times p_r(n^p) + E\,Y_{c^p}(Z=0)\times p_r(c^p) \\ &= E\,Y_{n^p}(Z=0,S^p=0)\times(1-\pi) + E\,Y_{c^p}(Z=0,S^p=1)\times\pi \\ &= E\,Y_{n^p}(Z=1,S^p=0)\times(1-\pi) + E\,Y_{c^p}(Z=0,S^p=1)\times\pi \end{aligned}$$

Therefore, $E\,Y_{c^p}(Z=0,S^p=1) = (EY(Z=0) - E\,Y_{n^p}(Z=1,S^p=0)(1-\pi))/\pi$

and the Average Causal Effect among the participants is given by

$$E\,(Y_{c^p}\ (Z=1,S^p=\ 1))\ \text{-}\ (\frac{E(Y(Z=0))\text{-}(1-\pi)\times E\,(Y_{n^p}\ (Z=1,S^p=\ 0))}{\pi}) \qquad (2.3)$$

Assuming that the distribution of $Y(Z=0)$, $Y_{n^p}(Z=1,\ S^p=0)$, and $Y_{c^p}(Z=1,$ $S^p=1)$ has the same variance, Note that sample means $\overline{Y}(Z=0)$, $\overline{Y}_{n^p}(Z=1,S^p=0)$, and $\overline{Y}_{c^p}(Z=1,S^p=1)$ are unbiased estimate of $E(Y(Z=0))$, $E(Y_{n^p}(Z=1,S^p=0))$, and $E(Y_{c^p}(Z=1,S=1))$, respectively. Bloom proposed the following unbiased moment estimator of the ACE

$$ACE = \overline{Y}_{c^p}\ (Z=1,S=\ 1)\ \text{-}\ (\frac{\overline{Y}(Z=0)\ \text{-}\ (1-\hat{\pi})\times\ \overline{Y}_{n^p}\ (Z=1,S=\ 0)}{\hat{\pi}}) \qquad (2.4)$$

Bloom further gave an estimate of standard error (SE) for the $A\hat{C}E$

$$\sqrt{\hat{\text{var}}\ (\overline{Y}_{c^p}\ (Z=1, S=1)) + \frac{\hat{\text{var}}(\overline{Y}(Z=0)) + (1-\hat{\pi})^2 \times \hat{\text{var}}(\overline{Y}_{n^p}\ (Z=1, S=0))}{\hat{\pi}^2}} \tag{2.5}$$

2.3 Principal Stratification in Two-Level Randomized Trials

In this section, we extend Principle Stratification to two-level randomized trials. A two-level randomized trial refers to one in which randomization is placed at a group (second) level, above the individual level, but participation can vary at both group and individual level, and moreover outcome is measured at individual level within each group. For example, physicians are randomly assigned to one of two interventions, and we assess the impact of these intervention conditions on individual patients, and individual patients can decide whether to participate. School-level randomized trial is another example where impact is assessed on students but intervention is randomized at school level, participation level can vary at school and individual levels.

Similar to the notation in Section 2.2, let Z represent random assignment of intervention at the 2nd level, where $Z = 0$ or 1, representing control and intervention, respectively. Let

$$Z_j = \begin{cases} =1 & \text{Intervention} \\ =0 & \text{Control} \end{cases}$$

be the indicator of intervention assignment for j^{th} group ($j = 1, \ldots, k$) in the trial. Further, let $S(Z)$ be the indicator of participation status after group randomization Z = 1 or 0. In a two-level randomized trial, $S = (S_j^g, S_{ij}^p)$ is a two dimensional binary indicator of participation at both second level and first level. This is a new notation introduced specifically in this dissertation to describe two-level randomized trials. In this notation, the second component S_{ij}^{p} indicates the participation level of individual subject i from group j, and the first component S_{j}^{g} represents participation level for group j. Therefore, $S = (1, 1)$ represents participation at both levels and $S = (0, 0)$ represents non-participation at neither level. In the following we use the notation ($S_{j}^{g}(Z_{j}), S_{ij}^{p}(Z_{j})$) to make S dependent on random assignment Z. Using unique combinations of potential mediator outcomes ($S_{j}^{g}(0)$, $S_{ij}^{p}(0)$, $S_{j}^{g}(1)$, $S_{ij}^{p}(1)$) we are able to defines subsets of individuals according to their potential participation status, thereby forming a principle stratification of the subjects. Table 4 summarizes the stratification with a total of 16 possible principal strata: ($S_{j}^{g}(0)$, $S_{ij}^{p}(0)$, $S_{j}^{g}(1)$, $S_{ij}^{p}(1)$ = (0, 0, 0, 0), (0, 0, 0, 1), ……..., (1, 1, 1,1). In this context, the notation for enumerating all possible principal strata in a two-level randomized trial is new, and these are summarized in Table 4. Column 1 of Table 4 is the name of each principal stratum as defined by potential mediator value and intervention condition at both group level (column 2, column 4) and individual level (column 3, column 5). The superscript p indicates the potential mediator outcome of individual level and the superscript g indicates the potential mediator outcome of group level. For example, individual subjects who are in the "group of never-takers" but are "individual complier" are denoted by $n^{g}c^{p}$. The causal effects can be defined at either individual level or group level within each principal stratum.

With a two-level trial, we consider that an individual in a group can have two potential outcomes, $Y(Z_{j} = 0)$ if the group was assigned to control and $Y(Z_{j} = 1)$ if under intervention. For the i^{th} individual at the j^{th} group, the corresponding notation of potential outcomes is $Y_{ij}(0)$, and $Y_{ij}(1)$ if assigned to control or intervention condition, respectively.

Table 4: Principal Stratification based on Potential Mediator Outcome in Two-level Randomized Trials

Principal Stratum by Level of Group and individual Participation $C^g C^p$	Potential Mediator Outcome			
	Control		Intervention	
	$S_j^g(0)$	$S_{ij}^p(0)$	$S_j^g(1)$	$S_{ij}^p(1)$
Group Never-takers Individual Never-takers ($n^g n^p$)	0	0	0	0
Group Never-takers Individual Complier ($n^g c^p$)	0	0	0	1
Group Complier Individual Never-takers ($c^g n^p$)	0	0	1	0
Group Complier Individual Compliers ($c^g c^p$)	0	0	1	1
Group Never-takers Individual Defier ($n^g d^p$)	0	1	0	0
Group Never-takers Individual Always-takers ($n^g a^p$)	0	1	0	1
Group Complier Individual Defier ($c^g d^p$)	0	1	1	0
Group Complier Individual Always-takers ($c^g a^p$)	0	1	1	1
Group Defier Individual Never-takers ($d^g n^p$)	1	0	0	0
Group Defier Individual Complier ($d^g c^p$)	1	0	0	1
Group Always-takers Individual Never-takers ($a^g n^p$)	1	0	1	0
Group Always-takers Individual Complier ($a^g c^p$)	1	0	1	1
Group Defier Individual Defiers ($d^g d^p$)	1	1	0	0
Group Defier Individual Always-takers ($d^g a^p$)	1	1	0	1
Group Always-takers Individual Defier ($a^g d^p$)	1	1	1	0
Group Always-takers Individual Always-takers ($a^g a^p$)	1	1	1	1

Note: $S_j^g(Z)$ = Group level Potential Mediator Outcome; $S_{ij}^p(Z)$ = Individual level Potential Mediator Outcome

Denote by $W_j(Z_j)$ the average potential value for the j th group, with

$$W_j(0) = \frac{1}{N_j}\sum_{i=1}^{N_j} Y_{ji} \ (Z_j = 0)$$

being the average potential value under control condition and

$$W_j(1) = \frac{1}{N_j}\sum_{i=1}^{N_j} Y_{ji} \ (Z_j = 1)$$

under the intervention. So we can in principle define group-level causal effect for each principal stratum defined in Table 4 as

$E\{(W(1) - W(0)) \mid C^g C^p\}, C^g C^p = c^g c^p, c^g c^p = (0,0,0,0), (0,0,0,1), \ldots\ldots.., (1,1,1,1)$.

The assumptions for randomized trials at individual-level (section 2.1) are restated here. Some, however, need modification in order to be applicable in two-level randomized trials.

- *Randomized Intervention Assignment*. All groups have the same probability of assignment to intervention, and that Z is drawn independently of all baseline variables prior to the assignment.
- *Stable Unit Treatment Value Assumption*. (Cox 1956, Rubin 1978). Potential outcomes for each group are unrelated to the intervention status of other groups; stratum membership of one group is unchanged regardless of other groups' assignment.

Note, however, groups assigned to control condition may not always have the same probability as groups in the intervention condition to participate active intervention.

Moreover, information on individual or group level participation cannot always be observed because of the nature of randomization and potential outcome. As the degree and amount of information about participation and compliance status vary from one study to another, the structure of principal strata also differs and so does the analysis. In the next section we discuss four such examples and illustrate four distinct models.

2.4 PS Models for Two-Level Randomized Trials with Active Intervention versus Control

This section focuses on four different models through four examples. In these examples intervention is assigned at the group level and participation status is determined by a combination of both group and individual level compliance. Across these examples, the amount of observable information on participation at each level varies. Table 5 summarizes the four examples and the associated models for analysis. In discussing these models, we present the fundamental characters and illustrate of each the examples from the literature.

Model 1 represents the situation where participation status is observed at individual level only, and is ignored (i.e., unmeasured) but can always be determined at group level. There, if individuals in a group assigned to control condition have a zero probability to receive the intervention, then Bloom's model for trials of single level randomization can be directly used. However, if individuals have the possibility to receive the intervention even when their group is assigned to control condition, the single level AIR model may be used to estimate individual level CACE (complier average causal effects) with additional assumptions. For illustration of single level AIR model, we consider the example of the Flu shot study (Hirano, el at., 2000).

Model 2 represents the situation where participation status is available at group level only, but ignored at individual level. There, when assigned to the intervention condition some groups would receive the intervention while others decline intervention and receive the same condition as if assigned to controls. In such cases, we have extended classic Bloom's model or AIR model to be applicable. For convenience, the extended models are called G-B (Group-Bloom) model or G-AIR (Group-AIR) model. We use the example of GA-gatekeeper study (Wyman, el at 2008) to illustrate the G-B model.

When intervention is only available to individuals whose group is assigned to and participates in intervention, we consider model 3 (GB-PB model), a combination of G-B model and individual-level Bloom model (PB). When individuals in any group have the same probability to receive the intervention, we consider model 4 (GA-PA model), a combination of G-AIR and individual level AIR model. We use the example of Good Behavior Game (Ialongo et al., 1999) to illustrate both model 3 and model 4.

Table 5: Summary of Models for Two-level Randomized Trials with Group and Individual Level Participation

Model		Restriction		Number of Principal Strata	Example			
		Group Level Participation	Individual Level Participation		Individual & Group Randomization	Intervention	Participation	Outcome
1 (participation status is available at individual level only)	Bloom's Model	Unmeasured	Only possible in intervention group	2 principal strata: c^p (Participation), n^p (Nonparticipation), (individual level)	Elementary school students in randomly assigned classrooms *Individual*: students; *Group*: classroom	Parent training, only available to families in assigned classrooms	Parent attendance in parent training (Ialongo et al., 1999)	Child aggressive behavior
	AIR Model		None	4 principal strata: a^p, c^p, n^p, d^p	Patients within physicians *Individual*: patients; *Group*: physicians	Physicians encouraged to have their patients get a flu shot	Patients can receive flu shot regardless of physician behavior (Hirano, el at. 2000)	Did patient get the flu?
2 (participation status is available at group level only)	G-Bloom's Model*	Only possible when assigned to intervention	Unmeasured	2 principal strata: c^g (Participantion), n^g (Nonparticipation), (group level)	Middle/High schools randomly assigned to the intervention *Individual*: school staff; *Group*: schools	Gatekeeper training program provided to all school staff in assigned schools	School staff receive training within intervention schools (Wyman, el at., 2008)	Frequency that school staff refers students for suicidal behavior
	G-AIR Model	None	Unmeasured	4 principal strata: a^g, c^g, n^g, d^g	Counties randomly assigned to one of	Community development team model	Level of implementatio n achieved by	Number of foster care placements

					two implementation strategies for an evidence-based practice *Individual*: foster care family; *Group*: county	versus individual county implementation intervention	county within 18 months (Chamberlain, el at. 2006)	made by county
3	GB-PB Model*	Only possible when assigned to intervention	Only possible when the group participates	4 principal strata: $n^g n^p$, $n^g c^p$, $c^g n^p$, $c^g c^p$	Elementary school students randomly assigned to classrooms and classrooms randomly assigned to intervention *Individual*: parent; *Group*: classroom	Parent training in behavior management	Level of program fidelity delivered by school counselors; Family attendance in training sessions (Ialongo et al., 1999)	Child aggressive behavior
4	GA-PA Model*	None	None	16 principal strata: $n^g n^p$, $n^g c^p$, $c^g n^p$, $c^g c^p$, $n^g d^p$, $n^g a^p$, $c^g d^p$, $c^g a^p$, $d^g n^p$, $d^g c^p$, $a^g n^p$, $a^g c^p$, $d^g d^p$, $d^g a^p$, $a^g d^p$, $a^g a^p$				

Note*: G = Group, B = Bloom, A = AIR

2.4.1. AIR Model

The flu shot study mentioned in Chapter 1 is an example to which the single level AIR model may be applied under additional assumptions about individual level compliance as presented in Hirano et al. (2000). Physicians were assumed to completely comply with their intervention assignment (group level). As a result, all patients were classified into four principal strata based on their compliance status (individual level) with their physicians (see Table 6). In order to extend the AIR model to the flu shot study, Hirano et al. (2000) applied specific assumptions, which we state below,

- *Monotonicity Assumption*. The monotonicity assumption at the individual level is the probability that defier is zero ($S_{ij}^{p}(Z_j = 1) \geq S_{ij}^{p}(Z_j = 0)$). In the case of the flu shot example, this simply eliminates the possibility that there are patients who would not want a flu shot if their physician would receive an encouragement letter, but otherwise would want a flu shot. In terms of probability, the assumption dictates

$$\Pr(S_{ij}^{p}(S_j^{g}(0) = 0) = 1 \ \ \& \ \ S_{ij}^{p}(S_j^{g}(1) = 1) = 0) = 0,$$

hence effectively removing the stratum of individual defier. Hirano et al. (2000) argued for the reasonableness of the monotonicity assumption. However, there would be unwanted consequences of the AIR model, as is shown later, when the monotonicity assumption is untrue.

- *Exclusion restriction*. The Average Causal Effect among always-taker ($c^g a^p$) and never-taker ($c^g n^p$) are zero. Consider Y to be a measured outcome at the individual level. We have under this exclusion restriction that

$$\begin{aligned} ACE(C = c^g a^p) &= E(Y(Z=1) \mid C^g C^p = c^g a^p) - E(Y(Z=0) \mid C^g C^p = c^g a^p) \\ &= E(Y_{c^g a^p}(Z=1)) - E(Y_{c^g a^p}(Z=0)) \\ &= \mu_{1a} - \mu_{0a} = 0 \end{aligned}$$

and

$$\begin{aligned} ACE(C = c^g n^p) &= E(Y(Z=1) \mid C^g C^p = c^g n^p) - E(Y(Z=0) \mid C^g C^p = c^g n^p) \\ &= E(Y_{c^g n^p}(Z=1)) - E(Y_{c^g n^p}(Z=0)) \\ &= \mu_{1n} - \mu_{0n} = 0 \end{aligned}$$

where μ_{1a} and μ_{0a} are the population means of potential outcome for always- taker under intervention and control condition, respectively; μ_{1n} and μ_{0n} are the population mean of never- taker under intervention and control condition, respectively. In our example, the first relation implies that the chance that a subject gets the flu given that they would have gotten the flu shot is the same regardless of whether their physician was sent the letter to encourage her to recommend her patients to have the flu shot. In the second class, those patients who would not get the flu shot regardless of their physician being encouraged or not, would also have the same chance of getting the flu.

Let us now assume the individual level monotonicity assumption holds for a particular trial. Due to the monotonicity assumption, we are able to ignore the stratum of defier, $(S_j^g(0), S_{ij}^p(0), S_j^g(1), S_{ij}^p(1))$ = (0, 1, 1, 0), and consider only the average causal effects within the stratum of complier $(S_j^g(0), S_{ij}^p(0), S_j^g(1), S_{ij}^p(1))$ = (0, 0, 1, 1). The "Complier Average Causal Effect" is naturally defined as

$$CACE = \mu_{1c} - \mu_{0c}.$$

where $\mu_{1c} = E(Y(Z=1) \mid C^g C^p = c^g c^p)$ and $\mu_{0c} = E(Y(Z=0) \mid C^g C^p = c^g c^p)$ are the population mean of complier under intervention and control condition, respectively. Note that the overall population means of potential outcome under intervention and control can be expressed as

$$E(Y(Z=1)) = \mu_1 = \pi_c \times \mu_{1c} + \pi_n \times \mu_{1n} + \pi_a \times \mu_{1a}$$

$$E(Y(Z=0)) = \mu_0 = \pi_c \times \mu_{0c} + \pi_n \times \mu_{0n} + \pi_a \times \mu_{0a} \tag{2.7}$$

respectively, and π_c, π_n, and π_a are the population proportion (probability) of complier, never-taker, or always-taker. It follows that

$$\begin{aligned}\mu_1 - \mu_0 &= (\pi_c \times \mu_{1c} + \pi_n \times \mu_{1n} + \pi_a \times \mu_{1a}) - (\pi_c \times \mu_{0c} + \pi_n \times \mu_{0n} + \pi_a \times \mu_{0a}) \\ &= \pi_c \times (\mu_{1c} - \mu_{0c}) + \pi_n \times (\mu_{1n} - \mu_{0n}) + \pi_a \times (\mu_{1a} - \mu_{0a}) \\ &= \pi_c \times (\mu_{1c} - \mu_{0c})\end{aligned}$$

$$CACE = \mu_{1c} - \mu_{0c} = \frac{\mu_1 - \mu_0}{\pi_c}. \tag{2.8}$$

In situations where $\pi_d > 0$, in contrast with the montonicity assumption, i.e. there are individuals whose compliance behavior would be opposite to intervention assignment, the overall population means of potential outcome would be given by

$$\mu_1 = \pi_c * \mu_{1c} + \pi_n * \mu_{1n} + \pi_a * \mu_{1a} + \pi_d * \mu_{1d}$$

$$\mu_0 = \pi_c * \mu_{0c} + \pi_n * \mu_{0n} + \pi_a * \mu_{0a} + \pi_d * \mu_{0d} \tag{2.7a}$$

It follows that

$$\begin{aligned}\mu_1 - \mu_0 &= \pi_c \times (\mu_{1c} - \mu_{0c}) + \pi_n^* \times (\mu_{1n} - \mu_{0n}) + \pi_a^* \times (\mu_{1a} - \mu_{0a}) + \pi_d \times (\mu_{1d} - \mu_{0d}) \\ &= \pi_c \times (\mu_{1c} - \mu_{0c}) + \pi_d \times (\mu_{1d} - \mu_{0d})\end{aligned}$$

and

$$CACE^{*} = \mu_{1c} - \mu_{0c} = \frac{\mu_1 - \mu_0 - \pi_d \times (\mu_{1d} - \mu_{0d})}{\pi_c} \tag{2.8a}$$

If intervention has causal effects among both complier and defier, μ_{1c} - μ_{0c} and $\mu_{1d} - \mu_{0d}$ are of opposite signs. As a result, CACE* > CACE, and CACE under the monotonicity assumption would be an underestimate of the true causal effects. In practice compliance behavior is not observable, so some practical assumptions are useful for estimating CACE*. For example, if we assume causal effects of the intervention with $\mu_{1d} - \mu_{0d} = -\lambda(\mu_{1c} - \mu_{0c})$, and $\lambda > 0$, the estimate for CACE* may be approximated by

$$CACE^{**} = \mu_{1c} - \mu_{0c} = \frac{\mu_1 \text{ - } \mu_0}{\pi_c - \lambda\pi_d}. \tag{2.8b}$$

In the case where intervention effects are opposite among compliers and defiers, the estimator in (2.8) will be an over-estimate of CACE*.

Table 6: Principal Stratification of Flu Shot Study

Principal Stratum by Level of Group and individual Participation $C^g C^p$	Potential Mediator Outcome			
	Control		Intervention	
	$S_j^g(0)$	$S_{ij}^p(0)$	$S_j^g(1)$	$S_{ij}^p(1)$
Group Complier Individual Never-takers ($c^g n^p$)	0	0	1	0
Group Complier Individual Compliers ($c^g c^p$)	0	0	1	1
Group Complier Individual Defiers ($c^g d^p$)	0	1	1	0
Group Complier Individual Always-takers ($c^g a^p$)	0	1	1	1

Note: $S_j^g(Z)$ = Group level Potential Mediator Outcome;
$S_{ij}^p(Z)$ = Individual level Potential Mediator Outcome

2.4.2. G-Bloom's Model

The GA gatekeeper study to be discussed in Chapter Three is an example of G-Bloom's model. In this study, randomization occurred at the school level, intervention was implemented at the school level by training of school staff to identify signs of suicide and to ask the youth whether they felt suicidal. Group level participation was based on whether or not training at the school occurred by a certain time, and later we examine other characteristics, such as the proportion of school staff that are trained as a function of time. Youth outcomes included whether they were referred to the school support staff to deal with suicide and other life threatening behaviors, whether they were suicidal, and whether they attended mental health treatment. These data were all collected at the student level in a deidentified fashion. Researchers also measured the percentage of staff who had been trained and the time of staff receiving training at each school in aggregate instead of which individual staff member was trained. It our first look at this problem, it is assumed that participation status of staff within school is the same as the participant status of their school. Also, no one in a school could receive the intervention if the school was a control (Bloom's model) and all youth were exposed to the effects of the training if the school participated in the intervention. It is also assumed that the participation status of a school, which is only observed for those randomly assigned to be trained, was independent of its assigned intervention condition. In the gatekeeper study, due to the requirement that control schools had a zero probability to receive intervention, participation status was only observable among the intervention schools but not control schools. Therefore, all schools were classified into two principal strata formed according to participation status at the group and individual levels, i.e., never taker and complier.

Table 7 illustrates the design of the intervention. Using the notation of Table 4, $S_j^g(1)$ represented the participation status of the j^{th} school if it had been assigned to the intervention condition. Participation was determined by the time at which the school would implement the gatekeeper program. Therefore the level of participation could be defined as the time the school started the gatekeeper program. A simpler approach is to dichotomize the timing of intervention into early ($S^g=1$) or later ($S^g=0$). This is the case in Table 7. If assigned to control school participation $S_j^g(0)=0$ because it had no chance to participate, hence information on the level of participation was missing. This is seen from Table 6 as there is only one row, corresponding to $(S_j^g(0), S_{ij}^p(0))$ = (0, 0). Once the school's participation status was determined, the staffs' participation status had to be the same as the school's. Therefore $S_j^g(1)=S_{ij}^p(1)$ is given in the only two columns of Table 7. The design resulted in only two strata, never-taker (later-adopter) ($n^g n^p$) or complier (early-adopter) ($c^g c^p$). For this example, the common assumptions warrant slightly different interpretation.

- *Monotonicity Assumption*. $\Pr(S_j^g(0)=1)=0$ and $\Pr(S_{ij}^p(0)=1)=0$. That is, no school in the control condition hence no staff in a control school can receive active intervention.

For now we consider the situation with no covariates at the group level to predict participation status, and therefore all groups have the same probability (π) of participating if they are assigned to the intervention condition. Consider W to be a measured outcome at the group level. Additionally, Average Causal Effect is defined within complier (early-adopter) stratum $(S_j^g(0), S_{ij}^p(0), S_j^g(1), S_{ij}^p(1))$ = (0, 0, 1, 1) as

$$ACE(\text{complier (early-adopter)}) = E(W_{c^g c^p}(Z=1)) - E(W_{c^g c^p}(Z=0))$$

Note that we have used the subscript $c^g c^p$ to indicate expectation within the stratum of complier (early-adopter) while omitting $S_{ij}^p(Z_j)$ at the same time. Because membership in the stratum of complier (early-adopter) among those assigned to control ($Z = 0$) is not completely observable, the following exclusion restriction is used to provide identifiability.

- *Exclusion restriction*. The Average Causal Effect among nonparticipants (never-taker(later-adopter): $(S_j^g(0), S_{ij}^p(0), S_j^g(1), S_{ij}^p(1))$ = (0, 0, 0, 0) is zero,

$$ACE\text{ (never-taker (later-adopter))} = E(W_{n^g n^p}(Z=1)) - E(W_{n^g n^p}(Z=0)) = 0$$

Note that

$$E(W(Z=0)) = E(W_{n^g n^p}(Z=0)) \times P_r(n) + E(W_{c^g c^p}(Z=0)) \times P_r(c)$$

Because a subject can only be either a complier (early-adopter) or a never-taker (later-adopter), randomized intervention assignment dictates that $P_r(c) = \pi$ and $P_r(n) = 1 - \pi$.

The preceding equation gives

$$E(W_{c^g c^p}(Z=0)) = \frac{E(W(Z=0)) - E(W_{n^g n^p}(Z=0)) \times (1-\pi)}{\pi}$$

$$= \frac{E(W(Z=0)) - E(W_{n^g n^p}(Z=1))(1-\pi)}{\pi}$$

As a result, the Average Causal Effect among the compliers (early-adopter) is given by

$$ACE = E(W_{c^g c^p}\ (Z=1)) - (\frac{E(W(Z=0))\text{-}E(W_{n^g n^p}(Z=1)) \times (1-\pi)}{\pi}) \quad (2.9)$$

Table 7: G-Bloom's Mode: GA Gatekeeper Study

Principal Stratum by Level of Group and Individual Participation $C^g C^p$	Potential Mediator Outcome			
	Control		Intervention	
	$S_j^g(0)$	$S_{ij}^p(0)$	$S_j^g(1)$	$S_{ij}^p(1)$
Group Never-takers Individual Never-takers ($n^g n^p$)	0	0	0	0
Group Complier Individual Compliers ($c^g c^p$)	0	0	1	1
Group Defier Individual Defiers ($d^g d^p$)	Not Possible	Not Possible	Not Possible	Not Possible
Group Always-takers Individual Always-takers ($a^g a^p$)	Not Possible	Not Possible	Not Possible	Not Possible

Note: $S_j^g\ (Z)$ = Group level Potential Mediator Outcome;

$S_{ij}^p\ (Z)$ = Individual level Potential Mediator Outcome

S = 0: later intervention

S = 1: early intervention

In equation (2.9) all expectation is based on observed data. For example, $E(W(Z=0))$ is based on those assigned to the control condition. Equation (2.9) allows for unbiased estimation of ACE based on moment estimators, assuming that $W(Z=0)$, $W_{n^g n^p}\ (Z\ =1)$, and $W_{c^g c^p}\ (Z\ =1)$ have the same variance. Note that the moment estimators

$$\overline{W}(Z=0)=\frac{1}{k}\sum_{j=1}^{k}W_j(0)\ ,$$

$$\overline{W}_{n^g n^p}\ (Z\ =1)=\frac{1}{k_1}\sum_{j\in n}W_j(1)\,,$$

and

$$\overline{W}_{c^g c^p}\ (Z\ =1)=\frac{1}{k_2}\sum_{j\in c}^{k_\pi}W_j(1)$$

are unbiased estimators of $E(W\,(Z\ =\ 0))$, $E(W_{n^g n^p}\ (Z\ =1))$, and $E(W_{c^g c^p}\ (Z\ =1))$. Of the latter two estimators, the summation is over the set of never-taker (later-adopter) and the subset of complier (early-adopter) within the control and intervention groups, respectively, and k_1 and k_2 are the number of groups in each set, respectively. Finally, an unbiased moment estimator of the causal effect within the stratum of complier (early-adopter) is

$$A\hat{C}E\ =\overline{W}_{c^g c^p}\ (Z\ =1)\text{-}\ (\frac{\overline{W}(\mathrm{Z}=0)\text{-}(1-\hat{\pi})\times\ \overline{W}_{n^g n^p}\ (Z\ =1)}{\hat{\pi}}) \qquad (2.10)$$

The standard error (SE) of $A\hat{C}E$ is approximately equals to:

$$\sqrt{\hat{\text{var}}(\overline{W}_{c^g c^p}\ (Z\ =1))\ +\frac{\hat{\text{var}}(\overline{W}(\text{Z}=0))\ +\ (1-\hat{\pi})^2\times\hat{\text{var}}(\overline{W}_{n^g n^p}\ (Z\ =1))}{\hat{\pi}^2}}$$

(2.11)

Had the groups under the control condition had a possibility to receive the intervention, then there would be four principal strata and AIR model would apply. Results presented in section 2.4.1 can be readily extended to group level analysis by using the group average instead of the individual average.

2.4.3. GB-PB Model/GA-PA Model

With GB-PB/GA-PA models, we consider the situation where participation status is observed at both group level and individual level. An example is the Family-School Partnership intervention for first graders in Baltimore, MD (Ialongo et al., 1999). In that study, children were randomly allocated to one of three classrooms in first grade; the three classrooms were randomized to a classroom-centered intervention, a parent training intervention, or a control, respectively. The parent training intervention and control conditions are a pair of conditions considered in an example for these two models. In the parent training intervention, a counselor in the school was trained to provide parent training in the child's behavior management and a home environment to support school achievement. These parent trainings were provided weekly in school. However, the class receiving the control condition cannot participate in the intervention, at least by assumption that there is no contamination across classes. Thus it is possible to examine participation at two levels, both at the class level represented by the school counselor, who may or may not deliver the program with full fidelity, and at the level of the family, who may or may not attend an adequate number of sessions to receive the benefit of this intervention. One of the critical issues with two-level participation is the interrelationship between the two levels. We consider and examine two situations below: individuals cannot participate when their group does not participate, or not restricted at both individual level participation and group level participation.

Under the first situation where the school counselor can deliver the intervention with fidelity or not, and parents of first graders cannot participate in the intervention if their class is assigned the control condition, subjects are classified into 4 principal strata, $n^g n^p$, $n^g c^p$, $c^g n^p$, or $c^g c^p$ (Table 8a), The resultant principal strata are assured by the follow modified monotonicity assumption.

- *Monotonicity Assumption* (Angrist et al., 1996). $S_j^g(0) = S_{ij}^p(0) = 0$. In other words, Pr ($S_j^g$ (0) = 1) = 0 and Pr($S_{ij}^p(0) = 1$) = 0.

When an individual's participation status is always the same as the associated group's participation status (2 principal strata), G-B model is appropriate. When this is not the case, we consider the GB-PB model in which we assume there is no group level causal effect among nonparticipant groups (Angrist et al., 1996).

- *Exclusion Restriction*. The Average Causal Effect among nonparticipants is zero. That is,

$ACE(S_j^g = 0) = E(W(Z = 1, S_j^g = 0)) - E(W(Z = 0, S_j^g = 0)) = 0$. This in turn can be expressed in two parts:

$$ACE(n^g n^p) = ACE(S_{ij}^p(0) = 0, S_j^g(0) = 0, S_{ij}^p(1) = 0, S_j^g(1) = 0) =$$
$$E(W(Z = 1, S_j^g = 0, S_{ij}^p = 0)) - E(W(Z = 0, S_j^g = 0, S_{ij}^p = 0)) = 0$$

and

$$ACE(n^g c^p) = ACE(S_{ij}^p(0) = 0, S_j^g(0) = 0, S_{ij}^p(1) = 0, S_j^g(1) = 1) =$$
$$E(W(Z = 1, S_j^g = 0, S_{ij}^p = 1)) - E(W(Z = 0, S_j^g = 0, S_{ij}^p = 0)) = 0.$$

We note that the exclusion restriction $ACE(n^g c^p) = 0$ may be replaced by adding $S_j^g(1) \geq S_{ij}^p(1)$, or equivalently $\Pr(S_j^g(1) = 0 \,\&\, S_{ij}^p(1) = 1) = 0$, to the preceding monotonicity assumption. We also note an additional possible condition for exclusion restriction:

$$ACE(c^g n^p) = ACE(S_j^g(0) = 0, S_{ij}^p(0) = 0, S_j^g(1) = 1, S_{ij}^p(1) = 0) =$$
$$E(W(Z = 1, S_j^g = 1, S_{ij}^p = 0)) - E(W(Z = 0, S_j^g = 0, S_{ij}^p = 0)) = 0$$

The preceding equation implies that there are no causal effects among those never- taker individuals. The use of the additional restriction would impact the estimation of causal effects as is seen below in our discussion.

The average causal effects in the compiler stratum (both group and individual complier $c^g c^p$) is given by:

$$ACE(S_j^g(0) = 0, S_{ij}^p(0) = 0, S_j^g(1) = 1, S_{ij}^p(1) = 1)$$

$$= E(W_{c^g c^p}(Z = 1, S_j^g = 1, S_{ij}^p = 1)) - E(W_{c^g c^p}(Z = 0, S_j^g = 0, S_{ij}^p = 0)). \qquad (2.12)$$

Let π_1 be the group level participation rate under the intervention condition, and π_2 be the individual level participation rate within groups assigned to the intervention condition. Because of randomization, the proportion of participation at group level is the same as that of complier, i.e., $P_r(c^g) = \pi_1$, and subsequently, $P_r(n^g) = 1 - \pi_1$. By the same argument, the proportion of participation among individuals is that of participation when assigned to an intervention class, $P_r(c^p) = \pi_2$ and $P_r(n^p) = 1 - \pi_2$.

Note that

$$\begin{aligned} E(W(Z=1)) - E(W(Z=0)) = \; & \pi_1\pi_2 E(W_{c^g c^p}(Z=1) - W_{c^g c^p}(Z=0)) \; + \\ & \pi_1(1-\pi_2)E(W_{c^g n^p}(Z=1) - W_{c^g n^p}(Z=0)) + \\ & (1-\pi_1)\pi_2 E(W_{n^g c^p}(Z=1) - W_{n^g c^p}(Z=0)) + \\ & (1-\pi_1)(1-\pi_2)E((W_{n^g n^p}(Z=1) - W_{n^g n^p}(Z=0)) \end{aligned}$$

Under the exclusion restriction,

$$\begin{aligned} E(W(Z=1)) - E(W(Z=0)) = \; & \pi_1\pi_2 E(W_{c^g c^p}(Z=1) - W_{c^g c^p}(Z=0)) \; + \\ & \pi_1(1-\pi_2)E(W_{c^g n^p}(Z=1) - W_{c^g n^p}(Z=0)) \end{aligned}$$

It follows

$$CACE(c^g c^p) = \frac{E(W(Z=1) - W(Z=0)) - \pi_1(1-\pi_2)E\{W(Z=1, S_j{}^g = 1, S_{ij}{}^p = 0) - W(Z=0, S_j{}^g = 0, S_{ij}{}^p = 0)\}}{\pi_1\pi_2}.$$

Under the additional exclusion restriction $ACE(c^g n^p) = 0$, it reduces to

$$CACE(c^g c^p) = \frac{E(W(Z=1) - W(Z=0))}{\pi_1\pi_2} \tag{2.13}$$

Because the group $(Z=0, S_j^g=0, S_{ij}^p=0)$ is not differentiable from the other strata, $E(W_{c^g c^p}(Z=0, S_j^g=0, S_{ij}^p=0))$ is not directly computable. We thus use the exclusion restriction to re-express $E(W_{c^g c^p}(Z=0, S_j^g=0, S_{ij}^p=0))$ as follows.

$$
\begin{aligned}
&E(W(Z=0, S_j^g=0, S_{ij}^p=0)) \\
&\quad = E(W_{n^g}(Z=0, S_j{}^g=0)) \times P(n^g) + E(W_{c^g}(Z=0, S_j^g=0)) \times P(c^g)
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
&E\,(W_{n^g}(Z=0, S_j^g=0)) \\
&\quad = EW_{n^g n^p}(Z=0, S_j^g=0, S_{ij}^p=0) \times P(n^p) + EW_{n^g c^p}(Z=0, S_j^g=0, S_{ij}^p=0) \times P(c^p) \\
&\quad = EW_{n^g n^p}(Z=0, S_j^g=0, S_{ij}^p=0) \times (1-\pi_2) + EW_{n^g c^p}(Z=0, S_j^g=0, S_{ij}^p=0) \times \pi_2
\end{aligned}
$$

$$
\begin{aligned}
&E(W_{c^g}(Z=0, S_j^g=0)) \\
&\quad = EW_{c^g n^p}(Z=0, S_j^g=0, S_{ij}^p=0) \times P(n^p) + EW_{c^g c^p}(Z=0, S_j^g=0, S_{ij}^p=0) \times P(c^p) \\
&\quad = EW_{c^g n^p}(Z=0, S_j^g=0, S_{ij}^p=0) \times (1-\pi_2) + EW_{c^g c^p}(Z=0, S_j^g=0, S_{ij}^p=0) \times \pi_2
\end{aligned}
$$

Using three equations from the exclusion restriction, we can show

$$
\begin{aligned}
&E(W_{c^g c^p}(Z=0, S_j^g=0, S_{ij}^p=0)) \\
&= \frac{E(W(Z=0, S_j^g=0, S_{ij}^p=0)) - E(W_{n^g}(Z=1, S_j^g=0))(1-\pi_1)}{\pi_1 \pi_2} \\
&\quad - \frac{E(W_{c^g n^p}(Z=1, S_j^g=1, S_{ij}^p=0))(1-\pi_2)\pi_1}{\pi_1 \pi_2}
\end{aligned}
$$

in which all terms can be estimated because data are observable within the respective strata. For example, $E(W(Z=0))$ and $E(W(Z=1))$ is based on those assigned to the control and intervention conditions, respectively. Equation (2.13) allows for unbiased estimation of CACE based on moment estimators, assuming that $W(Z=0)$, and $W(Z=1)$ have the same variance. Note that the moment estimators

$$\overline{W}(Z=0) = \frac{1}{k}\sum_{j=1}^{k} W_j(0) \ ,$$

$$\overline{W}(Z=1) = \frac{1}{k}\sum_{j=1}^{k} W_j(1),$$

are unbiased estimators of $E(W(Z = 0))$ and $E(W(Z = 1))$. Finally, an unbiased moment estimator of the complier average causal effect is:

$$CA\hat{C}E = \frac{\overline{W}(\mathrm{Z}=1) - \overline{W}(\mathrm{Z}=0)}{\hat{\pi}_1 \hat{\pi}_2} \tag{2.14}$$

The standard error (SE) of $CA\hat{C}E$ is approximately $\sqrt{\frac{\hat{\mathrm{var}}(\overline{W}(\mathrm{Z}=1)) + \hat{\mathrm{var}}(\overline{W}(\mathrm{Z}=0))}{\hat{\pi}_1{}^2 \hat{\pi}_2{}^2}}$

Table 8a: Baltimore Good Behavior Game: GB-PB Model for Intervention of First Graders in Baltimore (with restriction on individual level)

Principal Stratum by Level of Group and Individual Participation $C^g C^p$	Potential Mediator Outcome			
	Control		Intervention	
	$S_j^g(0)$	$S_{ij}^p(0)$	$S_j^g(1)$	$S_{ij}^p(1)$
Group Never-takers Individual Never-takers ($n^g n^p$)	0	0	0	0
Group Never-takers Individual Complier ($n^g c^p$)	0	0	0	1
Group Complier Individual Never-takers ($c^g n^p$)	0	0	1	0
Group Complier Individual Compliers ($c^g c^p$)	0	0	1	1

However, if more complicated examples under which both group level and individual level can have the alternative to receive the intervention, then all subjects will be classified into 16 principal strata based on assigned intervention status, group level participation status and individual level participation status (GA-PA Model, Table 8b). Analyses will be done at either the group level or individual level. For example, if we assume that an individuals' participation status is the same as their group participation status the causal models can be analyzed by using the same method as the AIR model or G-AIR model. The number of principal strata will be reduced from 16 to 4. Otherwise, more assumptions are required for analysis in order to reduce number of status beside monotonicity assumption and exclusion restriction assumption. The resultant principal strata are assured by the follow modified monotonicity assumption.

- *Monotonicity Assumption* (Angrist et al., 1996). The monotonicity assumption excludes the probability of having defiers from both group level and individual level, which assume $S_{ij}^{p}(S_{j}^{g}(1)=1) \geq S_{ij}^{p}(S_{j}^{g}(0)=0)$ and $S_{j}^{g}(1) \geq S_{j}^{g}(0)$. In our example, it simply eliminates the possibility that there are parents who would not want attend to training sessions if the school counselor would deliver the program with full fidelity, but otherwise would want attend to training sessions. In terms of probability, the assumption dictates

  $P_r(S_{ij}^{p}(S_{j}^{g}(0)=0)=1 \,\&\, S_{ij}^{p}(S_{j}^{g}(1)=1)=0)=0$ or

  $P_r(S_{j}^{g}(1)=0 \,\&\, S_{j}^{g}(0)=1)=0$, hence effectively removing the stratum including any individual defier or group defier. The number of principal strata will be reduced from 16 to 9 (Table 8b).

Table 8b: Baltimore Good Behavior Game: GB-PB Model for Intervention of First Graders in Baltimore (without restriction)

Principal Stratum by Level of Group and Individual Participation $C^g C^p$	Potential Mediator Outcome			
	Control		Intervention	
	$S_j^g(0)$	$S_{ij}^p(0)$	$S_j^g(1)$	$S_{ij}^p(1)$
Group Never-takers Individual Never-takers ($n^g n^p$)	0	0	0	0
Group Never-takers Individual Complier ($n^g c^p$)	0	0	0	1
Group Complier Individual Never-takers ($c^g n^p$)	0	0	1	0
Group Complier Individual Compliers ($c^g c^p$)	0	0	1	1
Group Never-takers Individual Always-takers ($n^g a^p$)	0	1	0	1
Group Complier Individual Always-takers ($c^g a^p$)	0	1	1	1
Group Always-takers Individual Never-takers ($a^g n^p$)	1	0	1	0
Group Always-takers Individual Complier ($a^g c^p$)	1	0	1	1
Group Always-takers Individual Always-takers ($a^g a^p$)	1	1	1	1

2.5 Incorporating Individual Characteristics through Regression Models

In general, intervention effects could be impacted by a subject's characteristics as well as other factors. As a result, individual's response varies. To reflect the between-individual variation in response, we can incorporate individual level covariates in the stratum mean of the Bloom model, as Little & Yau (1998) did. This can be done through adopting the linear model for the stratum-specific mean:

$$Y_i = \alpha_{c^p} c^p + \alpha_{n^p} n^p + \beta_{Zc^p} Z_i c^p + \beta'_{Xc^p} X_i c^p + \beta'_{Xn^p} X_i n^p + \varepsilon_i c^p + \varepsilon_i n^p \qquad (2.14)$$

where Y_i is a continuous response variable of individual i, ($c^p = 1$ & $n^p = 0$) if individual i is a complier (participation), and ($c^p = 0$ & $n^p = 1$) if individual i is a never-taker (non-participation). Further, α_{c^p} is intercept for compliers and α_{n^p} is intercept for never-taker; Z_i =1 indicates intervention condition and Z_i =0 is control condition; β_Z is the coefficient of intervention effects, representing average effect of intervention; X_i is a $p \times 1$ vector of individual level characteristics , and β'_X is $p \times 1$ vector of associated coefficients. The complier-average causal effect (ACE) is equal to β_{Zc^p} as given in equation (2.14). The indicators c^p and n^p are for the individual's principal strata membership. Recall that this class membership is independent of the intervention assignment and is only observed on those who are assigned to the intervention condition. However, the membership status for those under the control condition can be predicted under an assumption that Z is independent of the class membership and the participation rate is the same under both intervention and control conditions, the case for a well-designed randomized trial. Furthermore, it is possible to allow for participation (compliance) rate to vary with individual characteristics. Take a single level randomized trial for example. The log odds of participation (among those assigned to intervention) can be expressed as a linear function of the individual level covariates via a logistic regression model as suggested by Little and Yau (1998):

$$\log\left(\frac{\Pr(C_i^p = c^p)}{1-\Pr(C_i^p = c^p)}\right) = \gamma_0 + \gamma_1' X_{ic^p} \tag{2.15}$$

where X_{ic^p} is individual level covariates predictive of participation (compliance).

Assume that the distribution of Y is normally distributed with variance σ^2 and mean μ_{n^p} for never-taker (non-participation), $\mu_{c^p 0}$ for complier (participation) assigned to the control condition, and $\mu_{c^p 1}$ for complier (participation) assigned to the intervention condition. The likelihood based on the observed data then has the form:

$$\begin{aligned} L(\theta \mid \text{data}) \propto & \prod_{i\in\{Z_i=1,S_i=0\}} \pi_{n^p} g(y_i \mid \mu_{n^p}, \sigma^2) \times \prod_{i\in\{Z_i=1,S_i=1\}} \pi_{c^p} g(y_i \mid \mu_{c^p 1}, \sigma^2) \times \\ & \prod_{i\in\{Z_i=0,S_i=0\}} [\pi_{n^p} g(y_i \mid \mu_{n^p}, \sigma^2) + \pi_{c^p} g(y_i \mid \mu_{c^p 0}, \sigma^2)] \end{aligned} \tag{2.16}$$

where $\pi_{n^p} + \pi_{c^p} = 1$ and $\theta = (\pi_{n^p}, \pi_{c^p}, \mu_{n^p}, \mu_{c^p 1}, \mu_{c^p 0}, \sigma^2)$ is set of parameters in the model, and $g(y \mid \mu, \sigma^2)$ denotes the probability density of a normal distribution with mean μ and variance σ^2. Maximum likelihood estimates of the parameters are obtained by maximizing this mixture of likelihood functions of two principal strata of c^p (compliers) and n^p (never-takers). In particular, the estimation of π_{c^p} is via the logistic regression model of (2.15). The complier-average causal effect (ACE) is then estimated by $\hat{\mu}_{c^p 1} - \hat{\mu}_{c^p 0} = \hat{\beta}_{Zc^p}$, where $\hat{\beta}_{Zc^p}$ is the maximum likelihood estimates of β_{Zc^p}.

However, equation 2.14 requires the exclusion restriction assumption under which the causal effect is zero for never-taker. This may not hold in certain cases. If the never-taker-average causal effect does exist, then equation 2.14 can be re-written in a full model:

$$Y_i = \alpha_{c^p} c^p + \alpha_{n^p} n^p + \beta_{Zc^p} Z_i c^p + \beta_{Zn^p} Z_i n^p + \beta'_{Xc^p} X_i c^p + \beta'_{Xn^p} X_i n^p + \varepsilon_i c^p + \varepsilon_i n^p \tag{2.17}$$

where β_{Zc^p} is complier-average causal effect and β_{Zn^p} is never-taker-average causal effect. In reality, β_{Xc^p} and β_{Xn^p} may vary with individual characteristics.

The causal effect can be estimated by using marginal maximum likelihood estimates form the models defined in equations 2.14 and 2.15. However, these mixture likelihood functions involve individuals' class membership, which is unobservable for subjects in the control condition. Thus the estimation principal stratum membership is a key issue. To this end, the mixture model ties together the unobserved class membership in the control condition with individual level covariates and the marginal effects of intervention, and then maximizes the marginal maximum likelihood to estimate the regression coefficients. Although this mixture likelihood approach has been used in analysis of single level randomized studies, its use in multilevel randomized trials is relatively new and is considered in this dissertation. The following sections discuss the mixture likelihood approach in analysis of two-level randomized trials.

2.6 Mixtures and Marginal Maximum Likelihood Approach for Two-Level Randomized Trials

In this dissertation we have defined a two-level randomized trial as one where random assignment of intervention condition occurs at a group instead of individual level. Further, intervention status may change over time during the study period. Principal strata are defined by the combination of assigned intervention condition and the compliance of the subjects at both group and individual levels. The full likelihood function is the mixture of those associated with all possible principal strata. The mixture then leads to a marginal likelihood function on which maximum likelihood estimation is feasible. In the following we illustrate the regression model and its corresponding mixture likelihood function for the AIR models.

2.6.1 Two-Level Randomized Trials with Active Intervention versus Control

We assume that in the AIR model Y_{ij}, has same distribution in all principal strata $(C_j^g C_{ij}^p = c_j^g c_{ij}^p, c_j^g n_{ij}^p, c_j^g a_{ij}^p, c_j^g d_{ij}^p)$. In principle, individual membership in a principal stratum may not be observed, but can be predicted under a multinomial distribution in conjunction with individual-level covariates.

In this case, equation 2.14 can be extended as

$$
\begin{aligned}
Y_{ij} = {} & \beta'_{c_j^g n_{ij}^p}(c_j^g n_{ij}^p) + \beta'_{c_j^g c_{ij}^p}(c_j^g c_{ij}^p) + \beta'_{c_j^g a_{ij}^p}(c_j^g a_{ij}^p) + \beta'_{c_j^g d_{ij}^p}(c_j^g d_{ij}^p) + \\
& \beta'_{(c_j^g n_{ij}^p)x} x_{c_j^g n_{ij}^p}(c_j^g n_{ij}^p) + \beta'_{(c_j^g c_{ij}^p)x} x_{c_j^g c_{ij}^p}(c_j^g c_{ij}^p) + \beta'_{(c_j^g a_{ij}^p)x} x_{c_j^g a_{ij}^p}(c_j^g a_{ij}^p) + \beta'_{(c_j^g d_{ij}^p)x} x_{c_j^g d_{ij}^p}(c_j^g d_{ij}^p) + \\
& \beta'_{(c_j^g n_{ij}^p)Z} Z_j (c_j^g n_{ij}^p) + \beta'_{(c_j^g c_{ij}^p)Z} Z_j (c_j^g c_{ij}^p) + \beta'_{(c_j^g a_{ij}^p)Z} Z_j (c_j^g a_{ij}^p) + \beta'_{(c_j^g d_{ij}^p)Z} Z_j (c_j^g d_{ij}^p) +
\end{aligned}
$$

$$\varepsilon_{(c_j^g n_{ij}^p)b_j}(c_j^g n_{ij}^p) + \varepsilon_{(c_j^g c_{ij}^p)b_j}(c_j^g c_{ij}^p) + \varepsilon_{(c_j^g a_{ij}^p)b_j}(c_j^g a_{ij}^p) + \varepsilon_{(c_j^g d_{ij}^p)b_j}(c_j^g d_{ij}^p) +$$
$$\varepsilon_{(c_j^g n_{ij}^p)wi_j}(c_j^g n_{ij}^p) + \varepsilon_{(c_j^g c_{ij}^p)w_{ij}}(c_j^g c_{ij}^p) + \varepsilon_{(c_j^g a_{ij}^p)w_{ij}}(c_j^g a_{ij}^p) + \varepsilon_{(c_j^g d_{ij}^p)w_{ij}}(c_j^g d_{ij}^p) \qquad (2.18)$$

$$\text{where} \quad \begin{array}{c} \{c_j^g n_{ij}^p = 1 \,\&\, c_j^g c_{ij}^p = 0 \,\&\, c_j^g a_{ij}^p = 0 \,\&\, c_j^g d_{ij}^p = 0\} = \text{never-taker} \\ \{c_j^g n_{ij}^p = 0 \,\&\, c_j^g c_{ij}^p = 1 \,\&\, c_j^g a_{ij}^p = 0 \,\&\, c_j^g d_{ij}^p = 0\} = \text{complier} \\ \{c_j^g n_{ij}^p = 0 \,\&\, c_j^g c_{ij}^p = 0 \,\&\, c_j^g a_{ij}^p = 1 \,\&\, c_j^g d_{ij}^p = 0\} = \text{always-taker} \\ \{c_j^g n_{ij}^p = 0 \,\&\, c_j^g c_{ij}^p = 0 \,\&\, c_j^g a_{ij}^p = 0 \,\&\, c_j^g d_{ij}^p = 1\} = \text{defier} \end{array}$$

$\beta'_{c_j^g n_{ij}^p}$, $\beta'_{c_j^g c_{ij}^p}$, $\beta'_{c_j^g a_{ij}^p}$ and $\beta'_{c_j^g d_{ij}^p}$ are intercepts for never-taker, complier, always-taker and defier, respectively; $x_{c_j^g n_{ij}^p}$, $x_{c_j^g c_{ij}^p}$, $x_{c_j^g a_{ij}^p}$ and $x_{c_j^g d_{ij}^p}$ are individual-level characters for never-taker, complier, always-taker and defier; β'_xs are coefficients of x covariates; β'_Zs represent the intervention effect. Errors $\varepsilon_{(c_j^g n_{ij}^p)b_j}$ (never-taker), $\varepsilon_{(c_j^g c_{ij}^p)b_j}$ (complier) , $\varepsilon_{(c_j^g a_{ij}^p)b_j}$ (always-taker) and $\varepsilon_{(c_j^g d_{ij}^p)b_j}$ (defier) are assumed to be normally distributed with zero mean and the between groups variance $\sigma^2_{(c_j^g n_{ij}^p)b_j}$, $\sigma^2_{(c_j^g c_{ij}^p)b_j}$, $\sigma^2_{(c_j^g a_{ij}^p)b_j}$ and $\sigma^2_{(c_j^g d_{ij}^p)b_j}$. Similarly, errors $\varepsilon_{(c_j^g n_{ij}^p)w_{ij}}$, $\varepsilon_{(c_j^g c_{ij}^p)w_{ij}}$, $\varepsilon_{(c_j^g a_{ij}^p)w_{ij}}$ and $\varepsilon_{(c_j^g d_{ij}^p)w_{ij}}$ are assumed to be normally distributed with zero mean and the within groups variance $\sigma^2_{(c_j^g n_{ij}^p)w_{ij}}$, $\sigma^2_{(c_j^g c_{ij}^p)w_{ij}}$, $\sigma^2_{(c_j^g a_{ij}^p)w_{ij}}$ and $\sigma^2_{(c_j^g d_{ij}^p)w_{ij}}$. The principal strata membership can be predicted by using a multinomial logistic model:

$$\Pr(C_j^g C_{ij}^p = t_j^g t_{ij}^p) = \frac{\exp(\gamma_0 + \gamma_1' X_{it^g t^p})}{\sum_{t^g t^p \in \{c^g c^p, c^g a^p, c^g n^p, c^g d^p\}} \exp(\gamma_0 + \gamma_1' X_{it^g t^p})} \qquad (2.19)$$

Where $t^g t^p \in \{c^g c^p, c^g a^p, c^g n^p, c^g d^p\}$ and $X_{it^g t^p}$ is individual level covariates predictive of principal strata membership.

Similarly, assume that the distribution of Y is normal distributed with variance σ^2 and mean $\mu_{c^g n^p}$ for never-taker, $\mu_{c^g a^p}$ for always-taker, $\mu_{c^g c^p 0}$ for complier assigned to the control condition, and $\mu_{c^g c^p 1}$ for complier assigned to the intervention condition, $\mu_{c^g d^p 0}$ for defier assigned to the control condition, and $\mu_{c^g d^p 1}$ for defier assigned to the intervention condition. The likelihood based on the observed data then has the form for AIR model:

$$
\begin{aligned}
L(\theta \mid \text{data}) \propto & \prod_{i \in \{Z_i=1, S_i=0)} [\pi_{c^g n^p} g(y_{ij} \mid \mu_{c^g n^p}, \sigma^2) + \pi_{c^g d^p} g(y_{ij} \mid \mu_{c^g d^p 1}, \sigma^2)] \\
& \times \prod_{i \in \{Z_i=1, S_i=1)} [\pi_{c^g c^p} g(y_{ij} \mid \mu_{c^g c^p 1}, \sigma^2) + \pi_{c^g a^p} g(y_{ij} \mid \mu_{c^g a^p}, \sigma^2)] \\
& \times \prod_{i \in \{Z_i=0, S_i=0)} [\pi_{c^g n^p} g(y_{ij} \mid \mu_{c^g n^p}, \sigma^2) + \pi_{c^g c^p} g(y_{ij} \mid \mu_{c^g c^p 0}, \sigma^2)] \\
& \times \prod_{i \in \{Z_i=0, S_i=1)} [\pi_{c^g n^p} g(y_{ij} \mid \mu_{c^g n^p}, \sigma^2) + \pi_{c^g d^p} g(y_{ij} \mid \mu_{c^g d^p 0}, \sigma^2)]
\end{aligned}
\tag{2.20}
$$

where $\pi_{c^g n^p} + \pi_{c^g c^p} + \pi_{c^g a^p} + \pi_{c^g d^p} = 1$ and

$\theta = (\pi_{c^g n^p}, \pi_{c^g c^p}, \pi_{c^g a^p}, \pi_{c^g d^p}, \mu_{c^g n^p}, \mu_{c^g c^p 1}, \mu_{c^g c^p 0}, \mu_{c^g a^p}, \mu_{c^g d^p 1}, \mu_{c^g d^p 0}, \sigma^2)$ is a set of parameters in the model, and $g(y \mid \mu, \sigma^2)$ denotes the probability density of a normal distribution with mean μ and variance σ^2. Maximum likelihood estimates of the parameters are obtained by maximizing this function with respect to the parameters θ. The complier-average causal effect ($CACE$) is then estimated by $\hat{\mu}_{c^g c^p 1} - \hat{\mu}_{c^g c^p 0}$, where $\hat{\beta}_{(c_j^g c_{ij}^p)}$ are maximum likelihood estimates of $\beta_{(c_j^g c_{ij}^p)}$.

Under the assumptions of monotonicity and exclusion restriction, defier does not exist and a causal effect will not exist for the never-taker and always-taker class. Then equation 2.18 becomes

$$\begin{aligned} Y_{ij} \ = \ & \beta^{'}_{c_j^g n_{ij}^p}(c_j^g n_{ij}^p) + \beta^{'}_{c_j^g c_{ij}^p}(c_j^g c_{ij}^p) + \beta^{'}_{c_j^g a_{ij}^p}(c_j^g a_{ij}^p) + \\ & \beta^{'}_{(c_j^g n_{ij}^p)x} x_{c_j^g n_{ij}^p}(c_j^g n_{ij}^p) + \beta^{'}_{(c_j^g c_{ij}^p)x} x_{c_j^g c_{ij}^p}(c_j^g c_{ij}^p) + \beta^{'}_{(c_j^g a_{ij}^p)x} x_{c_j^g a_{ij}^p}(c_j^g a_{ij}^p) + \\ & \beta^{'}_{(c_j^g c_{ij}^p)Z} Z_j (c_j^g c_{ij}^p) + \\ & \varepsilon_{(c_j^g n_{ij}^p)b_j}(c_j^g n_{ij}^p) + \varepsilon_{(c_j^g c_{ij}^p)b_j}(c_j^g c_{ij}^p) + \varepsilon_{(c_j^g a_{ij}^p)b_j}(c_j^g a_{ij}^p) + \\ & \varepsilon_{(c_j^g n_{ij}^p)wi_j}(c_j^g n_{ij}^p) + \varepsilon_{(c_j^g c_{ij}^p)w_{ij}}(c_j^g c_{ij}^p) + \varepsilon_{(c_j^g a_{ij}^p)w_{ij}}(c_j^g a_{ij}^p) \end{aligned} \quad (2.21)$$

The $CACE$ as defined in equation 2.8 is equal to

$$CACE = \mu_{1c} - \mu_{0c} = \frac{\mu_1 \text{ - } \mu_0}{\pi_c} = \beta_{(c_j^g c_{ij}^p)}$$

Given the linear mixed regression models and its corresponding likelihood function, the AIR model above can be fitted using the Mplus software. The sandwich type estimators are used in Mplus to adjust for any correlation among responses of different individuals characterized by selected characters.

In the case of G-Bloom Model, equation 2.18 is simplified to

$$\begin{aligned} Y_{ij} \ = \ & \beta^{'}_{n_j^g n_{ij}^p}(n_j^g n_{ij}^p) + \beta^{'}_{c_j^g c_{ij}^p}(c_j^g c_{ij}^p) + \beta^{'}_{(n_j^g n_{ij}^p)x} x_{n_j^g n_{ij}^p}(n_j^g n_{ij}^p) + \beta^{'}_{(c_j^g c_{ij}^p)x} x_{c_j^g c_{ij}^p}(c_j^g c_{ij}^p) + \\ & \beta^{'}_{(n_j^g n_{ij}^p)Z} Z_j (n_j^g n_{ij}^p) + \beta^{'}_{(c_j^g c_{ij}^p)Z} Z_j (c_j^g c_{ij}^p) + b_j + \\ & \varepsilon_{(n_j^g n_{ij}^p)b_j}(n_j^g n_{ij}^p) + \varepsilon_{(c_j^g c_{ij}^p)b_j}(c_j^g c_{ij}^p) + \varepsilon_{(n_j^g n_{ij}^p)wi_j}(n_j^g n_{ij}^p) + \varepsilon_{(c_j^g c_{ij}^p)w_{ij}}(c_j^g c_{ij}^p) \end{aligned} \quad (2.22)$$

due to group variation may exist, we define $b_j \mid C_j^g C_i^p \sim N(0,\ \sigma_b^2)$ is the random effect of the group. The principal strata membership can be predicted by using a multinomial logit model with a group variation:

$$\Pr(C_j^g C_i^p = c_j^g c_i^p) = \frac{\exp(\gamma_0 + \gamma_{1b}^{'} X_{b_j c^g c^p} + \gamma_{1w}^{'} X_{w_{ij} c^g c^p} + \xi_j)}{1 + \exp(\gamma_0 + \gamma_{1b}^{'} X_{b_j c^g c^p} + \gamma_{1w}^{'} X_{w_{ij} c^g c^p} + \xi_j)} \tag{2.23}$$

where X_{b_j} is a vector of between-group covariates; $X_{w_{ij}}$ is a vector of within-group covariates and between-group residual ξ_j predictive of principal strata membership. ξ_j causes the logistic value vary across groups which meaning the proportion of compliers differs across groups. The likelihood based on the observed data then has the form:

$$\begin{aligned} L(\theta \mid \text{data}) \propto & \prod_{j=1}^{k1} \Big(\prod_{i \in \{Z_j=1, S_j=0)} \pi_{n^g n^p} g(y_{ij} \mid \mu_{n^g n^p}, \sigma^2) \Big) \\ & \times \prod_{j=1}^{k2} \Big(\prod_{i \in \{Z_j=1, S_j=1)} \pi_{c^g c^p p} g(y_{ij} \mid \mu_{c^g c^p 1}, \sigma^2) \Big) \\ & \times \prod_{\text{j}} \Big(\prod_{i \in \{Z_j=0, S_j=0)} [\pi_{n^g n^p} g(y_{ij} \mid \mu_{n^g n^p}, \sigma^2) + \pi_{c^p} g(y_{ij} \mid \mu_{c^g c^p 0}, \sigma^2)] \Big) \end{aligned} \tag{2.24}$$

where k_1 and k_2 are the number of groups in each principal stratum in intervention condition and $\pi_{n^g n^p} + \pi_{c^g c^p} = 1$.

2.6.2 Two-Level Randomized Trial with Random Time of Crossover from Control to Active Intervention

We now consider a special case of two-level intervention study in which groups are originally assigned to the control condition and at a later time change to intervention condition. We call this a dynamic wait-listed design or timed assignment of intervention, and its properties of ITT analyses have been investigated elsewhere (Brown et al., 2006). We introduce this study design in preparation for the application to be discussed in the next chapter.

Let

$$Z_{jt} = \begin{cases} =1 & \text{Intervention} \\ =0 & \text{Control} \end{cases}$$

be the indicator of intervention assignment for $j = 1, \ldots, k$, the number of groups in the trial at time t.

- *Randomized Intervention Assignment*. We assume that the Z_{jt} is exchangeable, that is, all groups have the same probability of assignment to intervention at each different time interval, and that Z is drawn independent of all the other random variables in the study.
- *Monotonicity Assumption*. Pr ($S_{jt}^{g}(0) = 1) = 0$. That is, no one in the control group at time t can receive the active intervention before that group receives the intervention.
- *Stable Unit Treatment Value Assumption*. (Cox 1956, Rubin 1978). All potential outcomes for each individual are unchanged regardless of the assignment of all other units. However, the principal stratum to which each group belongs will not change over the time.

Under these assumptions, a regression model can be developed by including in equation 2.22 the effects of timed entrance into active intervention due to timed-assignment:

$$Y_{ijt} = \beta'_{n_j^g n_{ij}^p}(n_j^g n_{ij}^p) + \beta'_{c_j^g c_{ij}^p}(c_j^g c_{ij}^p) + \beta'_{(n_j^g n_{ij}^p)x} x_{c_j^g n_{ij}^p}(n_j^g n_{ij}^p) + \beta'_{(c_j^g c_{ij}^p)x} x_{c_j^g c_{ij}^p}(c_j^g c_{ij}^p) +$$
$$\beta'_{(c_j^g c_{ij}^p)Z} Z_j (c_j^g c_{ij}^p) + b_j + \beta'_{(c_j^g c_{ij}^p)T} T_j (c_j^g c_{ij}^p) +$$

$$\varepsilon_{(n_j^g n_{ij}^p)b_j t}(n_j^g n_{ij}^p) + \varepsilon_{(c_j^g c_{ij}^p)b_j t}(c_j^g c_{ij}^p) + \varepsilon_{(n_j^g n_{ij}^p)w_{ij} t}(n_j^g n_{ij}^p) + \varepsilon_{(c_j^g c_{ij}^p)w_{ij} t}(c_j^g c_{ij}^p) \tag{2.25}$$

where $\beta'_{(c_j^g c_{ij}^p)t}$ represents the intervention effect on compliers due to the delay in receiving active intervention, errors $\varepsilon_{(n_j^g n_{ij}^p)b_j t}$ (never-taker) and $\varepsilon_{(c_j^g c_{ij}^p)b_j t}$ (complier) are assumed to be normally distributed with zero mean and the between groups variance $\sigma^2_{(n_j^g n_{ij}^p)b_j t}$ and $\sigma^2_{(c_j^g c_{ij}^p)b_j t}$ for those groups assigned to start intervention at time t. Similarly, errors $\varepsilon_{(c_j^g n_{ij}^p)w_{ij} t}$ and $\varepsilon_{(c_j^g c_{ij}^p)w_{ij} t}$ are assumed to be normally distributed with zero mean and the within groups variance $\sigma^2_{(c_j^g n_{ij}^p)w_{ij} t}$ and $\sigma^2_{(c_j^g c_{ij}^p)w_{ij} t}$ for individuals starting intervention at time t. Equation 2.23 can be applied here to predict the principal strata membership for each group because we assume that the membership of each principal stratum will not change over time. The likelihood function will become:

$$\begin{aligned} L(\theta \mid \text{data}) \propto & \prod_t \Big(\prod_{j=1}^{k1}\Big(\prod_{i \in \{Z_j=1, S_j=0)} \pi_{n^g n^p} g(y_{ij} \mid \mu_{n^g n^p}, \sigma^2)\Big)\Big) \\ & \times \prod_t \Big(\prod_{j=1}^{k2}\Big(\prod_{i \in \{Z_j=1, S_j=1)} \pi_{c^g c^p} g(y_{ij} \mid \mu_{c^g c^p 1}, \sigma^2)\Big)\Big) \\ & \times \prod_t \Big(\prod_{\text{j}}\Big(\prod_{i \in \{Z_j=0, S_j=0)} [\pi_{n^g n^p} g(y_{ij} \mid \mu_{n^g n^p}, \sigma^2) + \pi_{c^p} g(y_{ij} \mid \mu_{c^g c^p 0}, \sigma^2)]\Big)\Big) \end{aligned} \tag{2.26}$$

In Chapter Three we apply the new methods developed in this chapter to a two-level randomized trial, the QPR gatekeeper study.

Chapter Three

Using the Two-Level Principle Stratification Method to Evaluate the QPR Gatekeeper Training Program

Four main models, which can be possibly formed at a two-level randomized trial by using Principal Stratification method, were developed and discussed in the previous chapter. Both the method of moment estimator and the marginal maximum likelihood estimator for mixtures have been discussed for these four models. This chapter will apply these models to a specific two-level randomized trial, the Georgia Gatekeeper Study. It is the first randomized trial of the gatekeeper training in a school-based setting, and it is a crossover design trial in which schools have a random time to change from control to intervention. The time when schools started to receive training varied due to participation at the first level. This allows us to examine whether intervention effects varied by participant status and whether the effects continued over time. However, based on the nature of the study design, in which the probability of all subjects within a group under control condition received intervention was zero and group-level numbers of student referred to receive mental health professional assessment were collected for each school due to de-identified issue, G-Bloom model will be applied to evaluate an intervention causal effect. The exclusion restriction assumption was investigated because there were time intervals where the intervention condition of schools was changed from control to active intervention. In particular, strong and weak exclusion restriction assumptions will be introduced for this type of data. These two assumptions will be compared in the following analyses.

This chapter will start with an introduction of the intervention that was used in the randomized trial, Gatekeeper Training Program. Then it will be followed by an introduction of study design. In the analysis section, Principal Stratification method will be used to evaluate the effect of the Gatekeeper Training Program. Analyses will be conducted for two situations, the first time period only and the entire four study periods. The results will be compared with two traditional methods, Intent-to-Treat (ITT) and As-Treated (AT). Similar analyses will be conducted separately by school type as well.

3.1 Introduction

Nearly 4,000 people aged 15–24 die by suicide each year in the United States. The gatekeeper training program is one type of intervention that has been designed and conducted to prevent youth suicides. The school gatekeeper training program is a school-based program that is designed to train all school staffs, who act as gatekeepers, in order to improve early identification of students at high risk for suicide and to facilitate timely referrals for mental health services. School gatekeepers can include any adults in the school (e.g., counselors, teachers, coaches, administrators) who are in a position to observe and interact with students. The main purpose of the gatekeeper training program is to increase awareness and knowledge about youth suicide risk, to directly ask troubled youth if they are suicidal, and to help suicidal youth to receive appropriate mental health services. Increased knowledge of risk factors for suicide, and changing attitude towards asking troubled youth if they are suicidal have the potential for increasing referral behavior and may promote early identification of suicidal students.

Although the gatekeeper training program has been widely applied to various communities under different social and environmental settings, it has not been rigorously tested and evaluated. Most studies simply reported that the training program is helpful (Nelson, 1987, Barrett, 1985, Spiritto, et al., 1988). Based on the reported finding so far, the program basically increases awareness of suicide warning signs, knowledge of treatment resources, and willingness to make referrals to mental health professionals among gatekeepers (Shaffer et al., 1988).

However, the studies to examine the effects of the program on actual number of referrals were rare. One randomized trial was conducted in Cobb County, Georgia, to evaluate the training effect of the QPR (*Question, Persuade, Refer)* (Quinnett, 1995) gatekeeper training program on knowledge of suicide, appraisal including willingness to assume a 'gatekeeper' role for suicide prevention, self-reported intervention behaviors with students, and improvement of early detection. The QPR training has been shown to clearly increase knowledge of suicide warning signs, intervention behaviors, appraisals including gatekeeper efficacy, and service access, as tested on adults by an intent-to-treat analysis (Wyman, el at., 2008). To date the effects on youth referrals have not been reported. The analyses in this chapter will exam whether the QPR can increase the number of middle and high school students referred to receive mental health assessment. Of special interest is the fact that there was variation in the timing and completeness of the training of adults in schools. These may be due to self-selection factors that are relatively stable in the schools, or this variation may be due to more transitory or even spurious factors in the schools. The transitory and permanent referral behavior for different schools will also be investigated here.

3.2 Method

The QPR gatekeeper training occurred with school staffs in the Cobb County School District in Georgia. Funding for this study came from a National Institute of Mental Health grant. The trial started at January 8, 2004. It is the first randomized trial of the gatekeeper training in a school-based setting within the U.S. The intervention was used in this trial is called QPR gatekeeper training program. This dissertation will examine the QPR training effect on the outcome of the trial, which was measured by the number of middle and high school students who have been referred to receive mental health assessment. This section will start with an introduction of the QPR training, followed by description of how the randomized trial has been conducted, and how the method of analyses has been chosen and what analysis strategy is.

3.2.1 Study Design and Participant Population

3.2.1.1 Introduction of the QPR Gatekeeper Training

The QPR (*Question, Persuade, Refer*) Gatekeeper training program is designed to train school staffs directly. It increases knowledge on suicide among school staffs in order to help potential suicidal students to access professional services. School staffs will learn three basic life-saving intervention skills to provide suicide prevention among youth.

The three intervention skills are: question a person who shows warning signs about suicide, persuade the person to pursue professional help, and refer and direct the person to appropriate resources (Quinnett, 1995, 1999).

The QPR is designed based on the belief or theory that those individuals would like to talk about their distress to someone around them or whom they trust about their feelings if they are at risk for self-destruction and violence. It is also believed that suicide "warning signs" can be recognized by someone who possesses enough suicide knowledge and is trained professionally.

Professional staffs in the district's Prevention/Intervention Center (PI/C), school counselors, and school staffs in Cobb County School District, GA received three levels of the QPR training program respectively. In order to play a role as a trainer and evaluator, professional staffs received more than 12 hours training before the study started. Then a counselor from each school that was assigned to the training program received more than 6 hours training from PI/C professional staffs after the study started. Finally, professional staffs and the counselors co-led a one and one-half hour gatekeeper training session for school staffs (gatekeepers) in assigned schools. The training covered important knowledge about youth suicide and how to identify students at high-risk. The training also taught school staffs how to ask a student about suicide, to persuade a student to obtain help, and to refer a student to receive professional help. About a half-year after the initial training, the school staffs were invited for a 30-minute refresher training.

3.2.1.2 Study Design

The random assignment of the time that each school would be trained occurred at the school level. Thirty-two middle and high schools in the Cobb County School District participated in the study. Three schools were excluded from the trial since they already received training before the trial began. All thirty-two schools were stratified into 4 strata by school type (middle vs. high school) and the rates of student crisis referrals during the 2002-2003 school-year which preceded the trial (low vs. high). Within each stratum, one half of the middle and high schools were randomly selected to receive the QPR training during the 2003-2004 school-years. For the 16 schools on the "waiting list", training was planned to start during the following school year, since the school district felt strongly that all schools should receive the training. This classic wait-listed design offered the opportunity to compare referral rates for suicide that were reported to a central district office, among trained and un-trained schools. No schools withdrew from the study.

However, it was apparent to the research team immediately, that additional information could be obtained by extending the time of the trial while continuing to schedule the remaining schools at random times to be trained. The investigators prepared and published a technical paper documenting the advantage in power resulting from this continued random assignment of crossover times for training (Brown et al., 2006). With the approval of the Data Safety and Monitoring Committee and the funding agency, the trial design was modified so that it could be extended. In summary, by the end of 2004-2005 school-year, all 16 early training schools received the training. The study used "dynamic wait-listed" or roll-out design for the remaining 16 schools on the "waiting list" after the first year of the study (Table 10). These remaining 16 schools on the waiting list were stratified into 4 strata by school type and the level (rate) of student crisis referral during the previous year. One school from each school-size/referral rate stratum was randomly selected and assigned into a block. The 4 schools in each block were promptly scheduled to receive training one after another. This formed 5 different training periods within this 2-year study.

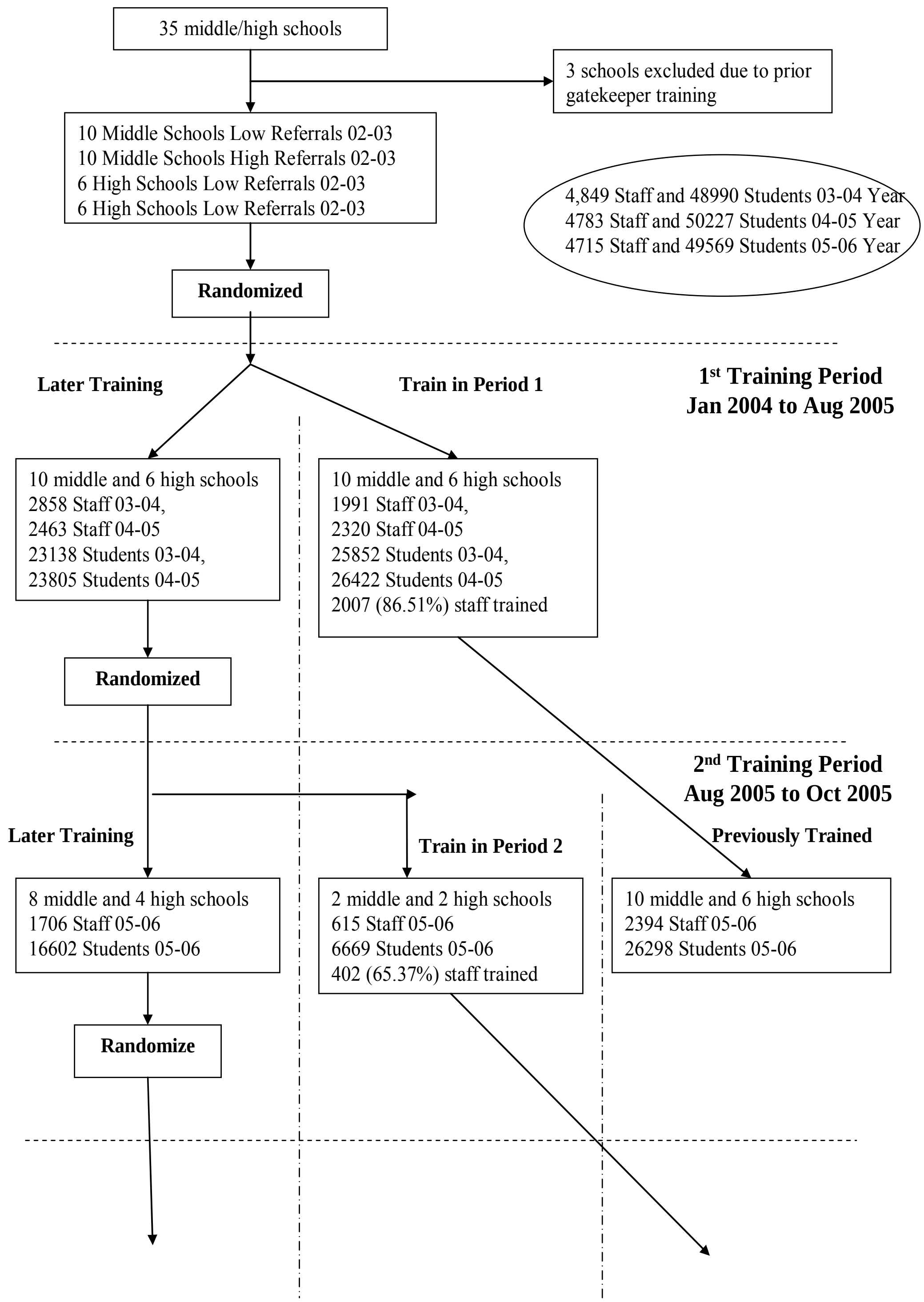
The Consort Diagram for Study Design (Figure 2) provides detail information of the study design, such as numbers of middle and high schools received training, and proportion of school staffs have been trained at each design time period.

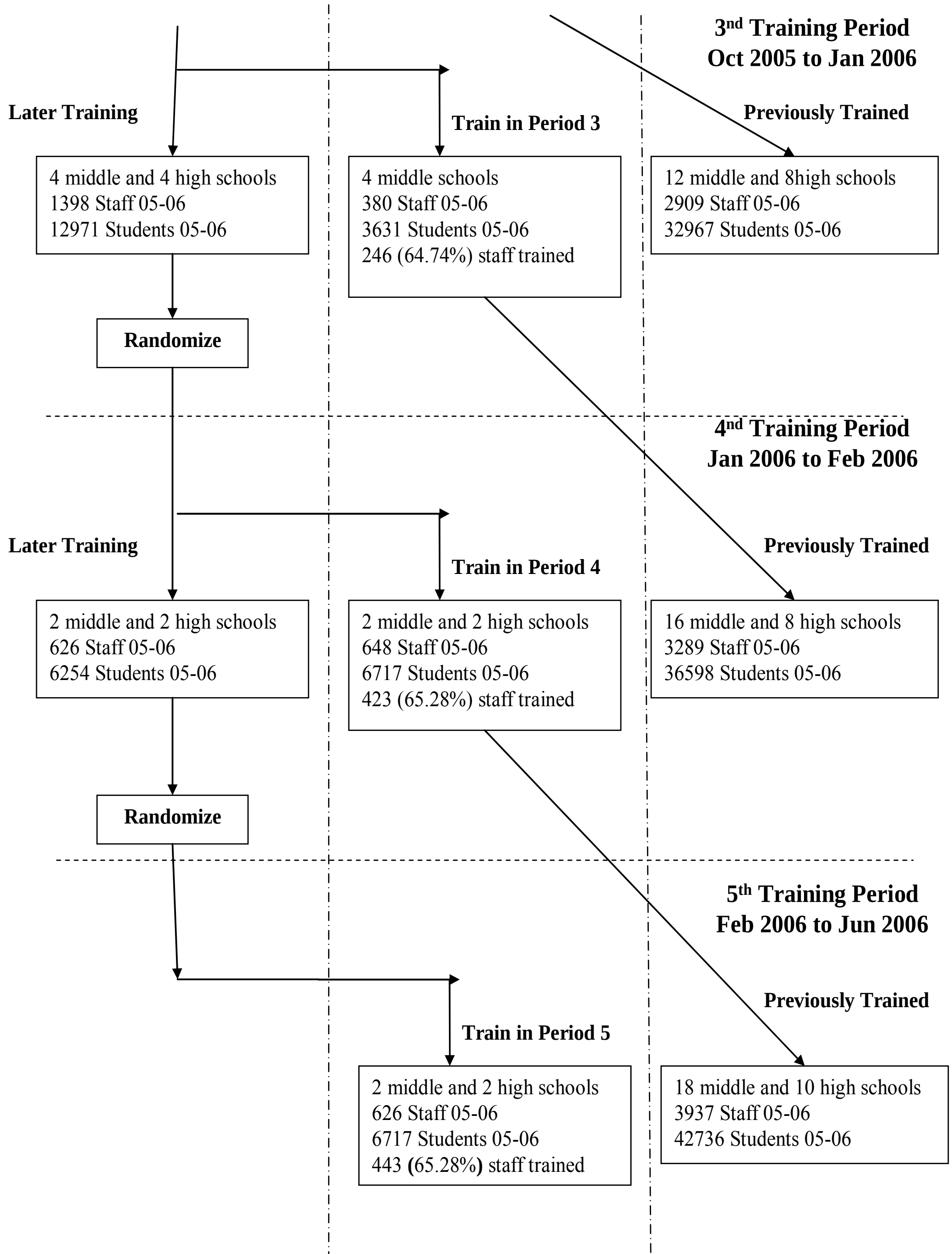
Table 10: Study Design of the QPR Gatekeeper Training Program

Year	Time Block	**Classic Wait-Listed Design**		**Dynamic Wait-Listed Design**	
	Period	**QPR Trained**	**Wait-Listed**	**QPR Trained**	**Wait-Listed**
1 Spring 04	1	**16 Trained: 14**	**16 Trained: 0**		
Fall 04 – Spring 05	1	**Trained: 2**	**Trained: 0**		
2	2			**20 Trained: 4**	**12 Trained: 0**
	3			**24 Trained: 4**	**8 Trained: 0**
	4			**28 Trained: 3**	**12 Trained: 1**
	5			**32 Trained: 4**	**12 Trained: 0**

Note: one middle school was scheduled training at period 4 but received training a few days into period 5; one high school was scheduled to be trained in period 5 but received training at the end of period 4

Figure 2: Consort Diagram for the Study Design of the QPR Gatekeeper Training Program





Note: one middle school was scheduled training at period 4 but received training at period 5, one high school was scheduled training at period 5 but received training at period 4.

3.2.1.3 Participant Population

The Cobb County School District is the second largest school system in Georgia. Its student population grows by nearly 2000 every school year. Over 97,000 students, in grades *Kg* - 12 were enrolled in the school system during the 2001-2002 school-year. 60% of them were White, 25% were African American, and 8% were Hispanic (Table 12).

A total of 48,990 students were in the schools that participated in the study during the 2003 – 2004 school-year. While these participants were all in grades 6 - 12, they have a similar distribution to the full population in terms of population ethnicity. The students were 56% White, 29% were African American, and remaining 15 % were Hispanic, Asian and others (Table 11). At the baseline, there were no significant differences between the 16 early training schools and the 16 wait-listed schools on race/ethnicity, gender, and grade level (Table 12).

Table 11: Ethnic Distribution of Cobb County School Students

	Total Student Population (2001-2002 school-years) (N =97,343)		Total Student Participants (2003 - 2004 school-years) (N = 48,990)	
	N	%	N	%
Ethnicity				
White	58,747	60.35	27,370	55.9
African American	24,267	24.98	14,295	29.2
Hispanic	7,953	8.17	4,164	8.50
Asian	3,572	3.67	1,767	3.61
Multi-Racial	2,541	2.61	1,277	2.61
American Indian	224	0.23	117	0.24

Table 12: Demographic Distribution of Students from the 32 Study Schools
(2003 - 2004 school-years)

	Training School (N =25,852)		Wait-Listed School (N = 23,138)	
	N	%	N	%
Ethnicity White	14,135	54.67	13,235	57.20
African American	8,035	31.08	6,260	27.06
Hispanic	2,126	8.22	2,038	8.81
Asian	871	3.37	896	3.87
Multi-Racial	622	2.41	655	2.83
American Indian	63	0.24	54	0.23
Gender Female	12,720	49.20	11,273	48.72
Male	13,132	50.80	11,865	51.28
Grade 6 grade	3763	14.56%	3737	16.15%
7 grade	3934	15.22%	3893	16.83%
8 grade	3954	15.29%	3845	16.62%
9 grade	4230	16.36%	3575	15.45%
10 grade	3635	14.06%	3011	13.01%
11 grade	3451	13.35%	2747	11.87%
12 grade	2885	11.16%	2330	10.07%

Among 13,080 staffs in the Cobb County schools during the 2001-2002 school-years, the majority (84.95%) were White, more than 12% was African American, and less than 2% were Hispanic (Table 13). Approximately two-thirds of the staffs were located in middle or high schools. This was the target population who took the QPR gatekeeper training course.

Of 13,080 staffs who held jobs in the 32 study schools, 4,853 received QPR training. However, we obtained training information only on 4,403 of them. Among those, 128 were administration staffs, 187 were supporting staffs, and the remaining 4,100 staffs were teachers. There were more full-time supporting staffs in training schools than in wait-listed schools (p = 0.002) (Wyman, el at., 2008). Among those school staffs, there were no significant differences in terms of gender, ethnicity, and years of experience between 16 early training schools and 16 wait-listed schools (Table 14).

Table 13: Ethnicity Distribution of Cobb County School Staffs

	Total Staff Population (2001-2002 school-years) (N =13,080)		Total Staff Participants (2003 - 2004 school-years) (N = 4,393)	
	N	%	N	%
Ethnicity				
White	11,111	84.95	3693	84.08
African American	1,631	12.47	581	13.22
Hispanic	191	1.46	72	1.64
Asian	94	0.72	19	0.43
Multi-Racial	41	0.31	21	0.48
American Indian	10	0.08	7	0.16

Table 14: Demographic Distribution of School Staffs from the 32 Study Schools
(2003 - 2004 school-years)

	Training School (N =2,263)		Wait-Listed School (N = 2,130)	
	N	%	N	%
Ethnicity White	1,843	81.44	1,850	86.87
African American	362	16.00	219	10.28
Hispanic	42	1.86	30	1.40
Asian	6	0.27	13	0.61
Multi-Racial	9	0.40	12	0.56
American Indian	1	0.04	6	0.28
Gender Female	1,575	69.60	1,481	69.21
Male	688	30.40	649	30.79
Years of Experience < 1 year	143	6.30%	132	6.10%
1 - 10 years	1187	52.40%	1,099	51.10%
11 - 20 years	504	22.30%	503	23.40%
21 - 30 years	354	15.60%	344	16.00%
more than 30 years	75	3.30%	62	2.90%

3.2.1.4 Measures

Within the school district, all crisis referrals were sent to the District's Prevention-Intervention Center (P/IC), which assessed the need for professional evaluation (Same-Day Assessment). A centralized record keeping system for reporting and referring youth for "life threatening behavior", either suicidal or homicidal ideation or behavior, had been in place for the past 15 years, and a crisis protocol system was in place to respond to the needs of these youth. An evaluated documentation was then completed by P/IC center staff for each referred student.

Due to de-identified issue, school level numbers of student referred to receive mental health professional assessment were collected. School level covariates such as gender, grade, and race/ethnicity were collected as well.

3.2.2 Methodology

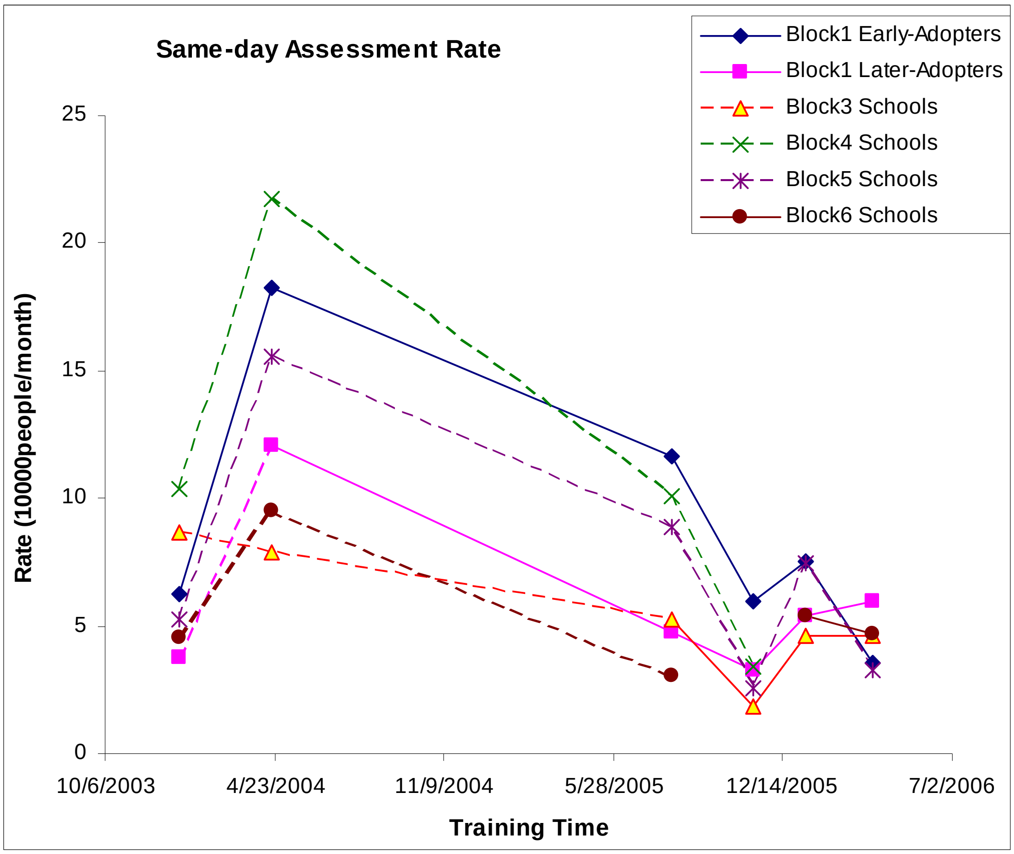
The Principal Stratification method has been used to examine the QPR training effect. Due to the nature of the study design, specifically, the G-Bloom model has applied in this example. The reason that the Principal Stratification method has been chosen is because (1) group level data have been collected; (2) sixteen schools had been selected and assigned to be trained in the QPR during the first period; (3) some schools under the QPR training condition started to receive training quickly while training in other schools was slow, and (4) the remaining sixteen schools had been assigned to the wait-listed condition. None of the 16 schools on the waiting-list had any chance to attend training.

The purpose of using this method is to compare the effectiveness of training by the intensity of the school's participation in training which could only be measured among the 16 schools selected for early training. We intended to estimate how many students among all referred from a school were due to the effect of QPR training. We also wanted to examine whether the QPR training has a different effect on the schools starting training early vs. later.

The evidence given by Figure 3 shows the Same-Day Assessment rate per 1000 people per month during the entire study period. Sixteen schools assigned to training group at first period were dichotomized as either “Early-Adopter school” or “Later-Adopter school” based on when they received training. A school is defines as an “Early-Adopter school” if its first training started within 81 day after beginning of the study (January, 2004) at the first period under training condition, otherwise, a school is a “Later-Adopter school”. This cut point is the median value of the first training time so that the observed compliance rate ($\hat{\pi}$) is 50%. Therefore, thirty-two schools have been divided into 6 groups. Both the Early-adopter group and the Later-adopter group consist of 8 training schools which received their first training within or after 81 days after training started. The schools in Block1 (Figure 3) are 4 schools selected to receive training during the second period. Block2, Block3 and Block4 each include 4 schools selected to receive training during the third, forth and fifth period respectively. The 6 lines in Figure 3 represent the variations of the Same-Day Assessment rate per 1000 people per month over the time for 6 different school-groups. The solid line denotes the school-groups under training condition, and the dash line indicates those under control condition. Note that the Early-adopter group has continuous solid lines while all the other schools begin with dashed lines and then convert to solid lines when they begin training. The Early-adopter group (black line) appears to have a higher Same-Day Assessment rate than the Later-adopter group (pink dashed and solid line) except at the last time point when referral rates are comparatively low overall (Figure 3). This suggests self-selection factors that are persistently different between the early adopter and later adopting schools. Another indication of the persistent differences across schools is the fact that the patterns of referrals are generally similar and parallel over time. Except for the schools in Block3, all school-groups show a similar temporal pattern in their Same-Day Assessment rates, which increases at the beginning and reaches the maximum at the second time point, then decreases towards the third time point, at which the rates are still higher than at the first time point. The rates continuously drop towards the fourth time point, and increase somewhat towards the fifth time point, and show certain departure from parallelism at the 6th time point. Contrarily, the Same-Day Assessment rate for Block3 schools decreases from the beginning until the third time point, then follows a similar pattern as other school-groups. Figure 3 does show that there is a continuing differential rate of referrals throughout for the early adopters in Block 1, and it also displays the dramatic difference in referral rates across time. When we examine whether there are changes in referral rates as a function of training time, that is when the curves change from dotted to solid lines, we do not perceive any major shifts that occur as a function of training.

There is heterogeneity in the timing of training and number of staffs trained in the QPR schools that were selected to be trained during the first period. One implication of this heterogeneity is that it may be a self-selection factor; some concerns, which may be due to the principal, school counselor, or the climate of the school, may lead some schools to adopt training much more readily and intensively than did other schools.

Figure 3. Same-Day Assessment Rates of 6 Block Schools across Time



3.2.2.1 Method

The moment estimate and mixtures and marginal maximum likelihood estimate notations of the causal effect within the stratum of complier (early-adopter) have been developed in Chapter Two. Those notations can be extended for this example.

$$\text{Let } \begin{cases} j = 1, 2, ..., 32 \text{ schools} \\ i = 1, 2, ..., N_j \text{ students in the } j^{th} \text{ school sector by gender, race/ethnicity and grade level} \\ T = 1, 2, 3, 4, 5 \text{ period} \\ t = 1, 2, ... 64 \text{ time interval} \end{cases}$$

$$Z_{jt} = \begin{cases} 0 & \text{if the } j^{th} \text{ school was assigned to the waiting list at time } t \\ 1 & \text{if the } j^{th} \text{ school was assigned to the QPR training condition at time } t \end{cases}$$

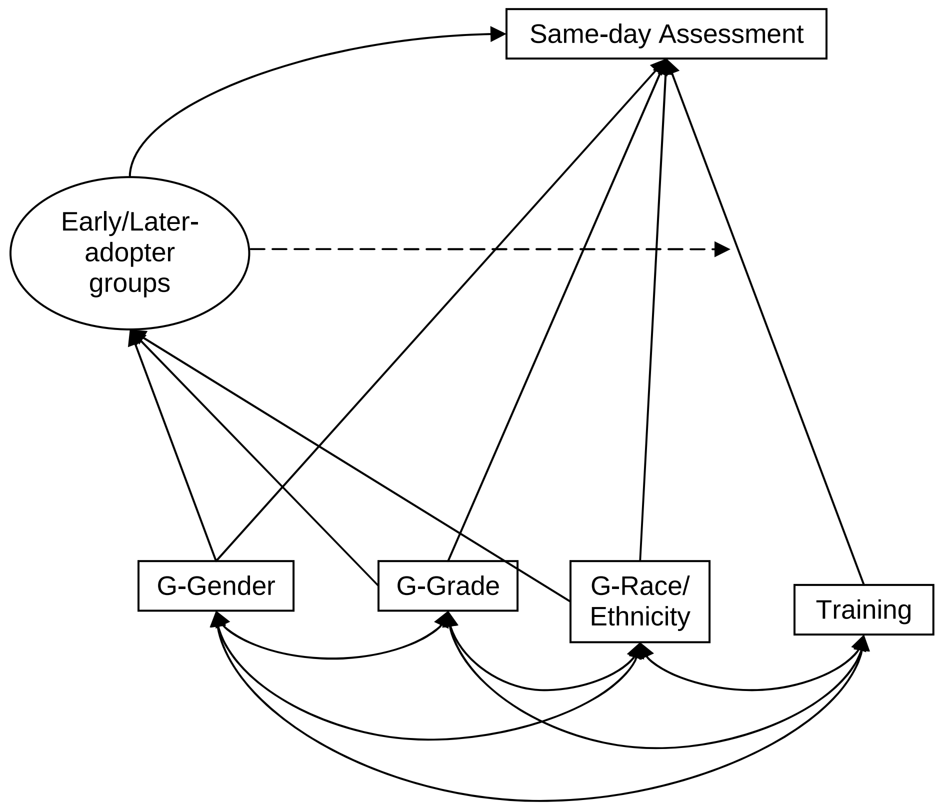
$$S_j^g = \begin{cases} 0 & \text{if the } j^{th} \text{ school started training later} \\ 1 & \text{if the } j^{th} \text{ school started training early} \end{cases}$$

Note that T represents the design study period and t represents the actual time point or day when a school receiving its training. In our analyses we have broken all the time intervals into sub-periods when any additional staff training occurred in any school. The post-treatment variable, S_j^g, is only observed if $Z_j = 1$ and S_{ij}^p is always equal to S_j^g under an intervention condition since we are unable to observe the training status of each individual school staff. Therefore, all schools will be classified into two group-level principal strata but the principal stratum membership of each school will not change over time. The two group-level principal strata are $C_j^g = c_j^g$, complier (early-adopter) stratum $(S_j^g(0), S_{ij}^p(0), S_j^g(1), S_{ij}^p(1)) = (0, 0, 1, 1)$, and $C_j^g = n_j^g$, never-taker (later-adopter) stratum $(S_j^g(0), S_{ij}^p(0), S_j^g(1), S_{ij}^p(1)) = (0, 0, 0, 0)$.

However, the principal stratum membership is always missing for schools in the wait-listed condition and can be treated as a mixture problem either without school level covariates used as predictors of principal stratum membership or with school level covariates used as predictors (Little & Yau, 1998).

Figure 4 shows a simplified schematic drawing of the model analysis presented in the study. The count variable, “Same-Day Assessment”, is regressed on the school level covariates “g-grade”, “g-gender”, “g-race/ethnicity” and the intervention variable “training”. The categorical latent variable C_j^g (Early/Later adopter groups) is the training-receiving status of a school with class 1 referring to “Early-adopter group” and class 2 referring to “Later -adopter group”. This variable is observable for the training condition, but unobservable for the wait-listed condition. The arrow from “Early/Later adopter groups” to the “Same-Day Assessment” indicates that the intercept of “Same-Day Assessment” across the classes of C_j^g (Early/Later adopter groups). The arrows from “Early/Later adopter groups” to “training” indicate that the slopes in regression of “Same-Day Assessment” on “training” vary across the classes of C_j^g (Early/Later adopter groups). The arrows from “g-grade”, “g-gender”, and “g-race/ethnicity” to C_j^g (Early/Later adopter groups) in Figure 4 represent the multinomial logistic regression of “Early/Later adopter groups” on these covariates.

Figure 4. A Simplified Schematic Drawing of the Model Analysis for the Georgia Gatekeeper Project



In this study, the response variable is the number of same day assessments from each school within each interval of time. Let us define $U_{jtzc^g}(x)$ as the number of same day assessments that are indexed by school level characteristics, i.e., in our case, x represent the cross classification by school, gender, race/ethnicity, and grade level. Note also that this total is indexed by time t, as well as principal stratum C^g and school j. Then we define $W_j = \sum_{tx} U_j$ to be total number of referred students for j^{th} school.

Note that the moment estimators developed from Chapter Two are:

$$\overline{W}(Z=0) = \frac{1}{k}\sum_{j=1}^{k} W_j(0) \ ,$$

$$\overline{W}_{n^g}\ (Z\ =1) = \frac{1}{k_1}\sum_{j\in n} W_j(1),$$

and

$$\overline{W}_{c^g}\ (Z\ =1) = \frac{1}{k_2}\sum_{j\in c}^{k_\pi} W_j(1)$$

The unbiased moment estimator of the causal effect within the stratum of complier (early-adopter) is:

$$A\hat{C}E\ =\ \overline{W}_{c^g}\ (Z\ =1)\text{-}\ (\frac{\overline{W}(\mathrm{Z}=0)\ \text{-}\ (1-\hat{\pi})\times\ \overline{W}_{n^g}\ (Z\ =1)}{\hat{\pi}})$$

The standard error (SE) of $A\hat{C}E$ is approximately equals to:

$$\sqrt{\mathrm{v\hat{a}r}(\overline{W}_{c^g}\ (Z\ =1))\ +\frac{\mathrm{v\hat{a}r}(\overline{W}(\mathrm{Z}=0))\ +\ (1-\hat{\pi})^2\times \mathrm{v\hat{a}r}(\overline{W}_{n^g}\ (Z\ =1))}{\hat{\pi}^2}}$$

However, students have been referred may vary by their characteristics. By applying the mixtures and marginal maximum likelihood estimate method, $U_{jtzc^g}(x)$ is assumed to have a Poisson distribution with a mean of λ_{jtxzc^g}. The overall model considered is given by the general set of predictors of this Poisson rate.

$$\begin{aligned}\log(\lambda_{jtxzC^g}) = offset_{jtxzC^g} + \alpha + \beta_x x + \\ \beta_{zx} ZX + \beta_{zc^g} ZC^g + \beta_{zt} ZT + \\ b_t + b_j\end{aligned} \tag{3.1}$$

On the first line of Equation 3.1, it includes an offset term corresponding to the number of students in that school with covariates x times the duration of the time for interval t, the effect of known covariates, and intercept. On the second line, three fixed effects have been listed as interactions between assignment and covariates, assignment and strata, and assignment and time. On the last line, two random effects have been listed to take into account of variation across time and schools. Both of these latter terms can include known covariates, i.e., the effect of longer periods of time and the contrast between middle and high schools.

As mentioned at the beginning of this chapter, the exclusion restriction assumption may be violated in the study due to different intervals in time where schools' training status changes from control to active intervention. Therefore, two additional assumptions, Strong Exclusion Restriction and Weak Exclusion Restriction are introduced.

The Strong Exclusion Restriction implies that no causal effect among later adopter groups exist at any time point. The Weak Exclusion Restriction implies that there is a causal effect among later adopter groups. Specifically, for all time periods T, all schools j, all categories of covariates x, and for late adopters $C^g = n^g$, if $\lambda_{jtx1n^g} = \lambda_{jtx0n^g}$, we say that the strong exclusion restriction holds. If on the other hand, this relationship only holds until the school is formally trained, or $\lambda_{jtx1n} = \lambda_{jtx0n}$ for time t where school j is converted to a fully trained condition, then we say the weak exclusion restriction applies.

3.2.2.2 Hypotheses

By applying the G-Bloom model of the Principal Stratification method to examine the QPR training effect. The following questions can be answered:

Hypothesis 1: the QPR training has an effect on increase of student "Same-Day Assessment" rates at the first training period, and, the effect is different for the schools start the training early vs. later.

Hypothesis 2: the QPR training effect varies by students' characteristics such as gender, race/ethnicity, and grade level at the first training period, and, the effect is same for all schools regardless their starting time of training.

Hypothesis 3: the QPR training effects persist over time, varies by school starting time of training (training early vs. later), and varies by students' gender, race/ethnicity, and grade level.

Hypothesis 4: the QPR training effects vary by school type, middle school vs high school.

3.2.2.3 Model Selection Strategy

In order to answer the questions listed above, the table below illustrates the specific sets of models that have been examined relative to the options in Equation 3.1 (all contain the same offset, which is ignored in this table). The analysis starts with strong exclusion restriction.

Step 1: fit a model that contains all main effects resulting from school level's covariates, gender, race/ethnicity, and grade, and training;

Step 2: set equal slopes on school level's covariates, gender, race/ethnicity, and grade for "Early-adopter group" and "Later-adopter group", then compare with the model from step 1 to test whether main effects on school level's characteristics are same between the Early-adopter groups and Later-adopter groups (referred as Model A in Table 15);

Step 3: remove school level's covariates, gender, race/ethnicity, and grade from the model in Step 2 one at a time, then compare it with the Step 2 model to test whether the number of students who have been referred are associated with those covariates (referred as Model B in Table 15);

Step 4: add the interaction effect between training and school level covariates to the model from Step 2, then compare it with the Step 2 model to test whether any interaction effect exists between school level's characteristics and training for the Early-adopter groups and Later-adopter groups (referred as Model C in Table 15);

Step 5: add the interaction effect between training and time to the model from Step 2 then compare it with the Step 2 model to test whether there are any training effect over time (referred to Model D in Table 15);

Step 6: add other interested covariates such as percentage of training staffs of each school to the model from Step 5, then compare it with the Step 5 model to test whether those covariates are significant (referred as Model E in Table 15);

Step 7: reach a final model that contains main effects of school level's covariates, gender, race/ethnicity, and grade with equal slopes for the "Early-adopter groups" and the "Later-adopter groups", interaction effect between training and time, and main effect of training. During this step, equal intercepts between the "Early-adopter groups" and the "Later-adopter groups" are tested to examine whether a baseline variation exists between different adopter classes (referred as Model F in Table 15).

All comparison tests listed above are based on likelihood ratio tests. Repeat all those seven steps under weak exclusion restriction which assume there may have a training effect among Later-adopter group. Within each step, the results will be compared parallelly for different types of exclusion restriction assumption to examine which assumption is appropriate to the data.

Table 15. Model Selection Strategy

Models	Main Effect of Training	Main Effects of School Level Covariates	Interaction Effects between Individual Level Covariates and Training	Other Main Effects	Moderation Effects	Effects of time	Options in Model 3.1
Model A		Gender, Race/Ethnicity Grade			Equal slopes between Early/Later-adopter groups		$\beta_{xc^g} = \beta_{xn^g}$
Model B		Gender, Race/Ethnicity Grade			Between Early/Later-adopter groups		β_x
Model C			Training with Gender, Race/Ethnicity Grade		Between Early/Later-adopter groups		β_{zx}
Model D					Between Early/Later-adopter groups	Time period	β_{zt}
Model E				Percentage of training	Between Early/Later-adopter groups		
Model F	Training	Gender, Race/Ethnicity Grade			Baseline different between Early/Later-adopter groups	Time period	$\alpha_{xc^g} = \alpha_{xn^g}$

3.3 Analysis

Mplus version 5.0 with TWOLEVEL, RANDOM, and MIXTURE analysis type was used to calculate the intervention causal effect. Due to the first training period being the longest period within the two-year study, our current analysis will examine the QPR training effect at the first period and the entire study period separately. This also allows us to examine whether the QPR training effect is persistent over time.

3.3.1 Analyses Limited to the First Study Period

After model selection, the final model for the first study period only was constructed under the weak exclusion restriction assumption and contains main effects of gender, grade level, and race/ethnicity, and random effect in school by training time. The results show that there is a significant difference on baseline between the Early-adopter groups and the Later-adopter groups, with intercept values of 10.497 (SE = 0.200) and -16.049 (SE = 0.858) (Table 16), respectively. The slopes of the “Same-Day Assessment” on the main covariate effects are not significantly different between the “Early-adopter groups” and the “Later-adopter groups”. Table 16 shows that female students have a mean of “Same-Day Assessment” rate less than 30% (β = -0.327, SE = 0.128), which is higher than male students, and the difference is significant (p = 0.011). On average, the “Same-Day Assessment” rate for middle schools is higher than high schools. The 8th Grade has the highest “Same-Day Assessment” rate (β = 0.445, SE = 0.198) within middle schools. Meanwhile, the “Same-Day Assessment” rate increases with increase of grade level within middle schools. However, this association is opposite within high schools, where the 9th Grade has the highest “Same-Day Assessment” rate (β = 0.247, SE = 0.255). The White has a higher “Same-Day Assessment” rate than the other race/ethnicity categories, except for Multi-race (β = 0.155, SE = 0.290). The “Same-Day Assessment” rates for African American and Hispanic are significantly different from the White (p =0.004, p = 0.227). The mean intervention effect is -0.596 (SE = 0.215, 95% CI =(-1.026, -0.166)) for the “Early-adopter group”, which suggests that the “Same-Day Assessment” rate in the training schools is almost 60% lower than the rate in the wait-listed schools. In contrast, the “Later-adopter group” has a mean intervention effect of 4.7 (SE = 0.852, 95% CI = (2.996, 6.404)), indicating that the “Same-Day Assessment” rate for the training schools in this group is almost 100 times higher than the rate in the wait-listed schools. The reason that the “Later-adopter group” has a higher intervention effect than the “Early-adopter group” may be due to the facts that their school systems and school staffs were well prepared to participate in the training and all training instructors have received feedback from the early training period. Finally, the results show that the “Same-Day Assessment” rates are highly associated with school level characters such as gender, race/ethnicity, and grade level, and the training has different impact on the “Early-adopter group” and the “Later-adopter group”.

The following section will apply the Poisson regression model to the entire four time periods. The results will also be compared with two traditional methods, Intent-to-Treat (ITT) and As-Treated (AT).

Table 16: Results of the Poisson Regression Model under Weak Exclusion Restriction Assumption during the First Period of Study

Comparison	Estimate	S.E.	Est./S.E.	P-Value
Gender Male vs Female	-0.327	0.128	-2.556	0.011
Grade 7^{th} vs 6th	0.300	0.187	0.1603	0.109
8^{th} vs 6th	0.445	0.198	2.258	0.024
9^{th} vs 6th	0.247	0.255	0.972	0.331
10^{th} vs 6th	-0.204	0.255	-0.717	0.473
11^{th} vs 6th	-0.317	0.268	-1.183	0.237
12^{th} vs 6th	-0.935	0.369	-2.534	0.011
Race/Ethnicity Asian vs White	-0.384	0.357	-1.076	0.282
African American vs White	-0.501	0.174	-2.873	0.004
Hispanic vs White	-0.492	0.222	-2.218	0.027
Multi-race vs White	0.155	0.290	0.534	0.593
Early-Adopter Group Training Status	-0.596	0.215	-2.774	0.006
Intercept	-10.479	0.200	-52.432	< 0.0001
Later-Adopter Group Training Status	4.700	0.852	5.520	< 0.0001
Intercept	-16.049	0.858	-18.714	< 0.0001

3.3.2 Entire Study Period

All of the analyses that have been done so far are only applied to the first training period, which is the longest training period within the two-year study. During this period, the schools in the wait-listed condition did not have a chance to participate in the training. The compliance status for the training schools can be defined by their first start time of training. However, the study design has been changed in order to increase the study power (Brown et al., 2006) and increase participation rate within each school; training time for each training period has been narrowed down. The 16 remaining schools have been randomly assigned to receive training at four different time blocks. Eventually, all 32 schools finished their training within these five time periods. The following analyses will focus on the first four time periods to evaluate the QPR Gatekeeper training program effect over time. The two traditional methods, ITT (Intent-to-treat) Analysis and AT (As-treated) Analysis, were used first to examine the intervention effect over time. Then, the advanced method (Principal Stratification method), was used under both strong exclusion restriction and weak exclusion restriction assumptions to compare with ITT and AT analyses to investigate whether we can gain more information from this advanced method and are able to better examine the intervention effect over time for this study.

3.3.2.1 Summary of Analyses for All 32 Schools

Comparison of all estimates obtained from different methods show that the ITT method underestimates on almost all parameters (Table 17). The estimators on gender give a consistent message that male students have 17% (β = -0.19, SE = 0.11) lower "Same-Day Assessment" rates than female students by using different methods. The estimators on grade levels show that middle schools have a higher "Same-Day Assessment" rate than high schools on average. The "Same-Day Assessment" rates increase along with grade level in middle school but they show an opposite direction in high schools. The 8^{th} and 9^{th} grade students have the highest "Same-Day Assessment" rates among middle schools and high schools respectively. The multi-race students have the highest "Same-Day Assessment" rate among all race/ethnicity students. White students are the majority group in the school district. Their "Same-Day Assessment" rate is the second to the highest. Compared to ITT analysis, AT analysis has a positive estimate on intervention effect which means that the QPR Gatekeeper training program increased the "Same-Day Assessment" rate in the training group compared to the wait-listed group. Contrarily, the results from the Principal Stratification method show that the intervention has a positive impact on "Later-adopter groups" rather than "Early-adopter groups", regardless of the ITT or AT training status. Time has a strong positive impact on both "Later-adopter groups" and "Early-adopter groups". Period 2 has the highest "Same-Day Assessment" rate.

The Principal Stratification method with AT status under weak exclusion restriction has the best fit due to the smallest value of BIC. The results show that female students have a mean “Same-Day Assessment” rate that is 17% higher than do male students (β = -0.19, SE = 0.11). By grade, the 8th grade students have the highest referral rate (β = -0.44, SE = 0.17). Among different race/ethnicity groups, the Multi-race students have the highest referral rate (β = 0.17, SE = 0.25), and Hispanic students have the lowest (β = -0.46, SE = 0.19). The mean intervention effect is -0.52 (SE = 0.25, 95% CI = (-0.82, -0.02)) (Table 18) for the “Early-adopter group”, which suggests that the “Same-Day Assessment” rate in the training schools is almost 60% lower than the rate in the wait-listed schools. However, the “Later-adopter group” has a mean intervention effect of 0.67 (SE = 0.31, 95% CI = (0.05, 1.29)), indicating that the “Same-Day Assessment” rate for the training schools in this group is almost 2 times higher than the rate in the wait-listed schools.

Table 17: Estimate Comparison for All 32 Schools over Four Periods

Comparison	Method					
			PS(ITT)		PS(AT)	
	ITT	AT	Strong	Weak	Strong	Weak
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Gender Male vs Female	-0.47 (0.10)***	-0.19 (0.11)	-0.19(0.11)	-0.19 (0.11)	-0.19 (0.11)	-0.19 (0.11)
Grade 7th vs 6th	-0.63 (0.15)***	0.25 (0.17)	0.25 (0.17)	0.25 (0.17)	0.25 (0.17)	0.25 (0.17)
8th vs 6th	-0.43 (0.15)*	0.44 (0.17)*	0.44 (0.17)*	0.44 (0.17)*	0.44 (0.17)*	0.44 (0.17)*
9th vs 6th	-0.64 (0.18)***	0.17 (0.20)	0.24 (0.22)	0.23 (0.21)	0.24 (0.21)	0.24 (0.21)
10th vs 6th	-0.90 (0.20)***	-0.15 (0.22)	-0.10 (0.23)	-0.10 (0.23)	-0.10 (0.23)	-0.08 (0.23)
11th vs 6th	-0.96 (0.18)***	-0.27 (0.22)	-0.21 (0.23)	-0.21 (0.23)	-0.21 (0.22)	-0.20 (0.22)
12th vs 6th	-1.14 (0.18)***	-0.82 (0.28)*	-0.75 (0.29)*	-0.76(0.29)*	-0.75 (0.29)*	-0.75 (0.29)*
Race/Ethnicity Asian vs White	-0.68 (0.35)**	-0.26 (0.30)	-0.26 (0.30)	-0.26 (0.30)	-0.26 (0.30)	-0.27(0.30)
African American vs White	-0.65 (0.14)***	-0.35 (0.14)*	-0.34 (0.14)*	-0.34 (0.15)*	-0.34 (0.14)**	-0.36 (0.14)*
Hispanic vs White	-0.75 (0.20)***	-0.44 (0.19)**	-0.43 (0.19)**	-0.44 (0.19)**	-0.44 (0.19)**	-0.46 (0.19)**
Multi-race vs White	-0.63 (0.42)	0.17 (0.25)	0.17 (0.25)	0.17 (0.25)	0.17 (0.26)	0.17 (0.25)
Early-Adopter Group Intercept			-11.02 (0.28)***	-14.72 (4.85)***	-10.88 (0.21)***	-10.73 (0.19)*
Later-Adopter Group Intercept			-11.42 (0.25)***	-10.66 (0.19)***	-11.49 (0.23)***	-11.89 (0.29)***

* p-value < 0.01; ** p-value < 0.05; *** p-value < 0.0001

Table 18: Effect Estimates for all 32 Schools over Four Periods

Method	Effect Estimate (SE) Early- Adopter group	Effect Estimate (SE) Later- Adopter group
Intent-to-treat (ITT)	-0.04 (0.13)	
As-treated (AT)	0.02 (0.17)	
Principal Stratification (ITT) Strong Exclusion Restriction	-0.09 (0.30)	0
Weak Exclusion Restriction	3.61 (4.84)	-0.77 (0.23)**
Principal Stratification (AT) Strong Exclusion Restriction	-0.35 (0.25)	0
Weak Exclusion Restriction	-0.38 (0.23)	0.69 (0.32)**

* p-value < 0.01
** p-value < 0.05
*** p-value < 0.0001

3.3.2.2 Summary of Analyses for 20 Middle Schools

The same logic described above has been applied to test the training effect on middle schools only. The results show that the "Same-Day Assessment" rates in middle schools have similar slopes for both gender and race/ethnicity between "Early-Adopter groups" and "Later-Adopter groups". The results also show that the "Same-Day Assessment" rates are highly associated with race/ethnicity and grade level but not with gender.

The weak exclusion restriction assumption has a better fit than the strong exclusion restriction assumption with ITT training status, but not with AT training status. ITT training status has a lower BIC value than AT training status.

Overall, the Poisson regression model with ITT training status under the weak exclusion restriction shows that male students haves a 16% (β = -0.18, SE = 0.15) lower rate than female students (Table 19). The "Same-Day Assessment" rates increase as students get older. The 8^{th} grade has the highest "Same-Day Assessment" rate within middle schools which is 54% higher (β = 0.43, SE = 0.17) than 6^{th} grade and 25% higher than 7^{th} grade. All race/ethnicity categories have a lower "Same-Day Assessment" rate compared to Whites. The "Same-Day Assessment" rates for African American (β = -0.84, SE = 0.18) and Hispanic students (β = -0.48, SE = 0.23) are significantly different from the Whites. African American students have the lowest rates among middle school students. Their rates are less than half of those of White students. The overall intervention has a strong positive effect on the "Early-adopter group" and a negative effect on the "Later-adopter group". The "Same-Day Assessment" rate has been increased more than

2.7 times (β = 1.02, SE = 0.50) (Table 20) in the training schools than wait-listed schools for “Early-adopter groups”. However, the “Same-Day Assessment” rate has been decreased by 56% (β = -0.82, SE = 0.38) in the “Later-adopter” training schools. The intervention effect persist over time

Table 19: Estimate Comparison for 20 Middle Schools over Four periods

Comparison	Method					
			PS(ITT)		PS(AT)	
	ITT	AT	Strong	Weak	Strong	Weak
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Gender Male vs Female	-0.18 (0.15)	-0.18 (0.15)	-0.18 (0.15)	-0.18 (0.15)	-0.18 (0.15)	-0.18 (0.15)
Grade 7th vs 6th	0.25 (0.17)	0.25 (0.17)	0.25 (0.17)	0.25 (0.17)	0.25 (0.17)	0.25 (0.17)
8th vs 6th	0.43 (0.17)*	0.43 (0.17)*	0.43 (0.17)*	0.43 (0.17)*	0.43 (0.17)*	0.43 (0.17)**
Race/Ethnicity Asian vs White	0.59 (0.44)	-0.60 (0.44)	-0.59 (0.44)	-0.60 (0.44)	-0.60 (0.44)	-0.60 (0.44)
African American vs White	-0.84 (0.18)***	-0.81 (0.18)***	-0.84 (0.18)***	-0.84 (0.18)***	-0.81 (0.18)***	-0.81 (0.18)***
Hispanic vs White	-0.48 (0.23)**	-0.47 (0.23)**	-0.48 (0.23)**	-0.48 (0.23)**	-0.47 (0.23)**	-0.46 (0.23)**
Multi-race vs White	-0.02 (0.33)	-0.02 (0.33)	-0.02 (0.33)	-0.02 (0.33)	-0.02 (0.33)	-0.02 (0.33)
Early-Adopter Group Intercept			-11.50 (0.39)***	-13.45 (2.12)***	-11.01 (0.23)***	-11.01 (0.22)
Later-Adopter Group Intercept			-10.98 (0.30)***	-10.13 (0.23)***	-11.17 (0.285)***	-11.55 (0.43)***

* p-value < 0.01
** p-value < 0.05
*** p-value < 0.0001

Table 20: Effect Estimates for 20 Middle Schools over Four Periods

Method	Effect Estimate (SE)	
	Early- Adopter group	Later- Adopter group
Intent-to-treat (ITT)	0.35 (0.21)*	
As-treated (AT)	0.06 (0.21)	
Principal Stratification (ITT) Strong Exclusion Restriction	0.57 (0.39)	0
Weak Exclusion Restriction	1.02 (0.50)*	-0.82 (0.38)*
Principal Stratification (AT) Strong Exclusion Restriction	-0.04 (0.27)	0
Weak Exclusion Restriction	-0.14 (0.28)	0.58 (0.56)

* p-value < 0.01
** p-value < 0.05
*** p-value < 0.0001

3.3.2.3 Summary of Analyses for 12 High Schools

Overall, the Poisson regression model with ITT training status under weak exclusion restriction shows that female students have more than 20% higher (β = -0.22, SE = 0.17) “Same-Day Assessment” rate than male students in the high schools (Table 21). The 12th grade has the lowest “Same-Day Assessment” rate (β = -0.97, SE = 0.25) among students. The highest rate among race/ethnicity groups in high schools was found for American Indian students (β = 0.95, SE = 1.00). African American students have a consistently higher rate than White students in both “Early-Adopter group” and “Later-Adopter group”. Similar to middle schools, Hispanic students have a consistently lower rate than White students in both “Early-Adopter group” and “Later-Adopter group”. The overall intervention has a negative effect on the “Early-Adopter group” but a positive effect on the “Later-adopter group”. The Early-adopter training schools show a 18% decrease in the “Same-Day Assessment” rate compared to the wait-listed schools (β = -1.73, SE = 0.51) (Table 22); while the “Later-adopter” training schools have a “Same-Day Assessment” rate 3 times of the “Later-adopter” training schools (β = 1.23, SE = 1.57) (Table 22). The impact of the intervention increases over time in both the “Early-Adopter group” and the “Later-Adopter group”.

Table 21: Estimate Comparison for 12 High Schools over Four Periods

Comparison	Method					
	ITT	AT	PS(ITT)		PS(AT)	
			Strong	Weak	Strong	Weak
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Gender Male vs Female	-0.29 (0.15)	-0.20 (0.15)	-0.20 (0.162)	-0.22 (0.167)	-0.20 (0.16)	-0.21 (0.17)
Grade 10th vs 9th	-0.57(0.18)*	-0.81 (0.21)***	-0.32 (0.20)	-0.30 (0.20)	-0.31(0.20)	-0.30(0.20)
11th vs 9th	-0.73 (0.19)***	-1.02 (0.23)***	-0.43 (0.21)**	-0.42 (0.21)**	-0.44 (0.21)**	-0.42 (0.21)**
12th vs 9th	-1.39 (0.28)***	-1.58 (0.31)***	-0.98 (0.25)***	-0.97 (0.25)***	-0.96 (0.25)***	-0.96 (0.25)***
Race/Ethnicity American Indian vs White	-0.94 (5.07)	-0.50 (2.94)	0.92 (0.99)	0.95 (1.00)	0.95 (0.99)	0.98 (1.00)
Asian vs White	-2.94 (8.26)	-1.99 (2.19)	0.07 (0.40)	0.10 (0.41)	0.12 (0.40)	0.09 (0.40)
Africa American vs White	-0.09 (0.17)	-0.41 (0.23)	0.13 (0.21)	0.19 (0.21)	0.196 (0.215)	0.19 (0.21)
Hispanic vs White	-0.98 (0.42)*	-1.01 (0.38)**	-0.39 (0.34)	-0.45 (0.33)	-0.44 (0.33)	-0.46 (0.33)
Multi-race vs White	-0.67 (0.87)	-0.88 (0.97)	0.37 (0.39)	0.41 (0.40)	0.43 (0.40)	0.44 (0.40)
Early-Adopter Group Intercept	-10.49 (0.17)	-10.45 (0.16)	-10.17 (0.44)***	-9.35 (0.30)***	-10.31 (0.43)***	-9.77 (0.33)***
Later-Adopter Group Intercept			-11.84 (0.31)***	-12.80 (1.58)***	-11.51 (0.23)***	-11.84 (0.31)***

* p-value < 0.01; ** p-value < 0.05; *** p-value < 0.0001

Table 22: Effect Estimates for 12 High Schools over Four Periods

Method	Effect Estimate (SE) Early- Adopter group	Effect Estimate (SE) Later- Adopter group
Intent-to-treat (ITT)	-0.23 (0.19)	
As-treated (AT)	-0.003 (0.23)	
Principal Stratification (ITT) Strong Exclusion Restriction	-0.80 (0.35)**	0
Weak Exclusion Restriction	-1.73 (0.51)*	1.23 (1.57)
Principal Stratification (AT) Strong Exclusion Restriction	-0.72 (0.37)**	0
Weak Exclusion Restriction	-1.34 (0.57)**	0.63 (0.38)

* p-value < 0.01
** p-value < 0.05
*** p-value < 0.0001

3.4 Conclusion

Overall, the analyses show that the Principal Stratification method with a weak exclusion restriction assumption is the best model because of the nature of the study design, variation in starting time of school training, and the characteristics of schools (e.g. whether willing to receive the training earlier or later). Moreover, the training had a different impact on “Early-Adopter” and “Later-adopter” schools. The intervention effect is also different for middle and high schools. The training had a strong-long term positive impact on “Same-Day Assessment” rates. The “Same-Day Assessment” rate is also highly associated with school level characteristics. However, there are still limitations for this study. First, the outcome, the number of “Same-Day Assessments” from schools, was rare. Secondly, limited information was collected from student level which was used to predict the adopter class for the wait-listed schools.

Chapter Four
Conclusion

This study has extended current causal inference procedures from single-level principal stratification to two-level stratification in randomized trials. It applied the developed models on a two-level randomized trial in which stratification is determined at two levels to evaluate the causal effect of the intervention. This chapter will summarize the new development arising from this study, the contribution of this study in both methodology and application, the limitations, and the research goals in the future.

4.1 Methodological Contributions

Most previous applications of Principal Stratification have focused on studies where participation or compliance status is determined by a single post-treatment variable, even though randomization can be at either group or individual level (Vinokur et al., 1995) (McDonald et al., 1992). Moreover analyses of the causal effects have been done at the individual level even in the case of multilevel randomized trials such as the flu shot study (McDonald et al., 1992). Such an approach has ignored the group level participation status. In this dissertation, methods were developed for multi-level randomized trials where principal stratification membership may be determined by multilevel post-treatment variables. The following summarizes the main contributions of this dissertation.

The new method defines multi-dimensional (2) post-treatment variable (S) which allows us to combine the participation status from both the group level and the individual level and to list all possible principal strata. Several possible models that can be applied to two-level randomized trials where participation can be at the individual level only, group level only, or both individual and group levels have been discussed. The new method modifies assumptions that underly general causal inference for different situations in principal stratification. Moreover, it discusses how to obtain a moment or mixture and marginal likelihood estimate of causal effect for all possible models by listing relative equations and maximum likelihood functions. A very useful feature of mixture and marginal likelihood estimation is that it takes account of covariate effects. Understanding the complex influence of covariates is essential to understanding intervention mechanisms because those covariates may confound the intervention. In general, including covariates that are good predictors of compliance increases precision in estimation of compliance status, increases the power to detect complier average causal effect (CACE), decreases sensitivity if CACE estimates to violation of underlying assumptions, and increases identifiability of CACE when critical indentifying assumptions are relaxed (Booil, 2008).

Finally, the idea of defining the post-treatment variable (S) to multiple dimensions permits the method to be applied to multilevel randomized trials with more than two intervention arms or with more than two levels of randomization. For example, study subjects can be classified into more than 16 principal strata because higher dimensional post-treatment outcome can be created.

4.2 Limitations

However, this study did not consider the impact of intraclass correlation (ICC) on variance inflation in the estimation of CACE in multi-level randomized trials because it has been discussed in other literature (Booil, 2008). Power issues have not been explicitly discussed in this study, even though power is a great concern in planning randomized trials because sample size will be restricted due to the limitation of group level randomized trials (Booil, 2008).

4.3 Application of Findings

The primary example used in this dissertation, the Georgia Gatekeeper Study, started with a classic randomized design. However, due to variation in implementation of groups selected to participate in the intervention, the study design switched to the dynamic waitlist design which implied varying exposure level to schools that adopted the intervention at different times. This variation invited the application of the Principal Stratification method. This variation in exposure could be caused by the organization of the school system or the attitude or willingness of school administration regarding participation in the intervention. There was evidence of baseline variation between the "Early-adopter group" and the "Later-adopter group" among all 32 middle and high schools. Similarly, the baseline variations between the "Early-adopter group" and the "Later-adopter group" also existed in both middle schools and high schools when analyses were done separately.

As to the outcome of QPR (*Question, Persuade, Refer)* (Quinnett, 1995) services, females were 17% more likely to be referred than males. Middle schools had higher rates of referral than high school on average. Seventh, 8th, and 9th grades had higher rates of referral then 6th, 11th and 12th grades did. Multi-race and White students were more likely to be referred than Americans Indians. Females had higher referral rates than males in both middle schools and high schools. The referral rates increased in middle schools but decreased in high schools by grade level.

In this example, schools were broken down into a smaller unit within each training period in order to maintain the training schedule.

4.4 Further Discussion

Evaluation of the intervention effect is the objective of clinical trials. However, compliance is an important issue existing in prevention trials and other research studies. Evaluation of causal effect, taking account of non-compliance, is what Principal Stratification method aims to achieve. Applying the Principal Stratification method to multilevel randomized trials is current and important among researchers because it controls selection bias. The Principal Stratification method can be applied in any scientific fields which are able to conduct randomized trials, especially, in social science and mental health because group level randomized trials are common designs used in mental health prevention studies or community social behavior studies.

There are some alternative methods which address non-compliance issues such as as-treated analysis and per-protocol (PP) analysis. However, the randomization may be broken by using these two analyses, because as-treated analysis is the case where the subjects in the control condition are not measured for participation and the as-treated analysis may compare the participants in the intervention condition with a non-intervention group, which combines the non- participants in the intervention condition with all subjects in the control condition. The PP analysis focuses on the effect of compliance to the assigned treatment protocol (Ten Have, et al., 2008). When participation can be measured in the control condition, a PP analysis may compare participation in the intervention condition with the participation in the control condition. The exclusion of non- participation under the PP approach distinguishes it from the as-treated method. In the case where the control condition is not measured for participation, the PP analysis may contrast the participation in the intervention condition with all subjects in the control condition, excluding the non- participants in the intervention condition.

Principal Stratification method and propensity score method are two other methods used to reduce selection bias by equating groups based on a set of known covariates. They both estimate intervention causal effect by comparing potential outcomes under control and intervention condition and reduce overt bias which may be attributable to observed confounders by adjusting these covariates (Ten Have, et al., 2008). The propensity score method was introduced by Rosenbaum and Rubin (1983). It is defined as the conditional probability of intervention given background variables. However, the main difference between these two methods is that the Principal Stratification method is also able to adjust for hidden bias which is unobserved confounders such as compliance behaviors based on randomization, so it relies on having some "instrument" (what subjects were randomized) that affects the intervention level that subjects receive. The propensity score method does not assume anything was randomized, but instead relies on an assumption of un-confounded treatment assignment: it assumes that there is no hidden bias between the intervention and control conditions, and it assumes that only observed variables affect the intervention level that subjects receive (Angrist, et al., 1996; Rubin, 2001; Posner, et al., 2001; Landrum & Ayanian, 2001). Therefore, the Principal Stratification method can help better identify meaningful relationships between treatment compliance and non-compliance with respect to the effect of treatment on outcome. The following table (Table 23) summarizes the comparison of these alternative methods.

Further improvement on the modified Principal Stratification method presented in this research is still warranted due to the limitations discussed previously. Future studies should focus on how to make the Principal Stratification method more flexible for use on multilevel randomized trials due to variations of participation status at different levels, complex memberships of principal strata, and requirements of more sophisticated models. To apply the Principal Stratification method to longitudinal trials where the participation status of an individual subject or group may change over time, further development of the method would also be needed. Future studies should also address the issues of power and power calculation requirements of the Principal Stratification method, as currently there are no theoretical and practical bases for such exercises. Expansion of the Principal Stratification method to other types of analyses (e.g., survival analysis) and other research fields (e.g., pharmaceutical) may also draw extensive interests. It may also be attractive to seek alternative methods to control the selection bias when designing similar community trials such as those studies described in this research. We do believe that the Principal Stratification method has the potential to be applied to other relevant research fields and cover more situations and is thus a viable method, if appropriate development is conducted.

Table 23. Comparison of Alternative Methods

Methods	Strength	Limitation
Intent-to-treat (ITT)	• Includes all randomized subjects	Does not control selection bias
As-Treated (AT)	• Considers non-compliance	Broken randomization
Per-Protocol (PP)	• Considers non-compliance	Broken randomization
Propensity Score	• Reduces selection bias • Control observed confounders	Does not require randomization
Principal Stratification	• Reduces selection bias • Controls observed and unobserved confounders	Needs additional assumptions to limit the numbers of principal strata

References

Angrist, J. D., Imbens, G. W., & Rubin, Donald B. (1996). *Identification of causal effects using instrumental variables*. Journal of the American Statistical Association,

Barnard, J., Frangakis, Constantine E., Hill, J., & Rubin, Donald B. (2003). *Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City (with discussion)*. Journal of the American Statistical Association.

Brown, C. Hendricks, Liao, Jason (1999). *Principles for Designing Randomized Preventive Trials in Mental Health: An Emerging Developmental Epidemiology Paradigm*. American Journal of Community Psychology.

Brown, C. Hendricks, Wyman. Peter A, Guo. Jing, Peña. Juan. (2006). *Dynamic wait-listed designs for randomized trials: new designs for prevention of youth suicide*. Clinical Trials.

Bloom, Howard S. (1984). *Accounting For No-Shows In Experimental Evaluation Designs*. Evaluation Review.

Cox, D.R.(1958). *The Planning of Experiments*, John Wiley and Sons, New York.

Frangakis, Constantine E., Rubin, Donald B. (2002). *Principle Stratification in Causal Inference*. Biometrics.

Frangakis, Constantine E., Rubin, Donald B., Zhou, X. H. (2002). *Clustered encouragement design with individual noncompliance: Bayesian inference and application to advance directive forms*. Biostatistics.

Grilli L., Mealli F. (2008). *Nonparametric Bounds on the Causal Effect of University Studies on Job Opportunities Using Principal Stratification*. Journal of Educational and Behavioral Statistics.

Have, Thomas R. Ten., Normand, Sharon-Lise T., Marcus, Sue M., Brown, C. Hendricks, Lavori, Philip., and Duan, Naihua (2008). *Intent-to-treat vs. Non-intent-to-treat Analyses under TreatmentNon-adherence in Mental Health Randomized Trials.* Psychiatric Annals.

Hirano, Keisuke, Imbens, Guido W., Rubin, Donald B., Zhou, Xiao-Hua (2000). *Assessing the Effect of an Influenza Vaccine in an Encouragement Design*. Biostatistics.

Hong, G. & Raudenbush, S. W. (2006). *Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data*. Journal of the American Statistical Association

Ialongo, Nicholas S., Werthamer, Lisa, Kellam, Sheppard G., Brown, C. Hendricks, Wang, Song B., & Lin, Yu H. (1999). *Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression, and antisocial behavior.* American Journal of Community Psychology, 27, 599-641.

Jo, Booil (2002). *Estimation of Intervention Effects with Noncompliance: Alternative Model Specification*. Journal of Educational and Behavioral Statistics.

Jo, Booil， Asparouhov, Tihonmir， Muthen. Bengt O． (2008). *Intention-to-Treat Analysis in Cluster Randomized Trials with Noncompliance*． Statistics in Medicine.

Jo, Booil, Asparouhov, Tihonmir, Muthen. Bengt O., Ialongo, Nicholas S., Brown, C. Hendricks. (2008). *Cluster Randomized Trials With Treatment Noncompliance*. Psychological Method.

Jo, Booil. (2008). *Causal Inference in Randomized Experiments With MediationalProcesses*. Psychological Method.

Landrum, M.B. and Ayanian, J.Z. (2001). *Causal effect of amulatory specialty care on mortality following myocardial infarction: A comparison of propensity score and instrumental variable analyses*. Health Services and Outcomes Research Methodology

Posner, M.A., Ash, A.S., Freund, K.M., Moskowitz, M.A., and Shwartz, M. (2001). *Comparing standard regression, propensity score matching, and instrumental variables methods for determining the influence of mammography on stage of diagnosis.* Health Services and Outcomes Research Methodology.

Quinnett, P. (1995). *QPR: Ask a question, save a life*. Spokane, WA: QPR Institute and Suicide Awareness/Voices of Education.

Rosenbaum, Paul. R., & Rubin, Donald B. (1983a). *Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome*. Journal of the Royal Statistical Society, Series B (Methodological).

Rosenbaum, Paul. R., & Rubin, Donald B. (1983b). *The central role of the propensity score in observational studies for causal effects*. Biometrika.

Rubin, Donald B. (1974). *Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies*. Journal of Educational Psychology.

Rubin, Donald B. (1978). *Bayesian-Inference for Causal Effects – Role of Randomization*. Annals of Statistics.

Rubin, Donald B. (1978). *Randomization Analysis of Experimental Data – the Fisher Randomization Test – Comment*. Journal of the American Statistical Association.

Rubin, D.B. (2001). *Using propensity scores to help design observational studies: Application to the tobacco litigation*. Health Services and Outcomes Research Methodology.

Stuart, Elizabeth A. & Green, Kerry M. (2008). *Using Full Matching to Estimate Causal Effects in Nonexperimental Studies: Examining the Relationship Between Adolescent Marijuana Use and Adult Outcomes*. Developmental Psychology.

Sobel, Michael E. (2006). *What Do Randomized Studies of Housing Mobility Demonstrate?: Causal Inference in the Face of Interference*. Journal of the American Statistical Association.

Vinokur, Amiram D., Price, Richard H., and Schul, Yaacov. (1995). *Impact of IOBS Intervention on Unemployed Workers Varying in Risk for Depression*. American Journal of Community Psychology.

Wyman, Peter A., Brown, C. Hendricks, Inman, Jeff, Cross, Wendi, Schmeelk-Cone, Karen, Guo, Jing, Pena, Juan B.(2008). *Randomized Trial of a Gatekeeper Program for Suicide Prevention: 1-Year Impact on Secondary School Staff.* Journal of Consulting and Clinical Psychology.

Appendices

Appendix A: Compute a Chi-square Difference Test

Following are the steps needed to compute a chi-square difference test based on log-likelihood values and scaling correction factors obtained with the MLR estimator.

1. Estimate the nested and comparison models using MLR. The printout gives log-likelihood values L0 and L1 for the H0 and H1 models, respectively, as well as scaling correction factors c0 and c1 for the H0 and H1 models, respectively. For example,

   L0 = -2,606, c0 = 1.450 with 39 parameters (p0 = 39)
   L1 = -2,583, c1 = 1.546 with 47 parameters (p1 = 47)

2. Compute the difference test scaling correction where p0 is the number of parameters in the nested model and p1 is the number of parameters in the comparison model.

   cd = (p0 * c0 - p1*c1)/(p0 - p1)
   = (39*1.450 - 47*1.546)/(39 - 47) = 2.014

3. Compute the chi-square difference test (TRd) as follows:

   TRd = -2*(L0 - L1)/cd
   = -2*(-2606 + 2583)/2.014 = 22.840

4. S-plus Code:

```
ChisqTest <- function( L0, C0, P0, L1, C1, P1) {
	cd <- (P0*C0 - P1*C1)/(P0-P1)
	df <- P0 - P1
	TRD <- -2*(L0 -L1)/cd
	PValue <- 1-pchisq(abs(TRD), abs(df))
	return( TRD, df, PValue )
}
```

```
Example:
#First period
#strong assumption
ChisqTest (-14604.866, 2.460, 28, -14610.212, 3.482, 16) #2
ChisqTest (-14604.866, 2.460, 28, -14604.878, 2.500, 27) #3
```

Appendix B: M-plus Code – ITT Model 01: Main effects within first training period

```
TITLE:  ITT-Model01_median_time;
 DATA: FILE IS J:\data\cpi.numericB.txt;

  Define:

    t1 =1; t2 =1;
    if Tcom eq 0 then t2=0;              ! early training group
    if Tcom eq 1 then t1=0;              ! later training group

    G7 = ( Grade EQ 7);                  ! create dummy variables for grade level
    G8 = ( Grade EQ 8);
    G9 = ( Grade EQ 9);
    G10 = ( Grade EQ 10);
    G11 = ( Grade EQ 11 );
    G12 = (Grade EQ 12);

    R2 = ( Race EQ 1);                    ! create dummy variables for race/ethnicity
    R3 = ( Race EQ 2);
    R4 = ( Race EQ 3);
    R5 = ( Race EQ 4);
    R6 = ( Race EQ 5);

                                          ! create dummy variables for interaction terms
    SM = S*M;                             ! -- status * gender
    SG7 = S*G7;                           ! -- status * grade
    SG8 = S*G8;
    SG9 = S*G9;
    SG10 = S*G10;
    SG11 = S*G11;
    SG12 = S*G12;
    SR2 = S*R2;                           ! -- status * race/ethnicity
    SR3 = S*R3;
    SR4 = S*R4;
    SR5 = S*R5;
    SR6 = S*R6;

  VARIABLE: NAMES ARE
                    School Male Grade Race Period Status Middle HiLevel
                    Duration Timing TimeInt CPI LPopDur PopDur ST
                    TimSchTr com Tcom Perc  First NotYet TNA TA
                     AFirst ANotYet ATNA ATA;
```

```
USEVARIABLES ARE  CPI Status ST  Male LPopDur    G7 - G12  R2 - R6   t1 t2;
       ! LPopDur = log (population*duration of training period
       ! ST = School*Time sequence

   COUNT is CPI;                          ! means CPI is a count data and follow a
                                          !  Poisson distribution

   CLUSTER = ST;                          !random effect for School * timing sequence

   WITHIN =  Male G7 - G12   R2 - R6   LPopDur ;
                                          ! main effect for Gender, Grade, and
                                          !  Race/ethnicity,
                                          ! offset = log (population * duration)
  BETWEEN = Status;                       ! Status = training status

  Class = c1(2);                          ! two levels

  TRAINING =t1 t2 ;                       ! later training group will have a similar
                                          !distribution as early training group

  Missing is ALL (999);

  USEOBS ARE  ( Period EQ 1 );        ! using period 1 data only


ANALYSIS: TYPE = TWOLEVEL random MIXTURE;

MODEL:
  %WITHIN%
    %OVERALL%

      CPI on  Male G7 - G12 R2 - R6   LPopDur @1;

       %c1#1%                             ! later training group
       CPI on Male G7 - G12  R2 - R6    LPopDur @1;


      %c1#2%                              !early training group
       CPI on  Male   G7 - G12 R2 - R6   LPopDur @1;



%BETWEEN%
```

```
%OVERALL%
     CPI on S;
     c1#1 on s @0;                    ! class level is independent with training status

      %c1#1%
      CPI on S @0;                     ! strong assumption: no training effect is
                                       !allowed

      %c1#2%
      CPI on S;                        ! only testing training effect on early training
                                       !group

OUTPUT: TECH1 TECH2;
```

Appendix C: M-plus Code – ITT Model 02: Main effects with equal slops within first training period

```
TITLE:  ITT-Model02_median_time_slops_eq;
  DATA: FILE IS J:\data\cpi.numericB.txt;

          t1 =1; t2 =1;
        if Tcom eq 0 then t2=0;              ! early training group
        if Tcom eq 1 then t1=0;              ! later training group

        G7 = ( Grade EQ 7);                  ! create dummy variables for grade level
        G8 = ( Grade EQ 8);
        G9 = ( Grade EQ 9);
        G10 = ( Grade EQ 10);
        G11 = ( Grade EQ 11 );
        G12 = (Grade EQ 12);

        R2 = ( Race EQ 1);                   ! create dummy variables for race/ethnicity
        R3 = ( Race EQ 2);
        R4 = ( Race EQ 3);
        R5 = ( Race EQ 4);
        R6 = ( Race EQ 5);

                                             ! create dummy variables for interaction terms
        SM = S*M;                            ! -- status * gender
        SG7 = S*G7;                          ! -- status * grade
        SG8 = S*G8;
        SG9 = S*G9;
        SG10 = S*G10;
        SG11 = S*G11;
        SG12 = S*G12;
        SR2 = S*R2;                          ! -- status * race/ethnicity
        SR3 = S*R3;
        SR4 = S*R4;
        SR5 = S*R5;
        SR6 = S*R6;

      VARIABLE: NAMES ARE
                          School Male Grade Race Period Status Middle HiLevel
                          Duration Timing TimeInt CPI LPopDur PopDur ST
                          TimSchTr com Tcom Perc  First NotYet TNA TA
                           AFirst ANotYet ATNA ATA;
```

```
USEVARIABLES ARE  CPI Status ST  Male LPopDur    G7 - G12  R2 - R6   t1 t2;

    ! LPopDur = log (population*duration of training period
    ! ST = School*Time sequence

  COUNT is CPI;                          ! means CPI is a count data and follow a
                                         !  Poisson distribution

  CLUSTER = ST;                          !random effect for School * timing sequence

  WITHIN =  Male G7 - G12   R2 - R6   LPopDur ;
                                         ! main effect for Gender, Grade, and
                                         !  Race/ethnicity,
                                         ! offset = log (population * duration)
  BETWEEN = Status;                      ! Status = training status

  Class = c1(2);                         ! two levels

  TRAINING =t1 t2 ;                      ! later training group will have a similar
                                         !distribution as early training group

  Missing is ALL (999);

  USEOBS ARE  ( Period EQ 1 );           ! using period 1 data only


ANALYSIS: TYPE = TWOLEVEL random MIXTURE;

MODEL:
  %WITHIN%
    %OVERALL%

     CPI on  Male G7 - G12 R2 - R6   LPopDur @1;

     %c1#1%                              ! later training group
     CPI on Male G7 - G12  R2 - R6    LPopDur @1;
     CPI on Male (M1);
                              ! define the slops on gender are same between two
                              !groups

  CPI on    R2   (b7); ! define the slops on race/ethnicty are same between two
                       ! groups
  CPI on    R3   (b8);
  CPI on    R4   (b9);
```

```
      CPI on   R5  (b10);
      CPI on   R6  (b11);

      CPI on   G7  (c7); ! define the slops on grade level are same between two groups
      CPI on   G8  (c8);
      CPI on   G9  (c9);
      CPI on   G10  (c10);
      CPI on   G11  (c11);
      CPI on   G12  (c12);

    %c1#2%                              !early training group
    CPI on  Male   G7 - G12 R2 - R6   LPopDur @1;

      CPI on Male (M1); ! define the slops on gender are same between two groups

      CPI on   R2  (b7); ! define the slops on race/ethnicty are same between two
                          !groups
      CPI on   R3  (b8);
      CPI on   R4  (b9);
      CPI on   R5  (b10);
      CPI on   R6  (b11);

      CPI on   G7  (c7); ! define the slops on geade level are same between two groups
      CPI on   G8  (c8);
      CPI on   G9  (c9);
      CPI on   G10  (c10);
      CPI on   G11  (c11);
      CPI on   G12  (c12);

     %BETWEEN%
       %OVERALL%
          CPI on S;
          c1#1 on s @0;              ! class level is independent with training status

           %c1#1%
           CPI on S @0;               ! strong assumption: no training effect is
                                      !allowed

           %c1#2%
           CPI on S;                  ! only testing training effect on early training
                                      !group

OUTPUT: TECH1 TECH2;
```

Appendix D: M-plus Code – ITT Model 09: Main effects with interactions within first training period

```
TITLE:  ITT-Model02_median_time_slops_eq;
  DATA: FILE IS J:\data\cpi.numericB.txt;

        t1 =1; t2 =1;
        if Tcom eq 0 then t2=0;              ! early training group
        if Tcom eq 1 then t1=0;              ! later training group

        G7 = ( Grade EQ 7);                  ! create dummy variables for grade level
        G8 = ( Grade EQ 8);
        G9 = ( Grade EQ 9);
        G10 = ( Grade EQ 10);
        G11 = ( Grade EQ 11 );
        G12 = (Grade EQ 12);

        R2 = ( Race EQ 1);                   ! create dummy variables for race/ethnicity
        R3 = ( Race EQ 2);
        R4 = ( Race EQ 3);
        R5 = ( Race EQ 4);
        R6 = ( Race EQ 5);

                                             ! create dummy variables for interaction terms
        SM = S*M;                            ! -- status * gender
        SG7 = S*G7;                          ! -- status * grade
        SG8 = S*G8;
        SG9 = S*G9;
        SG10 = S*G10;
        SG11 = S*G11;
        SG12 = S*G12;
        SR2 = S*R2;                          ! -- status * race/ethnicity
        SR3 = S*R3;
        SR4 = S*R4;
        SR5 = S*R5;
        SR6 = S*R6;

      VARIABLE: NAMES ARE
                        School Male Grade Race Period Status Middle HiLevel
                        Duration Timing TimeInt CPI LPopDur PopDur ST
                        TimSchTr com Tcom Perc  First NotYet TNA TA
                         AFirst ANotYet ATNA ATA;
```

```
USEVARIABLES ARE  CPI Status ST  Male LPopDur    G7 - G12  R2 - R6   t1 t2
              SM  SR2-SR6 SG7-SG12 t1 t2;
            ! LPopDur = log (population*duration of training period
     ! ST = School*Time sequence

  COUNT is CPI;                        ! means CPI is a count data and follow a
                                       !  Poisson distribution

  CLUSTER = ST;                        !random effect for School * timing sequence

  WITHIN =  Male G7 - G12   R2 - R6   LPopDur ;
                                       ! main effect for Gender, Grade, and
                                       !  Race/ethnicity,
                                       ! offset = log (population * duration)
  BETWEEN = Status;                    ! Status = training status

  Class = c1(2);                       ! two levels

  TRAINING =t1 t2 ;                    ! later training group will have a similar
                                       !distribution as early training group

  Missing is ALL (999);

  USEOBS ARE  ( Period EQ 1 );        ! using period 1 data only


ANALYSIS: TYPE = TWOLEVEL random MIXTURE;


MODEL:
 %WITHIN%
   %OVERALL%

    CPI on  Male G7 - G12 R2 - R6  SM  SR2-SR6 SG7-SG12 LPopDur @1;
     %c1#1%  ! later training group
     CPI on Male G7 - G12  R2 - R6   SM  SR2-SR6 SG7-SG12 LPopDur @1;
    CPI on SM (M);

    CPI on SR2 (R2);
    CPI on SR3 (R3);
    CPI on SR4 (R4);
    CPI on SR5 (R5);
    CPI on SR6 (R6);
```

```
        CPI on SG7 (G7);
        CPI on SG8 (G8);
        CPI on SG9 (G9);
        CPI on SG10 (G10);
        CPI on SG11 (G11);
        CPI on SG12 (G12);



        %c1#2%
         CPI on  Male   G7 - G12 R2 - R6   SM  SR2-SR6 SG7-SG12 LPopDur @1;
        CPI on SM (M);

        CPI on SR2 (R2);
        CPI on SR3 (R3);
        CPI on SR4 (R4);
        CPI on SR5 (R5);
        CPI on SR6 (R6);

        CPI on SG7 (G7);
        CPI on SG8 (G8);
        CPI on SG9 (G9);
        CPI on SG10 (G10);
        CPI on SG11 (G11);
        CPI on SG12 (G12);


  %BETWEEN%
    %OVERALL%
      CPI on S;
          c1#1 on s @0;                  ! class level is independent with training status

           %c1#1%
           CPI on S @0;                   ! strong assumption: no training effect is
                                          !allowed

           %c1#2%
           CPI on S;                      ! only testing training effect on early training
                                          !group

OUTPUT: TECH1 TECH2;
```

Appendix E: M-plus Code – ITT Model 13: Main effects within first training period with weaken condition

```
TITLE:  ITT-Model13_median_time_weaken;
 DATA: FILE IS J:\data\cpi.numericB.txt;

  Define:

    t1 =1; t2 =1;
    if Tcom eq 0 then t2=0;             ! early training group
    if Tcom eq 1 then t1=0;             ! later training group

    G7 = ( Grade EQ 7);                 ! create dummy variables for grade level
    G8 = ( Grade EQ 8);
    G9 = ( Grade EQ 9);
    G10 = ( Grade EQ 10);
    G11 = ( Grade EQ 11 );
    G12 = (Grade EQ 12);

    R2 = ( Race EQ 1);                   ! create dummy variables for race/ethnicity
    R3 = ( Race EQ 2);
    R4 = ( Race EQ 3);
    R5 = ( Race EQ 4);
    R6 = ( Race EQ 5);

                                         ! create dummy variables for interaction terms
    SM = S*M;                            ! -- status * gender
    SG7 = S*G7;                          ! -- status * grade
    SG8 = S*G8;
    SG9 = S*G9;
    SG10 = S*G10;
    SG11 = S*G11;
    SG12 = S*G12;
    SR2 = S*R2;                          ! -- status * race/ethnicity
    SR3 = S*R3;
    SR4 = S*R4;
    SR5 = S*R5;
    SR6 = S*R6;

  VARIABLE: NAMES ARE
                  School Male Grade Race Period Status Middle HiLevel
                  Duration Timing TimeInt CPI LPopDur PopDur ST
```

```
                    TimSchTr com Tcom Perc  First NotYet TNA TA
                     AFirst ANotYet ATNA ATA;

USEVARIABLES ARE  CPI Status ST  Male LPopDur    G7 - G12  R2 - R6   t1 t2;
    ! LPopDur = log (population*duration of training period
    ! ST = School*Time sequence

  COUNT is CPI;                      ! means CPI is a count data and follow a
                                     !  Poisson distribution

  CLUSTER = ST;                      !random effect for School * timing sequence

  WITHIN =  Male G7 - G12   R2 - R6   LPopDur ;
                                     ! main effect for Gender, Grade, and
                                     !  Race/ethnicity,
                                     ! offset = log (population * duration)
  BETWEEN = Status;                  ! Status = training status

  Class = c1(2);                     ! two levels

  TRAINING =t1 t2 ;                  ! later training group will have a similar
                                     !distribution as early training group

  Missing is ALL (999);

  USEOBS ARE  ( Period EQ 1 );       ! using period 1 data only


ANALYSIS: TYPE = TWOLEVEL random MIXTURE;

MODEL:
  %WITHIN%
    %OVERALL%

     CPI on  Male G7 - G12 R2 - R6   LPopDur @1;

      %c1#1%                         ! later training group
      CPI on Male G7 - G12  R2 - R6    LPopDur @1;


     %c1#2%                          !early training group
      CPI on  Male   G7 - G12 R2 - R6   LPopDur @1;
```

```
%BETWEEN%
   %OVERALL%
        CPI on S;
        c1#1 on s @0;                    ! class level is independent with training status

         %c1#1%
         CPI on S;                       ! weaken assumption: training effect is
                                           !allowed

         %c1#2%
         CPI on S;                         ! only testing training effect on early training
                                           !group


OUTPUT: TECH1 TECH2;
```

Appendix F: M-plus Code – AS Model 28: Main effects within first training period with weaken condition under As-treat status

```
TITLE:  ITT-Model28_as_median_time;
 DATA: FILE IS J:\data\cpi.numericB.txt;

  Define:

    t1 =1; t2 =1;
    if Tcom eq 0 then t2=0;            ! early training group
    if Tcom eq 1 then t1=0;            ! later training group

    AS = 1;
    if NotYet eq 1 then AS = 0; !create a As-treated training status

    G7 = ( Grade EQ 7);                ! create dummy variables for grade level
    G8 = ( Grade EQ 8);
    G9 = ( Grade EQ 9);
    G10 = ( Grade EQ 10);
    G11 = ( Grade EQ 11 );
    G12 = (Grade EQ 12);

    R2 = ( Race EQ 1);                  ! create dummy variables for race/ethnicity
    R3 = ( Race EQ 2);
    R4 = ( Race EQ 3);
    R5 = ( Race EQ 4);
    R6 = ( Race EQ 5);

                                        ! create dummy variables for interaction terms
    SM = AS *M;                            ! -- AS * gender
    SG7 = AS *G7;                          ! -- AS * grade
    SG8 = AS *G8;
    SG9 = AS *G9;
    SG10 = AS *G10;
    SG11 = AS *G11;
    SG12 = AS *G12;
    SR2 = AS *R2;                          ! -- AS * race/ethnicity
    SR3 = AS *R3;
    SR4 = AS *R4;
    SR5 = AS *R5;
    SR6 = AS *R6;
```

```
VARIABLE: NAMES ARE
                    School Male Grade Race Period Status Middle HiLevel
                    Duration Timing TimeInt CPI LPopDur PopDur ST
                    TimSchTr com Tcom Perc  First NotYet TNA TA
                     AFirst ANotYet ATNA ATA;

USEVARIABLES ARE  CPI Status ST  Male LPopDur    G7 - G12  R2 - R6   t1 t2;
    ! LPopDur = log (population*duration of training period
    ! ST = School*Time sequence

  COUNT is CPI;                          ! means CPI is a count data and follow a
                                         !  Poisson distribution

  CLUSTER = ST;                          !random effect for School * timing sequence

  WITHIN =  Male G7 - G12   R2 - R6   LPopDur ;
                                         ! main effect for Gender, Grade, and
                                         !  Race/ethnicity,
                                         ! offset = log (population * duration)
  BETWEEN = Status;                      ! Status = training status

  Class = c1(2);                         ! two levels

  TRAINING =t1 t2 ;                      ! later training group will have a similar
                                         !distribution as early training group

  Missing is ALL (999);

  USEOBS ARE  ( Period EQ 1 );           ! using period 1 data only


ANALYSIS: TYPE = TWOLEVEL random MIXTURE;

MODEL:
  %WITHIN%
    %OVERALL%

      CPI on  Male G7 - G12 R2 - R6   LPopDur @1;

       %c1#1%                            ! later training group
       CPI on Male G7 - G12  R2 - R6    LPopDur @1;


      %c1#2%                             !early training group
```

```
        CPI on  Male   G7 - G12 R2 - R6   LPopDur @1;



  %BETWEEN%
     %OVERALL%
        CPI on AS;
        c1#1 on AS @0;                  ! class level is independent with training status

         %c1#1%
         CPI on AS @0;                  ! strong assumption: no training effect is
                                          !allowed

         %c1#2%
         CPI on AS;                     ! only testing training effect on early training
                                          !group

OUTPUT: TECH1 TECH2;
```

Appendix G: M-plus Code – ITT Model 88: median time 4period

```
TITLE:  ITT-Model88_median_time_4period;
 DATA: FILE IS J:\data\cpi.numericB.txt;

  Define:

    t1 =1; t2 =1;
    if Tcom eq 0 then t2=0;              ! early training group
    if Tcom eq 1 then t1=0;              ! later training group

    G7 = ( Grade EQ 7);                  ! create dummy variables for grade level
    G8 = ( Grade EQ 8);
    G9 = ( Grade EQ 9);
    G10 = ( Grade EQ 10);
    G11 = ( Grade EQ 11 );
    G12 = (Grade EQ 12);

    R2 = ( Race EQ 1);                   ! create dummy variables for race/ethnicity
    R3 = ( Race EQ 2);
    R4 = ( Race EQ 3);
    R5 = ( Race EQ 4);
    R6 = ( Race EQ 5);

                                         ! create dummy variables for interaction terms
    SM = S*M;                            ! -- status * gender
    SG7 = S*G7;                          ! -- status * grade
    SG8 = S*G8;
    SG9 = S*G9;
    SG10 = S*G10;
    SG11 = S*G11;
    SG12 = S*G12;
    SR2 = S*R2;                          ! -- status * race/ethnicity
    SR3 = S*R3;
    SR4 = S*R4;
    SR5 = S*R5;
    SR6 = S*R6;

  VARIABLE: NAMES ARE
                  School Male Grade Race Period Status Middle HiLevel
                  Duration Timing TimeInt CPI LPopDur PopDur ST
                  TimSchTr com Tcom Perc  First NotYet TNA TA
                   AFirst ANotYet ATNA ATA;
```

```
USEVARIABLES ARE  CPI Status ST  Male LPopDur    G7 - G12  R2 - R6   t1 t2;
      ! LPopDur = log (population*duration of training period
      ! ST = School*Time sequence

   COUNT is CPI;                       ! means CPI is a count data and follow a
                                       !  Poisson distribution

   CLUSTER = ST;                       !random effect for School * timing sequence

   WITHIN =  Male G7 - G12   R2 - R6   LPopDur ;
                                       ! main effect for Gender, Grade, and
                                       !  Race/ethnicity,
                                       ! offset = log (population * duration)
  BETWEEN = Status;                    ! Status = training status

  Class = c1(2);                       ! two levels

  TRAINING =t1 t2 ;                    ! later training group will have a similar
                                       !distribution as early training group

  Missing is ALL (999);


ANALYSIS: TYPE = TWOLEVEL random MIXTURE;

MODEL:
  %WITHIN%
    %OVERALL%

      CPI on  Male G7 - G12 R2 - R6   LPopDur @1;

       %c1#1%                          ! later training group
       CPI on Male G7 - G12  R2 - R6    LPopDur @1;


      %c1#2%                           !early training group
       CPI on  Male   G7 - G12 R2 - R6   LPopDur @1;



%BETWEEN%
   %OVERALL%
        CPI on S;
```

```
c1#1 on s @0;                    ! class level is independent with training status

 %c1#1%
 CPI on S @0;                     ! strong assumption: no training effect is
                                  !allowed

 %c1#2%
 CPI on S;                        ! only testing training effect on early training
                                  !group

OUTPUT: TECH1 TECH2;
```

Appendix H: M-plus Code – ITT Model 132: AS main effects 4period ms

```
TITLE:  ITT-Model132_AS_main_effects_4period_ms;
 DATA: FILE IS J:\data\cpi.numericB.txt;

  Define:

    t1 =1; t2 =1;
    if Tcom eq 0 then t2=0;                 ! early training group
    if Tcom eq 1 then t1=0;                 ! later training group

    G7 = ( Grade EQ 7);                     ! create dummy variables for grade level
    G8 = ( Grade EQ 8);
    G9 = ( Grade EQ 9);
    G10 = ( Grade EQ 10);
    G11 = ( Grade EQ 11 );
    G12 = (Grade EQ 12);

    R2 = ( Race EQ 1);                      ! create dummy variables for race/ethnicity
    R3 = ( Race EQ 2);
    R4 = ( Race EQ 3);
    R5 = ( Race EQ 4);
    R6 = ( Race EQ 5);

                                            ! create dummy variables for interaction terms
    SM = S*M;                               ! -- status * gender
    SG7 = S*G7;                             ! -- status * grade
    SG8 = S*G8;
    SG9 = S*G9;
    SG10 = S*G10;
    SG11 = S*G11;
    SG12 = S*G12;
    SR2 = S*R2;                             ! -- status * race/ethnicity
    SR3 = S*R3;
    SR4 = S*R4;
    SR5 = S*R5;
    SR6 = S*R6;

  VARIABLE: NAMES ARE
                    School Male Grade Race Period Status Middle HiLevel
                    Duration Timing TimeInt CPI LPopDur PopDur ST
                    TimSchTr com Tcom Perc  First NotYet TNA TA
                     AFirst ANotYet ATNA ATA;
```

```
USEVARIABLES ARE  CPI Status ST  Male LPopDur    G7 – G8  R3 - R6   t1 t2;
       ! LPopDur = log (population*duration of training period
       ! ST = School*Time sequence

   COUNT is CPI;                          ! means CPI is a count data and follow a
                                          !  Poisson distribution

   CLUSTER = ST;                          !random effect for School * timing sequence

   WITHIN =  Male G7 – G8   R3 - R6   LPopDur ;
                                          ! main effect for Gender, Grade, and
                                          !  Race/ethnicity,
                                          ! offset = log (population * duration)
  BETWEEN = Status;                       ! Status = training status

  Class = c1(2);                          ! two levels

  TRAINING =t1 t2 ;                       ! later training group will have a similar
                                          !distribution as early training group

  Missing is ALL (999);

  USEOBS ARE  (Middle EQ 1 AND RACE NE 1);

ANALYSIS: TYPE = TWOLEVEL random MIXTURE;

MODEL:
  %WITHIN%
    %OVERALL%

      CPI on  Male G7 – G8   R3 - R6   LPopDur @1;

       %c1#1%                             ! later training group
       CPI on Male G7 – G8   R3 - R6   LPopDur @1;


      %c1#2%                              !early training group
       CPI on  Male   G7 – G8   R3 - R6   LPopDur @1;



%BETWEEN%
   %OVERALL%
```

```
CPI on S;
c1#1 on s @0;                    ! class level is independent with training status

 %c1#1%
 CPI on S @0;                      ! strong assumption: no training effect is
                                   !allowed

 %c1#2%
 CPI on S;                         ! only testing training effect on early training
                                   !group

OUTPUT: TECH1 TECH2;
```

Appendix I: M-plus Code – ITT Model 176: AS main effects 4period hs

```
TITLE:  ITT-Model176_AS_main_effects_4period_hs;
 DATA: FILE IS J:\data\cpi.numericB.txt;

  Define:

    t1 =1; t2 =1;
    if Tcom eq 0 then t2=0;          ! early training group
    if Tcom eq 1 then t1=0;          ! later training group

    G7 = ( Grade EQ 7);              ! create dummy variables for grade level
    G8 = ( Grade EQ 8);
    G9 = ( Grade EQ 9);
    G10 = ( Grade EQ 10);
    G11 = ( Grade EQ 11 );
    G12 = (Grade EQ 12);

    R2 = ( Race EQ 1);                ! create dummy variables for race/ethnicity
    R3 = ( Race EQ 2);
    R4 = ( Race EQ 3);
    R5 = ( Race EQ 4);
    R6 = ( Race EQ 5);

                                      ! create dummy variables for interaction terms
    SM = S*M;                         ! -- status * gender
    SG7 = S*G7;                       ! -- status * grade
    SG8 = S*G8;
    SG9 = S*G9;
    SG10 = S*G10;
    SG11 = S*G11;
    SG12 = S*G12;
    SR2 = S*R2;                       ! -- status * race/ethnicity
    SR3 = S*R3;
    SR4 = S*R4;
    SR5 = S*R5;
    SR6 = S*R6;

  VARIABLE: NAMES ARE
                  School Male Grade Race Period Status Middle HiLevel
                  Duration Timing TimeInt CPI LPopDur PopDur ST
                  TimSchTr com Tcom Perc  First NotYet TNA TA
                   AFirst ANotYet ATNA ATA;
```

```
 USEVARIABLES ARE  CPI Status ST  Male LPopDur    G10 – G12  R2 - R6   t1
t2;
       ! LPopDur = log (population*duration of training period
       ! ST = School*Time sequence

   COUNT is CPI;                          ! means CPI is a count data and follow a
                                          !  Poisson distribution

   CLUSTER = ST;                          !random effect for School * timing sequence

   WITHIN =  Male G10 – G12  R2 - R6 LPopDur ;
                                          ! main effect for Gender, Grade, and
                                          !  Race/ethnicity,
                                          ! offset = log (population * duration)
  BETWEEN = Status;                       ! Status = training status

  Class = c1(2);                          ! two levels

  TRAINING =t1 t2 ;                       ! later training group will have a similar
                                          !distribution as early training group

  Missing is ALL (999);

  USEOBS ARE  (Middle EQ 0);

 ANALYSIS: TYPE = TWOLEVEL random MIXTURE;

 MODEL:
  %WITHIN%
    %OVERALL%

      CPI on  Male G10 – G12  R2 - R6 LPopDur @1;

       %c1#1%                             ! later training group
       CPI on Male G10 – G12  R2 - R6 LPopDur @1;


      %c1#2%                              !early training group
       CPI on  Male   G10 – G12  R2 - R6 LPopDur @1;



 %BETWEEN%
```

```
%OVERALL%
  CPI on S;
  c1#1 on s @0;                ! class level is independent with training status

   %c1#1%
   CPI on S @0;                ! strong assumption: no training effect is
                               !allowed

   %c1#2%
   CPI on S;                   ! only testing training effect on early training
                               !group

OUTPUT: TECH1 TECH2;
```

About the Author

Jing Guo originally came from Beijing, P. R. China. She received a Bachelor's Degree in Applied Electronic Technology from Beijing Union University, Beijing, China and a M.S. in Statistics from University of Georgia, Athens, U.S. She entered the Ph.D. program at University of South Florida in 2004.

While in the Ph.D. program at the University of South Florida, Ms. Guo also work as a research staff at the Prevention Science and Methodology Group and a statistical data analyst at Policy and Services Research Data Center, Department of Mental Health Law & Policy, Louis de la Parte Florida Mental Health Institute, University of South Florida. She, as a biostatistician, has participated in many clinical trials and administrative studies, and provided critical statistical support.