THE APPLICATION OF PROPENSITY SCORE ESTIMATES IN HIERARCHICAL LINEAR MODELS FOR CAUSAL INFERENCE

by

PATRICIA A. ECKARDT

A dissertation submitted to the Graduate Faculty in Educational Psychology in partial fulfillment
of the requirements for the degree of Doctor of Philosophy, The City University of New York

2010

This manuscript has been read and accepted for the
Graduate Faculty in Educational Psychology in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

David Rindskopf

Date
Chair of Examining Committee

Mario Kelly

Date
Executive Officer

David M. Rindskopf, Ph.D.

Irvin Schonfeld, Ph.D.

Jay Verkuilen, Ph.D.
Supervision Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

THE APPLICATION OF PROPENSITY SCORE ESTIMATES IN HIERARCHICAL LINEAR MODELS FOR CAUSAL INFERENCE

by

Patricia A. Eckardt

Advisor: David M. Rindskopf, Ph.D.

This research investigated a causal estimate of the impact of zero tolerance policy adoption on individual students' cognitive outcomes by modeling multilevel propensity score estimates within a potential outcomes framework. This estimate was obtained using a large, nationally representative non-experimental sample. Proponents of zero tolerance policy assert that the mandatory expulsion of students for listed offenses leads to a learning environment that supports cognitive growth for the remaining students. Results indicated that zero tolerance policies do not have the desired positive effect on not-at-risk students' cognitive outcomes.

Acknowledgments

Thank you to David Rindskopf, my dissertation chair and mentor for all he has taught me these past four years. All I do going forward in my research is grounded in the academic foundation that you have given me, thank you. Thank you to Jay Verkuilen for his expertise and insight throughout this process. Thank you to Irvin Schonfeld for introducing me to propensity score estimates my first year of graduate studies, and for his attention to detail. Jennifer Hill and Michael Sobel, thank you for exposing me to Rubin's causal model framework. The thoughtful insight and comments of Howard Everson and Mary Clare Lennon have added to this research. I thank you both for sharing your individual expertise with me.

To my parents, Philip and Ann Suydam, thank you for always supporting my decisions, no matter where they took me. You are wonderful parents, thank you for a strong foundation. My grandparents and ancestors, after your hard work and struggle, we have arrived, thank you. My siblings: Joan, Eileen, John, and Diane, we have done well. My children, Dr Ann Eckardt and Ms Sarah Eckardt, thank you for being the greatest gifts of my life. All I do is better because you are here. You are my treasures, my keys to heaven. Ronald Eckardt, my husband and best friend, thank you for sharing this life with me, and for always inspiring me to grow. It is through your love for me that I have been able to do any and all things that I try. I only hope that I give to you all that you give me, words will never express the depths of love and gratitude that I have for you. Finally thank you to God, my accomplishments are yours, as you carry me through the stages of my journey. All that we do here is for your glory Lord.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

In 1989 school districts in California, New York, and Kentucky decided to institute “zero tolerance” policies, and began mandated expulsion of students (defined as suspension for more than 15 days) for gang-related activity, such as drug possession and sales, and weapons possession. By 1993, zero tolerance policies had been adopted across many states in the U.S., and were often expansive in translation to now include smoking and school disruption among the list of zero tolerance offenses (Skiba & Knesting, 2001).

Zero tolerance became national school policy with the signing of the Gun- Free Schools Act of 1994 by the Clinton Administration. The law mandates a one- calendar year expulsion for possession of a firearm, and the referral of law-violating students to the criminal justice system. The National Center on Educational Statistics reported that between 79% and 94% of schools nationally implement zero tolerance for at least one serious infraction (Heaviside, Rowand, Williams, & Farris, 1998).

Questions have been raised regarding the effect of zero tolerance policy on students’ cognitive outcomes. These questions of policy impact concern non-expelled students as well as expelled students in all schools (Ewing, 2000, Skiba & Peterson, 1999).

Chapter II provides a description of the major methodological challenges in studying zero tolerance policy effects, and foundational definitions of the methodologies used (causal inference in potential outcomes framework, propensity score estimates, and hierarchical linear modeling). In Chapter III the sample is introduced, an overview of the analysis and the key causal questions of interest are given, and framework for modeling causal estimates is described. Chapter IV provides results of modeling and analyses. Chapter V discusses implications of analyses, and suggests further research.

CHAPTER II: REVIEW OF THE LITERATURE

A. Zero Tolerance Policy and Major Methodological Challenges

When measured in the policy literature, zero tolerance treatment variables of interest are operationally defined as expulsion from school for serious violations, such as drug possession and sale, and weapons possession and use. Some school administrators consider the broader inclusion of violations such as any fighting, violence, or tobacco and alcohol use or possession. The analysis conducted will consider the narrowest and original definition of zero tolerance policy, which is limited to mandatory expulsion for student drugs or weapons infractions.

The limitation of a simple analysis measuring student outcomes after expulsion is that student expulsion occurs at the individual level and is not the treatment variable of interest in zero tolerance policy research. Students are expelled in both zero tolerance and non-zero tolerance policy schools alike. Zero tolerance policy is adopted at the school-level. Students attend a zero tolerance school or not. When examining the treatment variable at the individual level, the researcher is answering questions regarding the effects of expulsion, not the effects of a zero tolerance policy adoption. For zero tolerance policy causal effects to be estimated, the adoption of policy at the school-level needs to be included in the model.

The zero tolerance outcome variable of interest here is students' cognitive performance. The disciplinary strategy that has come to be known as zero tolerance is based on the belief that the removal of disruptive students (even if for a limited time) is not only effective but to a certain extent necessary to preserve the integrity of the learning environment (Ewing, 2000). Thus, it makes sense to explore the causal relationship between school discipline approaches and student cognitive outcomes. This argument is based on the belief that a student's cognitive outcome can be affected by the expulsion of other students, and what happens to individual students could be a mediator from that policy. A consideration of this belief will be modeled in the proposed analysis.

The analyses in the zero tolerance policy literature are limited. Multiple studies (Rausch & Skiba, 2004; Skiba & Peterson, 1999; Skiba, et al., 2003) reported descriptive statistics from observational data obtained from federal and state reports and small to moderate scale observational studies. The most recent independent published studies continue to examine the question of zero tolerance without the most appropriate methodology to answer causal questions. Feda (2008) conducted an observational study that analyzed teacher assault in relation to zero tolerance security measures and implementation, but did not account for non randomness of treatment assignment or address the hierarchical nature of the policy implementation. Farrell (2005) applied hierarchical analyses to zero tolerance survey data. However, Farrell (2005) did not use the natural grouping of students within schools as the nested structure, but rather constructed artificial groupings by behavior infractions and levels of connectedness, creating a latent variable analysis rather than modeling the non- independence of the student within-schools observations.

The position taken here is that there is a lack of appropriate modeling in the current literature of the non-experimental multi-level data used to examine zero tolerance policy.

Previous research has been an attempt to measure both a treatment (expelled or not expelled) and the effects of these policies (cognitive outcomes) at the individual student-level, thereby not accounting for the actual treatment assignment of interest, school adoption of zero tolerance policy. If the treatment variable of interest (zero tolerance policy adoption) is not measured at the level of assignment (school-level) accurate causal estimates will not be made. Furthermore, if non random treatment assignment at the group and individual level is not considered in the assumptions and modeling, the causal estimates produced may be biased. Lastly, the potential influence of peer treatment on individual students' outcomes needs to be considered in the model.

Appropriate methodologies to answer the causal inference questions about zero tolerance educational policies, and key terms and definitions associated with the methodologies used in this analysis, are outlined in the next subsections.

B. Propensity Score Estimates

The lack of randomization of school policy adoption and of students' assignment to specific schools are inferential limitations of educational research in observational settings. In an experiment, the assignment of treatments to subjects is controlled by the experimenter, who ensures that subjects receiving different treatments are comparable. In an observational study, this control is absent due to ethical, political, social or logistic opposition to randomization of treatment assignment. In the latter case, the investigator does not control the assignment of treatments and cannot ensure that similar subjects receive different treatments (Rosenbaum, 2002). Love (2004) describes the situation as wanting to compare groups who looked similar before they were exposed to treatments. One commonly used method of controlling for systematic differences involves grouping units into subclasses based on observed characteristics, and then directly comparing only treated and control units who fall within the same subclass. The use of the stratification methodology is limited. When the number of observed covariates increases, the number of subclasses grows exponentially. Consider the case with only two categories per covariate, so that there are 2^p subclasses for p covariates. If p is moderately large, some subclasses will contain no units. For example, if there are 10 such dichotomous covariates used in an analysis, then there are $2^{10} = 1,024$ subclasses of those covariates. Many subclasses will contain either treated or control units, but not both, making it impossible to form directly adjusted estimates for the entire population (Rosenbaum & Rubin, 1984).

One approach to solving this problem involves the use of propensity scores. The propensity score is defined as the conditional probability of assignment to the treatment group given a set of observed pretreatment variables (Rosenbaum & Rubin, 1984). The propensity score takes multiple covariates and creates a single covariate, the propensity score, for each unit in the study. The propensity score summarizes the information required to balance the distribution of the covariates. Specifically, subclasses formed from the scalar propensity score will balance all p covariates. For a binary treatment measure (such as adoption of zero tolerance or not), the joint distribution of all of the observed pretreatment covariates is balanced between the two treatment groups given the propensity score. However, the purpose of the propensity score is not to balance covariates, it is to assign probability to control and treatment units. The propensity score is constructed based on the probability of that particular unit being assigned to the treatment condition in the study, given all of its covariates (Gelman & Hill 2006).

When propensity score estimation is used with observational data, the only assumption necessary to interpret the estimates causally is the assumption of ignorability of treatment assignment. Ignorability of treatment assignment can be assumed if all the covariates that affect the treatment assignment have been accounted for, so that there are no unobserved covariates that will affect the estimates. If ignorability holds, one can obtain unbiased treatment effect estimates. If the assignment of schools to a zero tolerance policy or non-zero tolerance policy is ignorable (given the observed school-level pre-treatment covariates), then as a function of the propensity score, an unbiased estimate of a school propensity for selecting a zero tolerance policy can be made. In the same way, if the expulsion or non expulsion of students under zero tolerance school policy or non-zero tolerance school policy is ignorable given the observed student-level and school-level pretreatment covariates, then student propensities of being retained within a school can be estimated with the constructed school-level and student-level propensity scores.

In this analysis of non experimental data, a school's adoption of a zero tolerance policy is a school- specific process that depends on pretreatment covariates; similarly the decision to expel a student depends on pretreatment covariates at the student and school-level. In zero tolerance research theory, a school's adoption of a zero tolerance policy is associated mainly with school characteristics, such as school policies and procedures concerning discipline, administrator philosophies and behaviors, and staff beliefs and behaviors. Likewise, a child's probability of being expelled is likely to be associated with his or her behavioral history, demographic characteristics, cognitive functioning, emotional and social development, and previous home, community, and school disciplinary experiences (Skiba, et al., 2003, Skiba & Knesting, 2001).

Using the joint distributions of the school-level covariates and aggregated student-level covariates that influence adoption of zero tolerance policy, and the student-level covariates that influence a child's probability of being expelled, to estimate propensity scores for assignment to treatment will theoretically allow one to make causal inference about the average effects of these two treatments, if ignorability can be assumed. The naturally nested structure of the data was also considered in the modeling.

C. Hierarchical Linear Modeling

Hierarchical linear modeling (HLM), also known as multi-level modeling, can be thought of in two equivalent ways. One can think of multilevel modeling as a generalization of linear regression where the intercepts, and possibly slopes, are allowed to vary by group. Equivalently, one can think of HLM as a regression that includes a categorical input variable that represents group membership (Gelman & Hill, 2006). The interest in multilevel models in educational research comes from the fact that independent variables are often measured at a higher level of aggregation than is the outcome variable of interest. (Garner and Raudenbush, 1991).

Also, because these models explicitly consider the clustering of individuals within the higher level school units, they do not violate the assumption of independence of observations as ordinary least squares analysis does if used to analyze hierarchical data (Raudenbush and Bryk, 2002). The assumptions of hierarchical linear modeling include linearity (functional forms are linear at each level); normality (level-1 residuals are normally distributed and level-2 random effects have a multivariate normal distribution); homoscedasticity (level-1 residual variance is constant) and lastly independence (level-1 residuals and level-2 residuals are uncorrelated and the observations at highest level are independent of each other). The assumptions of hierarchical linear modeling are plausible with much educational data (such as samples of students grouped within classrooms), making hierarchical linear modeling a commonly used procedure among educational researchers (Hong & Raudenbush, 2006; Hong & Yu 2008; Raudenbush, 1993).

To answer zero tolerance policy causal questions, in addition to considering the hierarchical structure of the data and the cross level assignment mechanism, the possible influence of peer treatment on individual students’ potential outcomes needs to be considered.

D. Causal Modeling in a Potential Outcomes Framework

Rubin’s (1974) counterfactual causal model approaches causal modeling with a fundamental assumption of potential outcomes (Angrist & Pischke, 2008; Rosenbaum, 2002; Rubin, 1974). This theory asserts that there is one potential outcome associated with each treatment a subject may receive. For the following example, Y denotes an individual subject’s potential outcome, and Z denotes treatment assignment. If there are two alternate treatments (a binary treatment condition), then each subject has two potential outcomes, represented commonly in the potential outcomes literature as Y (0) and Y (1). Specifically, Y (0) is the outcome for a subject who does not receive the treatment (i.e. Z=0); while Y (1) is the outcome if the subject does receive the treatment (i.e. Z=1). However, one never gets to observe both of these potential outcomes, as the subject only can receive one treatment at one time (Rubin, 2004). Therefore one can never estimate the causal effect Y (1) – Y (0) for an individual, but only an average effect for a group of people.

In a randomized experiment, one can use the observed outcomes of the subjects in the sample to estimate an average treatment effect conditioned on treatment assigned. An unbiased estimate the average treatment effect $E[Y_i(1)-Y_i(0)] = E[Y_i(1)] - E[Y_i(0)]$ within an observed population sample is the difference in observed outcomes in the two distinct treatment groups. Therefore in a randomized experiment, if Z stands for treatment assignment, then the average treatment effect $E[Y_i(1)] - E[Y_i(0)]$, can be estimated as $E[Y_i(1)] = E[Y_i(1)|Z_i=1]$ and $E[Y_i(0)] = E[Y_i(0)|Z_i=0]$. One can then obtain an unbiased estimate of the average treatment effect by estimating $E[Y_i(1)]$ with $\bar{Y}_{z=1}$ and $E[Y_i(0)]$ with $\bar{Y}_{z=0}$.

The assumptions of the potential outcomes model are ignorability of treatment assignment and stable unit treatment value. Treatment assignment is ignorable when the potential outcomes are independent of the treatment variable (Rubin, 2004). The stable unit treatment value (SUTVA) is the assumption that there is only one potential outcome for each school associated with the policy it adopted, with no effect on an individual school’s potential outcome due to the policy the other schools adopted, or how that school’s specific policy was decided (Rubin, 1986). Given that the assumptions for Rubin’s counterfactual model of potential outcomes hold, then (within a one-level model) the definition of a student- level effect is simply the difference between the outcome of the student if the student was expelled and the outcome if the student was not expelled. Rubin (1990) cautioned that SUTVA becomes problematic when educational treatments are given to children who interact with one another.

In particular, zero tolerance policies are grounded in the belief that peers' treatment assignment will influence other students' potential outcomes. If students behaving badly receive swift punishment by removal from the classroom, a less disruptive environment for learning should be created, which in turn should promote positive cognitive outcomes for the remaining students.

CHAPTER III: METHOD

A. Sample Used in Analysis

The National Longitudinal Study of Adolescent Health (Add Health) is a longitudinal study of a nationally representative sample of adolescents in grades 7-12 in the United States that began during the 1994-95 school year. The Add Health cohort has been followed into young adulthood with four in-home interviews, the most recent in 2007-2008, when the sample was aged 24-32. In Wave I, Stage I the individual adolescent subjects were randomly selected from a stratified sample of all high schools in the US that met the eligibility criteria of at least 30 students in the school and 12th grade in the school; and from feeder schools (middle or junior high schools) that had a grade 7. The subjects consisted of 90,118 adolescent students with an accompanying survey of 144 school administrators regarding school policies, including school security and disciplinary procedures.

The Wave I, Stage II sample was a sub-sample (n=20,745) from the original core sample of 90,118 students. This sub-sample was selected to complete a more detailed in-home survey of students. Parents of the sub-sample were also surveyed at this visit. However, the parent surveys were not completed by all parents in the sub-sample, resulting in a parental subset (n= 17,670). This sub-sample of data was collected during the same time period as the Wave I Stage I data.

The treatment assignment variable of interest was specifically assignment of student to a zero tolerance school in the Wave I data. The survey data from the Wave I School Administrator questionnaire provided the information regarding individual school policy on zero tolerance. Questions in Wave I of the data ask the school administrator “In your school, what happens to a student who is caught with a first time occurrence of possessing an illegal drug at school, using an illegal drug at school, possessing a weapon?” These same questions then also have a second occurrence follow up question. Zero tolerance policy schools are defined as those that expel or have mandatory out of school suspension of over fifteen days students for first time infractions of these types (Skiba & Knesting, 2001). Using this definition (at the school-level), if a school reported a policy of expulsion for any of the above listed first-time infractions, it was defined as a zero tolerance policy school.

The measures in the Add Health dataset which have theoretically been identified as predictors of school adoption of zero tolerance policies were used as school-level covariates. The school-level covariates from Wave I used in the models are school type, school size, teacher characteristics, daily attendance, parent teacher organization involvement and characteristics, drop out rates, other school policies regarding discipline such as dress codes, school security measures, neighborhood and community characteristics, administrator and teacher beliefs regarding discipline and student academic achievement (the full list of school-level covariates is in APPENDIX B).

All covariates that effect both the treatment and the outcome variable of interest need to be measured and included in a causal analysis for the assumption of ignorability to hold (Gelman & Hill, 2006). Student-level covariates from Wave I included in the analysis were delinquency behavior, drug and alcohol use, self efficacy beliefs, victim of a crime, peer and friend behaviors, grades and success in school measures, history of expulsion or skipping a grade, student perception of teacher and peer relationships, home environment, relationship with parents, parent involvement in school, socioeconomic status of family, living arrangement at home, health history and problems, and perception of school safety (the full list of student-level covariates is in APPENDIX C) .

The Wave III sample was an in-home administered sub-sample of the Wave I sample. This sub-sample (n= 15,197) was then between 18 and 26 years of age. The current research estimated the causal effect of zero tolerance policies on students' cognitive outcomes. The cognitive outcome variable of interest is students' Add Health Picture Vocabulary Test (AHPVT) scores from Wave III. The AHPVT is an abbreviated version of the Peabody Picture Vocabulary Test (designed as a test of receptive vocabulary achievement and verbal ability) that was revised with age-standardized scores for adolescent respondents.

B. Overview of Analysis and Causal Questions of Interest

Hong and Raudenbush (2006) and Hong and Yu (2008) have attempted to estimate causal effects for non- experimental and nested educational data by applying propensity score estimates at one or multiple levels. Hong and Raudenbush (2006) have also modeled possible effects of peer treatment assignment on individual potential outcomes through a relaxation of SUTVA.

The policy issue addressed by Hong and Raudenbush (2006) was the effect of kindergarten retention policy on student cognitive outcomes. Hong and Raudenbush used multilevel propensity score stratification to approximate a two-stage experiment. First they blocked intact schools on covariates, and then, the second stage of the experiment had "at risk" students within-schools blocked on covariates. Theirs is a different policy question and population of interest than the zero tolerance policy questions. However, the methodological challenges for causal inference are comparable. Hong and Raudenbush's (2006) modeling of propensity score estimates within multilevel models was the guide for the methodologies proposed to answer the question of zero tolerance policy effects on student cognitive outcomes.

The causal modeling allowed school assignment and peer treatments to affect potential outcomes within a framework that accounted for the natural nesting of the data. The modeling included propensity score estimates at the school and student levels to remove the bias in the estimation of treatment effects on students within the schools.

Using the Add Health data sample, an estimate of the difference in causal effect of attending a zero versus a non-zero tolerance policy school for students without a risk of expulsion was made. The causal question of interest in the zero tolerance literature that was answered with this estimate is:

> What are the effects of zero tolerance policies treatment on the non-expelled student in comparison to the exposure of non-expelled students to a non-zero tolerance policy school regarding cognitive outcomes?

Put simplistically, does zero tolerance school policy make “good kids” cognitively better off?

C. Theoretical Framework for Models

Potential outcomes at student-level

For a binary assignment (expelled or not expelled) let $z_i = 1$ if student *i* is expelled and $z_i = 0$ if student is not expelled. Then, given *N* units overall in a school, there is a 1 x *N* vector of possible treatment assignments $\mathbf{z} = (z_1, z_2, \ldots, z_N) = (z_i, \mathbf{z}_{\_i})$, where $\mathbf{z}_{\_i}$ is the 1 x $(N - 1)$ vector of treatment assignments with z_i removed, for $z_i \in \{0,1\}, i = 1,\ldots, N$. Under these set conditions, subject *i* has 2^N potential outcomes, $Y_i(\mathbf{z})$, corresponding to all possible treatment assignments of the *N* subjects.

A contrast between any two of the 2^N potential outcomes for a given subject is a causal effect. Hong and Raudenbush (2006) note that SUTVA is a special case where $Y_i(\mathbf{z}) \equiv Y_i(z_i, \mathbf{z}_{\_i}) = Y_i(z_i)$, demonstrating that an individual student’s outcome, given all other student treatments, remains the same under SUTVA. In contrast, if SUTVA is not assumed, a change in treatment assignment of any subject changes the potential outcome of any other subject. Further constraints need to be imposed on the potential outcomes. The proposed models attempt to summarize the available school treatment assignment information in a way that will be useful for average causal estimates. This was done by modeling the impact of z (treatment assignment) on a single subject's potential outcome as operating through z_i (individual's treatment assignment) as well as through a simple function of $\mathbf{z}_i$, which will be denoted as $v(\mathbf{z})$, the treatment assignments of all other students within an individual's school, except the individual's assignment. This is represented as

$$\hat{Y}_i(\mathbf{z}) \equiv \hat{Y}_i(z_i,\ \mathbf{z}_{\_i}) = \hat{Y}_i[z_i,\ v(\mathbf{z})] \qquad (1)$$

" A generic estimand of causal effect is $E\{Y[z, v(\mathbf{z})] - Y[z', v(\mathbf{z'})]\}$, where z and z' are alternative treatment assignments for an individual and **z** and **z'** are alternative treatment assignments for all individuals within that school" (Hong & Raudenbush, 2006, p. 902). For the purposes of zero tolerance policy, let $v(\mathbf{z}) = 1$ if a school is a zero tolerance policy school and $v(\mathbf{z}) = 0$ if this is not the case.

Now, $Y(z, v(\mathbf{z}))$ can take on four possible values— $Y(1, 1)$, $Y(0, 1)$, $Y(1, 0)$, and $Y(0, 0)$. Two of these four possible values of potential outcomes provide the causal estimands which correspond to the question of interest regarding zero tolerance policy:

$E[Y(0, 1) - Y(0, 0)]$ is the average causal effect of attending a zero tolerance school versus a non-zero tolerance policy school if a student is not expelled. This is the causal effect of interest estimated in this analysis.

Conditioning on school assignment

To account for the grouping of the data in this analysis (of students grouped within schools), school assignment needs to be introduced to the proposed model. If school assignment had been at random, then $\mathbf{s} = (s_1, s_2 \ldots s_N)$, where the possible values of s_i could be the school identification numbers $j=1\ldots J$. This would then construct the potential student outcome for student i to be $Y_i\,[z_i, v\,(\mathbf{z}), \mathbf{s}]$. This addition of school assignment would modify the previous causal estimand to the form

$$E\{Y\,[z, v\,(\mathbf{z}), \mathbf{s}] - Y\,[z', v\,(\mathbf{z}'), \mathbf{s'}]\} \tag{2}$$

In which the average causal effect of treatment assignments z, v (z), and school assignment s is compared to treatment assignments *z', v (z'),* and school assignment *s'*.

However, the above estimand does not fit this data, as there is no random assignment of students to these schools. The model must account for non random assignment, while modeling the causal effects one wishes to estimate. The model will condition on the schools in the dataset. Therefore, the estimand of interest will be conditional on current school assignments, adding to the previous estimand of Y as

$$E\{Y\,[z, v(\mathbf{z}), \mathbf{s}^*] - Y[z', v(\mathbf{z'}), \mathbf{s}^*]|\mathbf{S} = \mathbf{s}^*\} \tag{3}$$

where $\mathbf{s}^*$ is the vector of school assignments observed in the sample. The rationale for conditioning on current school membership instead of on a hypothetical universe of schools is that the average causal effects of interest are of student membership within their current schools. The limitations of non randomness of assignment on causal inference will be considered with a propensity score estimate at school-level.

This leads to the discussion of the last two assumptions that need to be imposed to complete the theoretical framework for unbiased causal estimates in the analysis. These assumptions are that there is no interference between school units (SUTVA at the school-level of the model), and that treatment assignment at school-level (the adoption of a zero tolerance policy) is ignorable given the observed covariates.

SUTVA at school-level

The assumption of SUTVA at the school-level is plausible as schools in the proposed nationally sampled data set are located in distinct communities with different geographic conditions, various combinations of socioeconomic levels, governmental resources, political affiliations in the community profiles, and span the continental United States.

Here we assume that the identities and treatment assignments of students attending other schools are uninformative about student i's outcomes (SUTVA assumption at the school-level). We can now group $\mathbf{z}$ and $\mathbf{s}$ by school identification number, so that $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_J)$ and $\mathbf{s} = (\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_J)$, where $\mathbf{z}_j = (z_{1j}, z_{2j}, \ldots, z_{n_j j})$ is the 1 x n_j vector of treatment assignments of students assigned to school j and $\mathbf{s}_j = (j, j, \ldots, j)$ is the corresponding $1 \times n_j$ vector of school assignments. The relaxed form of SUTVA (Hong & Raudenbush, 2006) allows no interference between schools, so our causal estimand is

$$Yi[z_i,\ v(\mathbf{z}),\ \mathbf{s}^*] \ = \ Y_{ij}[z_{ij},\ v(\mathbf{z}_j),\ \mathbf{s}_j^*] \ \equiv \ Y_{ij}[z_{ij},\ v(\mathbf{z}_j)]. \tag{4}$$

Conditioning on covariates for ignorability of treatment assignment

In the Add Health dataset, zero tolerance policies were not assigned at random to schools, and students within schools were not assigned at random to be expelled. A propensity score is the conditioning on covariates an observed probability for receiving the treatment. As such, if there is no hidden bias, then the conditional distribution of treatment assignment is uniform, and the statistical methods used to analyze a randomized experiment may be used in the analysis (Rosenbaum, 2002).

Let $\mathbf{X}$ be a vector of student-level covariates and let $\mathbf{W}$ be a vector of school-level covariates. $\mathbf{W}$ includes the school-level aggregates of student-level covariates such as the demographic makeup of students attending a school. Causal inferences are possible if treatment assignments are ignorable within levels of covariates (where **X** = **x** and **W** = **w**) for schools such that

$$E[Y(z,\ v) \mid Z = z, V = v, \mathbf{X} = \mathbf{x}, \mathbf{W} = \mathbf{w}] = E[Y(z,\ v) \mid \mathbf{X} = \mathbf{x}, \mathbf{W} = \mathbf{w}]. \qquad (5)$$

in which case the conditional average causal effect (CACE)

$E[Y(z, v)|\mathrm{X} = \mathrm{x}, \mathbf{W} = \mathbf{w}] - E[Y(z', v')|\mathbf{X} = \mathbf{x}, \mathbf{W} = \mathbf{w}]$ is equivalent to the observed data estimand,

$E[Y(z, v)|Z = z, V = v, \mathbf{X} = \mathbf{x}, \mathbf{W} = \mathbf{w}] - E[Y(z', v')|Z = z', V = v', \mathbf{X} = \mathbf{x}, \mathbf{W} = \mathbf{w}]$.

CACE, being estimable from observed data, allows causal inference regarding the difference between students attending a zero tolerance school versus a non-zero tolerance school (Hong & Raudenbush, 2006).

D. Defining Causal Estimates of Interest

In the Add Health data analysis, the potential outcomes for an individual student i attending a school j, denoted by $Y_{ij}(\mathrm{z}_{\mathrm{ij}}, v_{\mathrm{j}})$, could hypothetically take on four values $Y_{ij}(1, 1)$, $Y_{ij}(0, 1)$, $Y_{ij}(1, 0)$, and $Y_{ij}(0, 0)$. For example $Y_{ij}(1, 1)$ represents the potential outcome of a student i who is expelled from a zero tolerance school, and $Y_{ij}(1, 0)$ is the potential outcome of a student i who is expelled from a non-zero tolerance policy school.

If one considers these potential outcomes carefully, in the context of the students in the Add Health dataset, it is clear that not all of the four potential outcomes listed above are defined for some of the students. A student who would likely be expelled in a zero tolerance policy school may not be at such a risk in a non-zero tolerance policy school, and another student may never be expelled (even within a school with a zero tolerance policy). To address this natural distinction into sub groupings of students within the data set, let q_1 denote a student's probability of being expelled under a zero tolerance policy, and let q_0 denote the student's probability of being expelled under a non-zero tolerance policy. Monotonicity, described by Angrist (1990), is applied here as the assumption that one's probability of being expelled in a zero tolerance policy school is always greater than or equal to that within a non-zero tolerance policy school. As monotonicity is plausible in this case, three subpopulations of students of interest can be identified within the data set

(A) Students at risk of being expelled in a non-zero tolerance school, as well as in a zero tolerance policy school ($q_1 \geq q_0 > 0$)

(B) Students at risk of being expelled only in a zero tolerance policy school ($q_1 > 0$, $q_0 \approx 0$)

(C) Students at no risk of being expelled in either type policy school ($q_1 = q_0 \approx 0$).

Table 1 lists the potential outcomes and the causal effects of interest.

Students in subpopulation (A) have four potential outcomes $Y(1,1)$, $Y(0,1)$, $Y(1,0)$, $Y(0,0)$. For subpopulation (A), to examine the impact of attending a zero tolerance policy school, the interest is in estimating the difference between these two effects, $E[Y(1, 1) - Y(0, 1)] - E[Y(1, 0)- Y(0, 0)]$, which will inform one as to whether the average expulsion effect on cognitive measures depends on whether or not a school has a zero tolerance policy. A continued contention of zero tolerance policy critics (Allen, 2004; Giroux, 2003a, 2003b; Skiba, et al., 2003) is that a school's adoption of a zero tolerance policy worsen troubled students' potential cognitive outcomes.

Students in subpopulation (B) have three potential outcome values $Y(1,1)$, $Y(0,1)$, $Y(0,0)$. The causal effect of particular interest in this dissertation is $E[Y(1, 1) - Y(0, 1)]$, which is the effect of being expelled in a zero tolerance school as opposed to not being expelled in a zero tolerance school on at risk children. The causal effects of interest in the ZT literature for students in subpopulations A and B are individual level effects.

The students in subpopulation (C) have two potential outcomes $Y(0,1)$, $Y(0,0)$, with $E[Y(0, 1)-Y(0, 0)]$ being the only estimand defined for this group which will inform one as to the impact of zero tolerance policy schools on “good students’” cognitive measures. This is the estimate obtained in this analysis.

Table 1

Potential Outcomes for Student Subpopulations and Causal Effects of Interest

Subpopulation	Probabilities of being expelled	Potential Outcomes	Causal effects of interest
A Students at risk of being expelled in a non-zero tolerance school	$q1 \geq q0 > 0$	Y(1,1), Y(0,1), Y(1,0), Y(0,0)	E[Y(1,0)- Y(0,0)] E[Y(1,1)-Y(0,1)] E[Y(1,1)-Y(0,1)]- E[Y(1,0)- Y(0,0)]
B Students at risk of being expelled only in a zero tolerance policy school	$q1 > 0$ and $q0 = 0$	Y(1,1), Y(0,1), Y(0,0)	E[Y(1,1)- Y(0,1)]
C Students at no risk of being expelled even in a zero tolerance policy school	$q1 = q0 = 0$	Y(0,1), Y(0,0)	E[Y(0,1)- Y(0,0)]

Note. Y(*i*,*j*). Student-level *i* = {0 not expelled; 1 expelled}. School-level *j*= {0 not a Zero tolerance policy school; 1 zero tolerance policy school}.

The Add Health study was not a randomized experimental design, and one must proceed with different assumptions in order to make a causal estimate regarding the treatment of interest. We denoted v(Z) = *V* as *the* random variable that takes on values *v*(z) = *v* = 0 for non-zero tolerance policy schools or *v*(z) = *v* = 1 for zero tolerance policy schools, that X was assigned to be a vector of child-level covariates and W as a vector of school-level covariates.

Let *Q* be the probability of a school having a tolerance policy. Conditioning on covariates, schools adopting a policy (*V*=1 or *V*=0) is influenced by *W* (school-level covariates) with a probability *Q* = *Q* (**W**); thus

$$Q = P(V = 1 \mid \mathbf{W}). \tag{6}$$

If the assignment of schools to a zero or non-zero tolerance policy is ignorable given the observed school-level pretreatment covariates **W**, then an unbiased estimate of school *j*'s propensity of selecting a zero tolerance policy, Q_j, can be made as a function of $\mathbf{W}_j$.

Let q be the probability that a student is expelled according to the school zero tolerance policy within the school that student attends. This could be written as

$$q_1 = P\left[Z = 1 \mid V = 1\right] \text{and } q_0 = P\left[Z = 1 \mid V = 0\right]. \tag{7}$$

If student expulsion or non-expulsion, under a zero or non-zero tolerance policy, is ignorable given the observed student-level pretreatment covariates **X** and the school-level covariates **W**, then student *i*'s propensity of being expelled in school *j*, denoted by q_{1ij} if the school has adopted a zero tolerance policy and by q_0ij otherwise, can be estimated as functions of $\mathbf{X}_i$ and $\mathbf{W}_j$. Hence $q_1 = q_1(\mathbf{X}, \mathbf{W})$ and $q_0 = q_0(\mathbf{X}, \mathbf{W})$, where

$$\begin{aligned} q_1 &= P(Z = 1 \mid V = 1, \mathbf{X}, \mathbf{W}) \\ q_0 &= P(Z = 1 \mid V = 0, \mathbf{X}, \mathbf{W}) \end{aligned} \tag{8}$$

A student's probability of receiving a specific treatment (expulsion or non-expulsion) under a specific tolerance policy (zero tolerance or not) can be expressed as

$P\ (Z = z, V = v|\ \mathbf{X}, \mathbf{W}\ ) = P\ (Z = z|\ V = v, \mathbf{X}, \mathbf{W}\ )\ P\ (V = v|\ \mathbf{W}\ ).$

This brings one to the following causal estimand:

1) The average causal effect of expulsion relative to non expulsion, conditioned on the covariates (**X** and **W**), under a non-zero tolerance policy for students in subpopulation A (students at risk of being expelled in a non-zero tolerance school), one may express this estimand

$$E\left[Y_{\mathrm{A}}(1,0) - Y_{\mathrm{A}}(0,0) \mid \mathbf{X}, \mathbf{W}\right]. \qquad (9)$$

2) the average conditional expulsion effect under a zero tolerance policy for students in subpopulation A (students at risk of being expelled in a non-zero tolerance school) and those in subpopulation B (students at risk of being expelled only in a zero tolerance policy school). Combining these subpopulations (A and B), produces a group of students who are at- risk of expulsion. Denoting this group of students as subpopulation AR (at risk), this estimand is expressed as

$$E\left[Y_{\mathrm{AR}}(1,\ 1) - Y_{\mathrm{AR}}(0,\ 1) \mid \mathbf{X}, \mathbf{W}\right]. \qquad (10)$$

For students in subpopulation A (students at risk of being expelled in a non-zero tolerance school), the difference between (9) and (10) can be considered an estimation of the extent to which the causal effect of the individual-level expulsion depends on the school-level expulsion policy, as these two expressions contrast the effect on student outcomes when expelled in a zero tolerance policy school (expression 10) as opposed to a non-zero tolerance policy school (expression 9).

3) the third estimand is the average causal effect of zero tolerance policy versus non-zero tolerance policy for students in subpopulation C (students at no risk of being expelled even in a

zero tolerance policy school), conditioning this estimand on school-level covariates **W**, this causal effect is defined

$$E\left[Y_{\mathrm{C}}\left(0,\ 1\right)-Y_{\mathrm{C}}\left(0,\ 0\right)\mid \mathbf{W}\right]. \tag{11}$$

Randomization would balance the joint distribution of school-level pretreatment covariates evenly across zero tolerance and non-zero tolerance schools (Fisher, 1925). If schools were assigned at random to a zero or non-zero tolerance policy, even though the within-school assignments of students to expulsion or non expulsion might not be random, then two propensity scores for every student, q_1 (student's probability of being expelled under a zero tolerance policy), and q_0 (student's probability of being expelled under a non-zero tolerance policy), could both be estimated. The propensity score function for q_1 estimated from the observed data of zero tolerance school students under ignorability would be also applicable to non-zero tolerance school students, and, the propensity score function for q_0 would apply to both non-zero tolerance and zero tolerance policy school- students.

Without randomization of the school-level policy adoption, as is the case with the Add Health observational study, only one of the two propensity scores for every student can be estimated from the observed data. This will affect the ability to estimate the causal effects defined in (9) and (10).

One may decompose Equation (9) as

$$E\left[Y_{\mathrm{A}}\left(1,0\right)-Y_{\mathrm{A}}\left(0,0\right)\mid \mathrm{V}=1,\mathbf{X},\mathbf{W}\right]P\left(\mathrm{V}=1\mid \mathbf{W}\right)+E\left[Y_{\mathrm{A}}\left(1,0\right)-Y_{\mathrm{A}}\left(0,0\right)\mid \mathrm{V}=0,\mathbf{X},\mathbf{W}\right]P\left(\mathrm{V}=0\mid \mathbf{W}\right) \quad (12$$

The limitation regarding estimation in this equation is with the first term, $\left[Y_{\mathrm{A}}\left(1,0\right)-Y_{\mathrm{A}}\left(0,0\right)\mid \mathrm{V}=1,\mathbf{X},\mathbf{W}\right]$, which cannot be directly estimated, because the propensity of expulsion under a non-zero tolerance policy is not estimable for students attending zero tolerance schools. One can estimate the second term for this population

$$\delta_{Z0} = E\left[Y_A(1,0) - Y_A(0,0) \mid \mathrm{V} = 0, \mathbf{X}, \mathbf{W}\right] \tag{13}$$

δ_{Z0} is the conditional effect of expulsion under a non-zero tolerance policy for students actually attending non-zero tolerance policy schools. The non-zero tolerance school students' propensity of being expelled if enrolled in a zero tolerance policy school is not directly estimable. The interaction effect between expulsion and zero tolerance school policy for those in subpopulation (A) is not directly estimable, as these students do not attend schools where the zero tolerance policy is in place.

Following this logic with the estimates of conditional average expulsion effects for students in the at-risk of expulsion subpopulation (AR), if one were to decompose expression (10) now as

$$\begin{aligned} &E\left[Y_{AR}(1,1) - Y_{AR}(0,1) \mid V = 1, \mathbf{X}, \mathbf{W}\right] \mathrm{P}(V{=}1 \mid \mathbf{W}) + \\ &E\left[Y_{AR}(1,1) - Y_{AR}(0,1) \mid V = 0, \mathbf{X}, \mathbf{W}\right] \mathrm{P}(V{=}0 \mid \mathbf{W}) \end{aligned} \tag{14}$$

one finds that the second term, $E\left[Y_{AR}(1,1) - Y_{AR}(0,1) \mid V = 0, \mathbf{X}, \mathbf{W}\right]$, cannot be estimated directly from the observational Add Health data, but can estimate its first term,

$$\delta_{Z1} = E\left[Y_{AR}(1,1) - Y_{AR}(0,1) \mid V = 1, \mathbf{X}, \mathbf{W}\right] \tag{15}$$

δ_{Z1} is the conditional effect of expulsion under the zero tolerance policy for students actually attending zero tolerance policy schools.

For those in subpopulation (C), the estimand as defined in (11) can still be estimated as

$$\delta_{V0} = E\left[Yc(0,1) - Yc(0,0) \mid \mathbf{W}\right]. \tag{16}$$

δ_{V0} is the conditional effect of attending a zero tolerance policy school versus a non-zero tolerance policy school for students at no risk for expulsion in either school type, and is the effect estimated with this analysis.

Causal Estimand Obtained with the Add Health Data Within this Causal Framework

The causal estimand that could be estimated within the proposed theoretical framework of ignorable treatment assignment is δ_{V0} (the conditional effect of attending a zero tolerance policy school versus a non-zero tolerance policy school for students at no risk for expulsion in either school type). The other two causal estimands, δ_{Z0} (the conditional effect of expulsion under a non-zero tolerance policy for students actually attending non-zero tolerance policy schools), and δ_{Z1} (the conditional effect of expulsion under the zero tolerance policy for students actually attending zero tolerance policy schools), are not valid estimands with this dataset. The individual level Add Health data regarding expulsion does not consider recidivism of individuals to expulsion (mandatory out-of-school suspension of over 15 days). The lack of distinction when assigning levels of treatment within the expelled treatment (i.e. low exposure to treatment is defined as expelled for 15 days, medium exposure to treatment is defined as expelled for 45 days, high exposure is defined as expelled for > 45 days) provides estimation effects of expulsion that do not answer the individual level causal policy questions of interest.

Theoretically, students at no risk of expulsion benefit from a learning environment where disruptions to learning are removed, such as illegal activities involving drugs and weapons. Following this logic, the removal of the students involved with the disruptive illegal activities, regardless of whether it is the same student removed repeatedly for such offenses, provides a learning environment conducive to good outcomes (Ewing, 2000). In the next section, the model used to obtain the causal estimate is described.

E. Obtaining Propensity Score Estimates

The causal estimate of interest which was estimated directly under the assumption of ignorability with the Add Health dataset was δ_{V0}. This causal estimand is of direct interest to answer the current research question about the effect of zero tolerance policies at an individual level.

Following expression (6) $Q = Q(\mathbf{W}) \equiv Q = P(V = 1 | \mathbf{W})$ a logistic regression was used to compute the estimate $\hat{Q}$, which is each school's conditional propensity of adopting a zero tolerance policy given the observed school-level covariates ($W_1 .... W_n$). Using conventional notation

$$\text{logit } \hat{Q}_j = \beta_0 + \beta_1 W_1 + \beta_2 W_2 + ... \ \beta_n W_n \tag{17}$$

logit $\hat{Q}_j$ is the log-odds of adopting a zero tolerance policy for school j . The log-odds is defined as the natural logarithm of the probability of adopting a zero tolerance policy divided by 1 – (the probability of adopting a zero tolerance policy); therefore logit $\hat{Q}_j$ = ln {p ($\hat{Q}_j$)/ [1- p ($\hat{Q}_j$)] }

β_0 represents the expected value of the log-odds for schools to be zero tolerance schools when all of the predictor variables equal zero.

$\beta_{1....}\ \beta_n$ represent the estimated change in the log-odds of a school being a zero tolerance policy school for every unit change in the independent variables in model (school-level covariates and aggregated student covariates).

On the basis of the logit of $\hat{Q}$, the sample of schools were divided into strata that had five sub classifications (quintiles). A minimum of five subclasses was used based on Cochran's (1968) observation that sub classification with five subclasses is sufficient to remove at least

90% of the bias for many continuous distributions. The quintiles were constructed arranging schools from lowest to highest on the estimated propensity score, and then dividing the schools into five strata of equal size.

The estimate of q_1 (the probability of being expelled from a zero tolerance policy school) is a function of school and student-level covariates for students attending zero tolerance schools. School-level covariates were included in the school-level propensity score estimates, and student-level covariates were included in the student-level propensity estimate of expulsion $\hat{q}_1$. As expulsion is a binary outcome, a logistic regression model was used for estimating $q_{1.}$ Using conventional notation, student-level covariates are ($x_{j} \ldots x_{nj}$ and $x_{1ij} \ldots x_{nij}$) in the below model. $\hat{q}_1$ estimates the propensity of each subject to be expelled from a zero tolerance policy school. Following Raudenbush and Bryk's (2002) notation,

$$\text{Level 1: Logit } \hat{q}_{1ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + .. \beta_{nj}x_{nij} \qquad (18)$$

Logit $\hat{q}_{1ij}$ is the log-odds of being expelled for student i in zero tolerance policy school j

β_{0j} represents the average log-odds for students with x=0 of being expelled in a zero tolerance policy school

$\beta_{1j} \ldots \beta_{nij}$ represent the estimated changes in the log-odds of being expelled in a zero tolerance policy school for each one unit change in the independent variable, while other variables are held constant.

In parallel to the estimation of q_1, q_0 was estimated through a logistic regression model for students attending non-zero tolerance policy schools. Again the covariates were represented as x's in the equation. The same logit link functions apply to these estimates.

$$\text{Level 1: Logit } \hat{q}_{0ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \beta_{nij}x_{nij} \qquad (19)$$

Logit $\hat{q}_{0ij}$ is the log-odds of being expelled for student *i* in non-zero tolerance policy school j

The initial propensity score estimates of $\hat{Q}$, $\hat{q}_0$, and $\hat{q}_1$ were examined for balance on covariates within each stratum. The goal of balancing is to verify that sub-classification on the estimated propensity score removes any initial bias on confounders. Balance on covariates was checked using multiple two-way analyses of variance (ANOVA), where treatment (zero tolerance school assignment, or expulsion, for Q and q respectively) was one factor, and the propensity score quintile to which the individual was assigned was a second factor (coded as a categorical variable with four levels), and each of the covariates (or confounders) was a dependent variable. If balance was achieved, there should not be a statistically significant main effect of treatment on the covariate. There should also not be a statistically significant interaction effect of the treatment variable by quintile (if these two conditions were not met, the propensity score would have been re-estimated by adding interaction terms and/or non-linear functions (e.g. square or cubic) of imbalanced covariates to the propensity score model). These balancing steps were repeated until balance was achieved or until no further improvement in balance could be made (Yanovitzky, Zanutto, and Hornik, 2005). At minimum, balance was achieved on identified key covariates. However, reliance on significance testing to check for covariate balance is sensitive to sample size (Shadish, Clark, and Steiner, 2008).

F. Identifying Empirical Subpopulations

The minimum value of the logit of q_1 (constructed from the dataset) for the expelled students attending zero tolerance schools was then used as the cutoff point to identify the non-expelled students in zero tolerance schools whose logit of q_1 was below this value. These identified non-expelled students with a logit below the minimum cutoff value, who did not have counterparts in the expelled group, were identified as students at an extremely low risk of expulsion under a zero tolerance policy. These students belong to subpopulation C (students unlikely to be expelled even in a zero tolerance policy school). Students whose logit of q_1 was above the cutoff value were considered to belong to subpopulations A or B (students who would be at risk of expulsion).

Subpopulation C members in the non-zero tolerance policy schools are hypothetical counterfactuals. Counterfactual estimation requires special assumptions (Hong & Raudenbush, 2006). The assumption was made that had a non-zero tolerance policy school instead adopted a zero-tolerance policy, its students would have been subject to similar selection criteria to be expelled, as defined by the propensity model specified for zero tolerance policy school students. Using this assumption, the estimated propensity model from the observed data of the zero tolerance policy schools' students was used as the same model to predict the propensity score for each student enrolled in a non-zero tolerance policy school. Specifically the minimum cutoff point of logit of $\hat{q}_1$ was taken (where below this value no expelled students were found) and used as the cutoff point for every student who attended non-zero tolerance policy schools as their risk of expulsion in a zero tolerance policy school. Using this method, the non-expelled students were identified in the non-zero tolerance policy schools who were highly unlikely to be expelled even if their schools had instead adopted a zero tolerance policy, and these students were assigned to subpopulation C. Recall that subpopulation C includes students in zero tolerance policy schools who were not expelled, and whose logit value of q_1 was below the minimum cutoff, leaving them without an expelled counterpart in their school.

A similar method for empirically identifying subpopulations can be used to identify students who attend zero tolerance policy schools, and have a risk of being expelled in non-zero tolerance policy schools. The assumption for the counterfactual estimate in this case is that had a zero tolerance policy school adopted a non-zero tolerance policy, its students would have been subject to similar criteria for expulsion that was defined in the propensity model of q_0 for students attending non-zero tolerance policy schools. After applying the estimated propensity model from the observed data of the non-zero tolerance policy schools' students to the students who attend zero tolerance policy schools, the zero tolerance policy students whose q_0 logit was above the maximum cut-off for non-expelled students in the non-zero tolerance policy students' model (expelled students without non-expelled student counterparts) could be identified as members of subpopulation A (students at risk of expulsion in non-zero tolerance policy schools). The remaining students who attended zero tolerance schools, who were not directly assigned to subpopulation C based on their q_1 logit ,or counterfactually to subpopulation A, could be assigned membership in subpopulation B. Lastly, the remaining students who attend non-zero tolerance policy schools, and whose q_0 logit was below the minimum cutoff for population A inclusion in that model, but above the cutoff for inclusion in population C when the q_1 logit model was applied as a counterfactual model to their data, could assigned population B. Table 2 contains a summary of methods to identify members of these sub-populations.

After the initial estimate of causal effects by propensity score stratification, the clustering of students within schools was modeled with a hierarchical linear model to account for student clustering within schools.

G. Multilevel Model for Causal Effect Estimation

Model for causal estimand of effect of zero tolerance policy relative to non-zero tolerance policy on Add Health Picture Vocabulary Test Scores for the students in subpopulation C (students at little risk of being expelled even in a zero tolerance policy school).

The estimand $\delta_{V0,}$ as defined in (16), was estimated by a statistical adjustment for a school's propensity of adopting a zero tolerance policy. The raw difference in the average Add Health Picture Vocabulary Test score outcome was observed across strata as described in the preceding section. Now, the analytical plan was to look at the variation of the expulsion effect across the schools, by using a hierarchical two-level regression model, with student at level 1 and school at level 2, as shown in Table 2, for child *i* in school j

$$\begin{aligned} Yij &= \gamma_0 + u_j + \delta_{v0} v_j + \sum_{h=2}^{n\_strata-1} \gamma_h Q_{hj} + e_{ij} \\ u_j &\sim N(0,\tau); e_{ij} \sim N(0,\sigma^2) \end{aligned} \tag{22}$$

In the above equation

Yij is the score of student *i* within school j on the Add Health Picture Vocabulary Test

γ_0 is the average AHPVT school score for non-ZTP students in stratum 1

$\delta_{v0} v_j$ represents the effect of school with a zero tolerance policy.

$\sum_{h=2}^{n\_strata-1} \gamma_h Q_{hj}$ represents the dummy indicators for the effect of strata assigned propensity for school j to adopt a zero tolerance policy.

e_{ij} represents the residual remaining after adjustment to AHPVT test score for ZTP school effect and strata sub classification effect around each student measure.

u_j represents the school-specific random effect on the AHPVT score.

This gives the conditional effect of zero tolerance policy in zero tolerance policy schools for children at no risk of being retained even though they attend a zero tolerance policy school.

Table 2

Identification of Subpopulations for Causal Estimates

Subpopulation A	Subpopulation B	Subpopulation C
From students who attend ZTP[a] schools(counterfactual);	From students who attend ZTP[a] schools (indirect);	From students who attend ZTP[a] schools (direct);
After applying the logit model for $q0$ (obtained from model on students in non-ZTP[a] schools) to ZTP[a] students: students whose $q0$ logit is greater than the minimum logit of $q0$ of expelled students attending non ZTP[a] schools.	Students not assigned to population C from $q1$ logit estimate (their $q1$ logit was above the minimum cut-off of expelled students), AND whose logit of $q0$ was below the cutoff for expelled students' in the observed data of $q0$ model.	*Non-* expelled students, with a $q1$ logit below the minimum cutoff value for expelled students' logit of $q1$, will not have counterparts in the expelled group.
From students who attend non-ZTP[a] schools (direct);	From students who attend non-ZTP[a] schools (indirect);	From students who attend non-ZTP[a] schools (counterfactual);
The non-expelled and expelled students from upper strata of logit of $q0$, whose $q0$ logit are greater than or equal to the minimum logit of $q0$ of expelled students.	Non- expelled students in non-ZTP[a] schools whose $q0$ logit was below the cut-off for expelled students in non-ZTP[a] schools, AND whose $q1$ logit was greater than or equal to the $q1$ logit minimum cut off for expelled students attending ZTP[a] schools.	After applying the q1 logit model to students attending non-ZTP[a] schools, those students whose logit fell below the minimum cutoff value for expelled students' $q1$ logit

Note. [a]ZTP = Zero Tolerance Policy.

CHAPTER IV: RESULTS

A. School Propensity of Adopting a Zero Tolerance Policy (ZTP), *Q*. (N=130)

Logistic regression was used to compute the estimate, *Q*, of each school's conditional probability of adopting a zero tolerance policy given the observed school-level covariates. A bar graph representing the distribution of schools according to the propensity score estimates of adopting a zero tolerance policy is in Figure 1. The majority of schools in the sample were Non-ZTP schools (N=110), with the remainder being ZTP schools (N=20). In Figure 1, the distributions of the estimates of *Q* of the two schools types are displayed; ZTP schools are represented across the upper half of the graph, and non-ZTP schools are represented across the lower half of the graph. The model used was good for prediction of a school's propensity to adopt a ZT policy; the model also exposed the lack of a uniform distribution of ZTP and non-ZTP schools.

On the basis of the logit of the *Q* estimate, the sample of schools (N=130) was divided into five strata. The results of the analysis are displayed in Table 3.

The distribution of covariates believed to effect the treatment or outcome variable was checked for balance in the strata of the logit of the *Q* estimate containing both ZTP and non ZTP schools. Balance was achieved on the18 school-level pretreatment covariates, with 90% of the main effects and interactions having significance levels > .05. The results of the propensity score estimates within strata differences are summarized in Table 4. The balance on covariates was achieved, with the 95% CI for the mean differences for all but two covariates containing zero. The differences between the mean value of covariates within quintiles of logit of propensity score distributions was statistically significant on six of the covariate results displayed.

Figure 1

Plot of Distribution of Estimated Propensity Score for ZTP School Adoption (N=130)

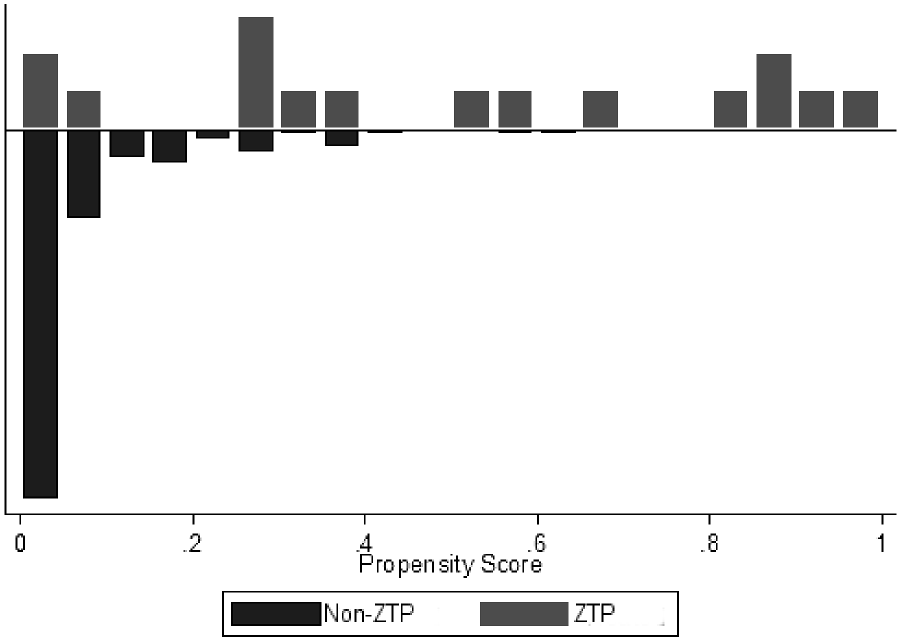


Table 3

Within-Stratum Distribution of Schools' Logit of Q Estimate and Average Propensity Estimate to Adopt a ZTP[a]

	ZTP[a] school				Non-ZTP[a] school			
Stratum	N	Mean	SD	Propensity	N	Mean	SD	Propensity
Q=1	3	-6.96	0.44	.001	23	-7.89	1.28	.001
Q=2	1	-4.75	0	.009	25	-4.72	0.43	.009
Q=3	3	-3.45	0.26	.031	23	-3.44	0.39	.032
Q=4	0				26	-2.28	0.45	.099
Q=5	13	0.80	1.58	.622	13	-0.62	0.60	.359

Note. [a]ZTP = Zero Tolerance Policy.

Table 4

Balance of Covariates in Logit of Q Estimate

		Significance level	
Covariate	ME covariate	ME quintile	covariate*quintile
Catchme: Catchment school	.64	.15	.18
A7:Average class size	.26	.35	.29
A12: % of New Teachers	.73	.57	.05
A13: % of Teachers w Masters	.45	.000*	.36
Ln_mehhl: Avg household inc	.95	.01	.91
Subdinc: % subsidized income	.71	.21	.32
Engprim:English primary home	.69	.32	.50
Atype: Public or private	.29	.01	.29
Rural: School location	.91	.25	.58
Suburb: School location	.76	.03	.04
Neast: School region	.74	.000*	.74
Midwest: School region	.86	.17	.83
South: School region	.93	.000*	.38
Jhsandhs: School grades	.38	.26	.38
Jhonly: School grades	.78	.42	.89
Avgatt: Daily attendance rate	.91	.98	.44
Loatt: Daily attendance rate	.58	.12	.62
Hidropou: High Dropout rate	.21	.23	.28

* <.001

in Table 4. The percentage of teachers with Master's degrees in the schools is presented in Figure 2, with the percentage being significantly higher in schools with a lower propensity to adopt a ZTP policy (variable a13); schools with a lower propensity for ZTP adoption had a significantly higher frequency of schools located in the Northeastern United States (variable neast), presented in Figure 3, while schools with a higher propensity to adopt ZTP policy had a statistically significant higher frequency of being located in the Southern United States (variable south), presented in Figure 4; ZTP schools also had a lower reported household income (variable Ln_mehhl), while non-ZTP schools had a higher frequency of being located in the suburbs (variable suburb).

Sample size was small for the ZTP schools (N=20), so mean differences across the schools were compared as well, and the results of the analysis are summarized in Table 5.

B. Student Propensity of Expulsion in Non-ZTP School, q_0.

Student propensity for expulsion within a non-ZTP school, q_0 was estimated through a logistic regression model for children attending non-ZTP schools only. A bar graph representing the distribution of students according to the propensity score estimates of being expelled from a non-ZTP school is presented in Figure 5. In Figure 5, expelled students (n=383) are represented across the upper half of graph and non-expelled students (n=10,654) are represented across the lower half of the graph.

On the basis of the logit of the q_0 estimate, the sample of students (n= 11037) was divided into five strata; the results of the analysis are summarized in Table 6. The students who were expelled had a high estimated propensity of being expelled (97% of expelled students were estimated as having a high propensity to be expelled). In contrast the students who were not expelled had lower estimated propensities for being expelled, with the non-expelled students belonging to the higher quintiles maintaining a lower propensity to be expelled than the expelled students within those quintiles.

The distribution of covariates believed to effect the treatment or outcome variable was checked for balance in the strata of the logit of q_0 estimate containing both expelled and non expelled students. After multiple trial models to improve balance, such as adding interactions and squaring non-linear terms, balance was best achieved on covariates without added terms. Balance was achieved on more than 99% of the 30 student-level pretreatment covariate main effects and interactions; the results of the analysis are summarized in Table7. However, 76 % of the main effect of quintile was not balanced (balance being statistically significant at an alpha of < .05). Students who had lower propensity estimates to be expelled had a decreased perception of autonomy to make decisions regarding weeknight behaviors, h1wp6, presented in Figure 6, and students who were not expelled had consistently higher beliefs in their perception of their decision-making autonomy during weeknights than their expelled counterparts, as presented in Figure 7 .

Figure 2

Estimated Mean Proportion of Teachers with Masters Degree by Quintiles of Logit of Estimated Q, the Propensity to be a ZTP School

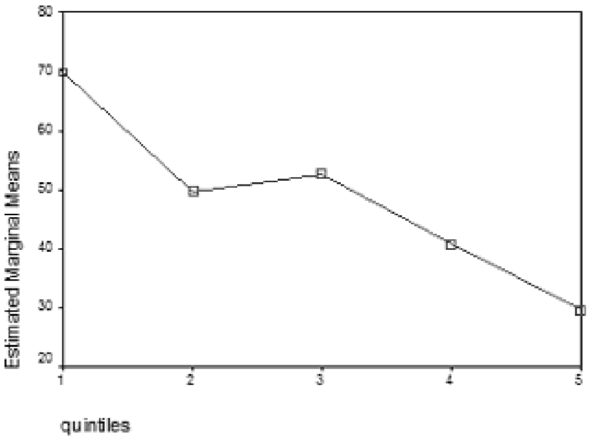


Figure 3

Estimated Mean Proportion of Schools Located in Northeast US by Quintiles Logit of Estimated Q, the Propensity to be a ZTP School

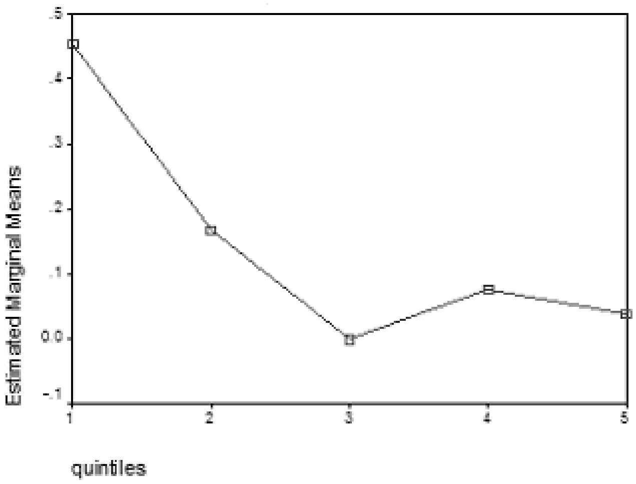


Figure 4

Estimated Mean Proportion of Schools Located in South US, by Quintiles of Logit of Estimated Q

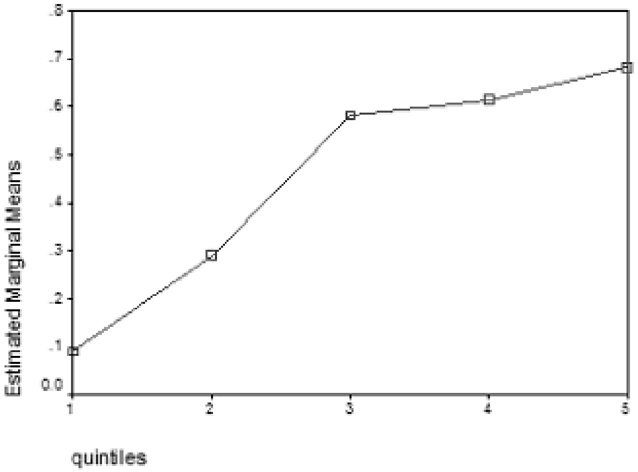


Table 5

Mean Differences Across Q Estimate Covariates for ZTP and Non-ZTP Schools

	ZTPvalue – nonZTPvalue		
Covariate	Mean Difference	Standard error	95% CI for Difference
Catchme	- .08	.14	-.20 - .36
A7	2.84	1.61	-.35 - 6.02
A12	-.66	4.70	-9.96 - 8.64
A13	11.75	7.50	-3.12 - 26.60
Ln_mehhl	.07	.12	-.18 - .31
Subdinc	-.02	.06	-.14 - .09
Engprim	-3.91	4.45	-12.72 - 4.91
Atype	-.16	.09	-.34 - .02
Rural	-.00	.12	-.25 - .24
Suburb	.06	.16	- .26 - .37
Neast	.10	.11	-.11 - .31
Midwest	.03	.14	-.25 - .31
South	-.19	.16	-.498 - .12
Jhsandhs	-.07	.08	-.23 - .08
Jhonly	.08	.17	-.25 - .41
Avgatt	-.00	.17	.34 - .34
Loatt	.12	.12	-.12 - .35
Hidropou	.16	.31	-.15 - .48

Figure 5

Graph of Propensity Estimates for Student Expulsion in Non-ZTP School, q_0 (n= 11037)

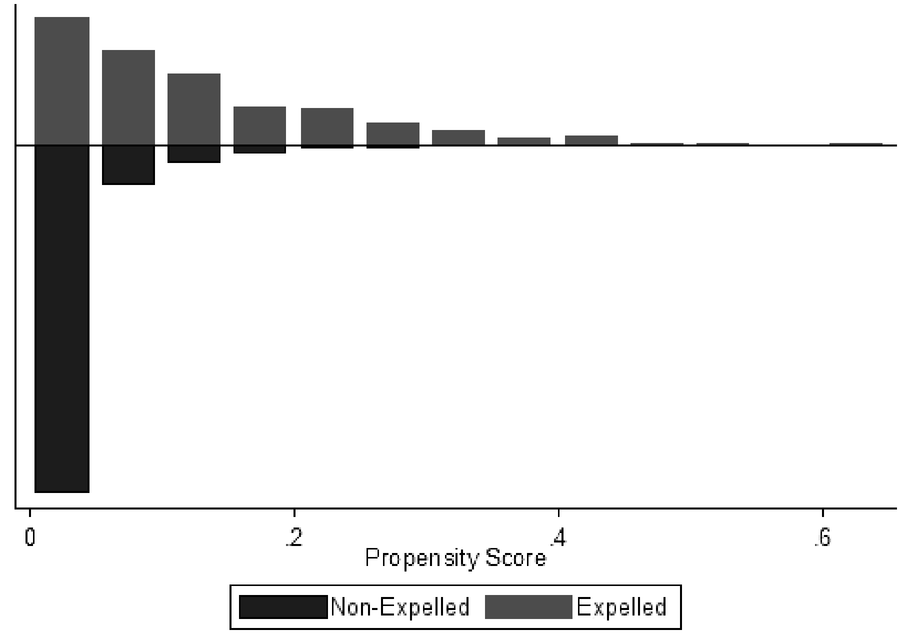


Table 6

Within-Stratum Distribution of the Logit of the Propensity and Average Propensity Estimate for Expulsion for Students Attending NON ZTP Schools (q_0)

	Expelled				Non-expelled			
Stratum	n	Mean	SD	Propensity	n	Mean	SD	Propensity
q_0=1	5	-5.44	.15	.004	2203	-5.96	.95	.003
q_0=2	5	-4.81	.13	.008	2202	-4.96	.19	.007
q_0=3	24	-4.31	.20	.013	2184	-4.30	.20	.013
q_0=4	79	-3.47	.27	.031	2128	-3.54	.25	.029
q_0=5	270	-1.78	.83	.172	1937	-2.22	.66	.114

Table 7

Balance of Covariates in Logit of q_0 *Estimate*

	Significance level		
Covariate	ME covariate	ME quintile	Interaction cov*quint
bio_sex:Gender	.78	.000*	.28
h1gi8:Race	.50	.11	.64
h1da1: Does Home Chores	.29	.40	.71
h1da8: Hours of Television	.63	.00	.65
h1gh23j: Doesn’t Eat Breakfast	.72	.28	.84
h1ed3: Ever Skipped a Grade	.88	.01	.83
h1ed5: Ever Repeated a Grade	.18	.000*	.43
h1ed11:Recent English Grade	.08	.000*	.34
h1ed12: Recent Math Grade	.08	.000*	.02
h1ed20: Feels Part of School	.20	.000*	.03
h1ed24: Feels Safe at School	.52	.000*	.26
Rmomed: Mother Education	.90	.000*	.47
h1wp6: Weeknight limits	.02	.13	.07
h1wp8: Dinner with Parent	.34	.000*	.64
h1pf5: Relationship with Father	.99	.15	.42
h1pf27: Seldom Sick	.49	.05	.27
h1co1: Ever had Sex	.56	.000*	.31
h1to1: Ever Smoked	.44	.000*	.52
h1to29: Friends Drink Alcohol	.85	.000*	.98
Smokemj: Ever Tried Marijuana	.15	.000*	.09
h1to51: Home Access Alcohol	.65	.01	.10
h1ds15: Ever Rowdy in Public	.04	.04	.09
h1fv6: Ever Jumped	.14	.000*	.02
h1pr1: Adults Care about Me	.22	.000*	.40
h1pr2: Teachers Care about Me	.92	.000*	.56
h1pr4: Friends Care about Me	.20	.000*	.02
h1ee1: Wants to go to College	.75	.000*	.05
h1ee4: Works Schoolweek	.68	.44	.98
pa55: Household Income	.82	.000*	.93
pa10: Parent Marital Status	.49	.000*	.27

*<.001

Figure 6

Estimated Mean Probability of Weeknight Autonomy Decision-Making by Quintile of Propensity for Student to be Expelled in a non-ZTP School

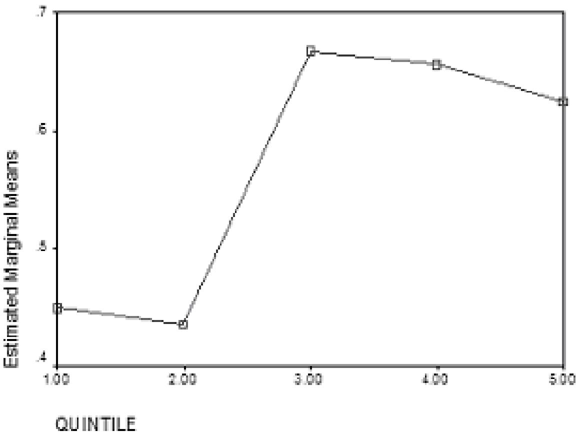
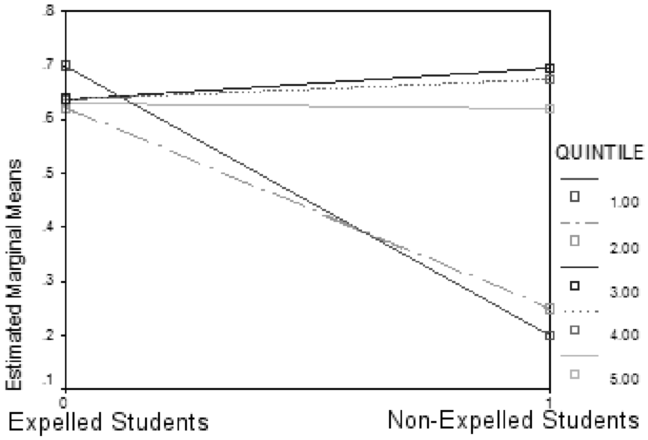


Figure 7

Estimated Means of Weeknight Autonomy Decision-Making for Expelled and Non-expelled Students by Quintile of Propensity for Student to be Expelled in a Non-ZTP School



Students who were expelled had a statistically significantly higher self-report of a history of being loud and rowdy in public places (variable h1ds15) than students who were not expelled (see Table 7); however, the self-report of loud and rowdy behavior in public places fluctuated across quintiles for both expelled and non-expelled students; results are presented in Figure 8.

C. **Student Propensity of Expulsion in ZTP School, q_1.**

Student propensity for expulsion within a ZTP school, q_1 was estimated through a logistic regression model for children attending ZTP schools only. A graph representing the distribution of students according to the propensity score estimates of being expelled from a ZTP school is presented in Figure 9. In Figure 9, expelled students are represented across the upper half of the graph and non-expelled students represented across the lower half of the graph.

On the basis of the logit of the q_1 estimate, the sample of students (n = 1098) was divided into five strata; the results are summarized in Table 8. The division of the distribution of expelled and not expelled students based on the propensity score estimates reveals the higher propensity for expulsion for students actually expelled than for the not expelled students (42 of the 44 expelled students have propensity score estimates in the fourth and fifth quintiles). In contrast, the non-expelled students have propensity score estimates distributed approximately evenly in all five quintiles. The propensity score estimate quintile stratification clearly demonstrates the lack of expelled student counterparts for the non-expelled students in quintiles one through four.

Figure 8
Estimated Mean Proportions of Students with Self-Reported Loud and Rowdy Public Behavior, for Expelled and Non-Expelled Students by Quintile

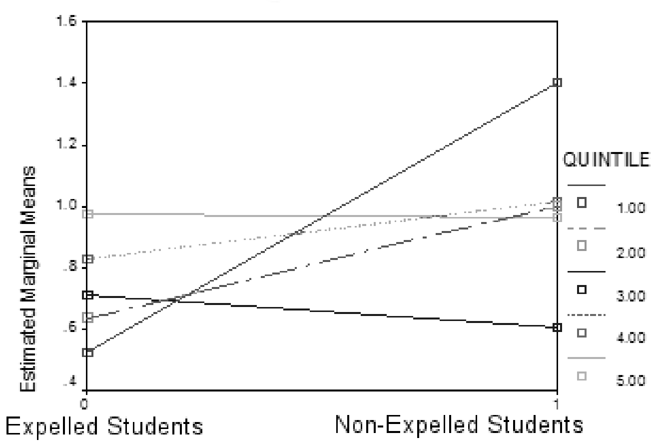


Figure 9

Propensity Estimates for Student Expulsion in ZTP School, q_1 (n= 1098).

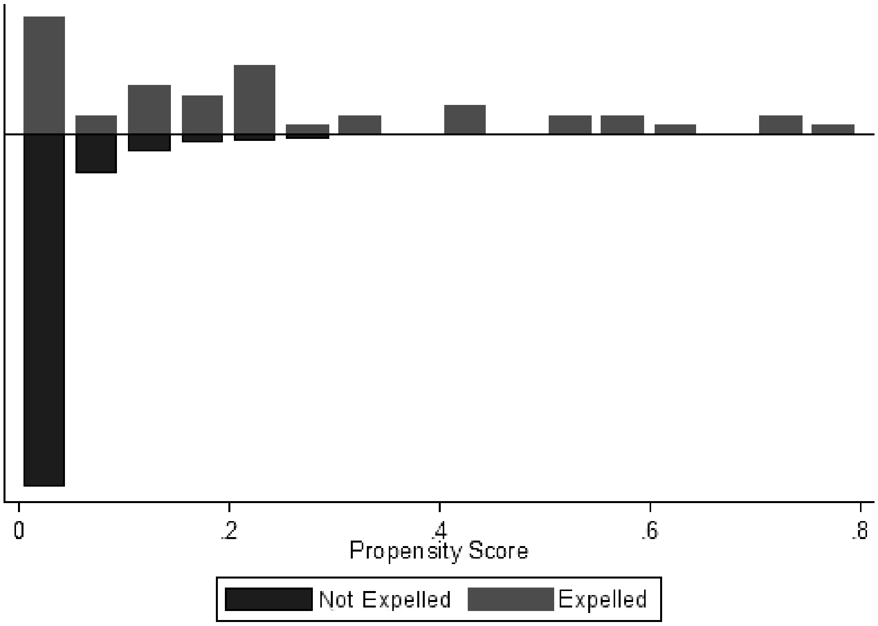


Table 8

Within-Stratum Distribution of the Logit of the Propensity and the Average Propensity Estimate for Expulsion in Students Attending ZTP Schools q_1

	Expelled				Non-Expelled			
Stratum	n	Mean	SD	Propensity	n	Mean	SD	Propensity
q_1=1	0				220	-7.01	.61	.001
q_1=2	2	-5.04	.32	.004	217	-5.67	.27	.003
q_1=3	0				220	-4.79	.28	.009
q_1=4	8	-3.66	.38	.035	211	-3.71	.36	.025
q_1=5	34	-.92	1.30	.330	186	-2.13	.78	.128

The distribution of covariates believed to effect the treatment or outcome variable was checked for balance in the strata of the logit of q_1 estimates containing both expelled (n=44) and non-expelled (n=1054) students. After multiple trial models to improve balance, such as adding interactions and squaring linear terms, it was observed that balance was best achieved on covariates without added terms. Balance was achieved on 100% of the 30 student-level pretreatment covariates main effects and interactions, with results summarized in Table 9; however 83% of the main effects of quintile were not balanced (balance being statistically significant at a significance level < .05).

Non-expelled students differed from their expelled counterparts, reporting having skipped a grade (variable h1ed3) with more frequency than expelled students. Expelled students were less likely to eat dinner with parents (variable h1wp8), reported having friends who drank alcohol more often, and their parents reported a lower household total income (variable pa55) than the non-expelled students in the higher quintiles. The average estimated logit for propensity for expulsion was higher for expelled students than for non-expelled students in quintile 5 (logits of q1 estimates -.92279 with standard error 1.2975, and -2.1349 with standard error .7792, respectively).

Table 9

Balance of Covariates in Logit of q_1 Estimate

	Significance level		
Covariate	ME covariate	ME quintile	Interaction cov*quint
bio_sex:Gender	.10	.000*	.13
h1gi8:Race	.84	.36	.89
h1da1: Does Home Chores	.36	.000*	.53
h1da8: Hours of Television	.54	.000*	.66
h1gh23j: Doesn't Eat Breakfast	.70	.000*	.76
h1ed3: Ever Skipped a Grade	.69	.000*	.25
h1ed5: Ever Repeated a Grade	.94	.000*	.26
h1ed11:Recent English Grade	.91	.000*	.32
h1ed12: Recent Math Grade	.76	.000*	.45
h1ed20: Feels Part of School	.48	.000*	.35
h1ed24: Feels Safe at School	.34	.07	.32
Rmomed: Mother Education	.36	.000*	.87
h1wp6: Weeknight limits	.98	.000*	.80
h1wp8: Dinner with Parent	.99	.000*	.65
h1pf5: Relationship with Father	.62	.03	.60
h1pf27: Seldom Sick	.77	.48	.93
h1co1: Ever had Sex	.88	.000*	.56
h1to1: Ever Smoked	.65	.00	.52
h1to29: Friends Drink Alcohol	.87	.000*	.55
Smokemj: Ever Tried Marijuana	.81	.000*	.28
h1to51: Home Access Alcohol	.37	.000*	.50
h1ds15: Ever Rowdy in Public	.66	.000*	.50
h1fv6: Ever Jumped	.99	.000*	.23
h1pr1: Adults Care about Me	.83	.000*	.64
h1pr2: Teachers Care about Me	.64	.000*	.60
h1pr4: Friends Care about Me	.53	.08	.83
h1ee1: Wants to go to College	.81	.000*	.04
h1ee4: Works Schoolweek	.94	.83	.87
pa55: Household Income	.72	.00	.10
pa10: Parent Marital Status	.79	.000*	.78

*<.001

D. Empirical Identification of Students in Subpopulation C.

The value of -5.216 was used as the cutoff point to identify the 440 non-expelled students in zero tolerance schools belonging to subpopulation C (students unlikely to be expelled even in a zero tolerance policy school, as there were no expelled students with logit less than or equal to -5.216). The same minimum value of the logit of $\hat{q}_1$ (-5.216) was used as the cutoff point to identify the 3,658 students attending non-ZTP schools belonging to subpopulation C, bringing the subpopulation C total sample size to 4,095. On the basis of the logit of q_1 for the sample of subpopulation C, the sample was divided into five strata, with results summarized in Table 10.

The distribution of covariates believed to effect the treatment or outcome variable was checked for balance in the strata of the logit of $\hat{q}_1$ estimate for the subpopulation C students, comparing students who attend ZTP schools and students who attend non ZTP schools. Balance was achieved on 88% of the 30 student-level pretreatment covariates main effects, 90% of the quintile main effects, and 99% of the interactions; the results are summarized in Table 11. Differences in means were considered to be unbalanced, and statistically significant at an alpha of $< .05$.

There were differences between students who attended ZTP schools and those who attended non-ZTP schools, summarized in Table 11 for students' reporting ever repeating a grade (variable h1ed5), recent math grade (variable h1ed12), feeling a part of school (variable h1ed20), feeling safe at school (variable h1ed24), ever having sex (variable h1co1), and reporting that teacher cares about me (variable h1pr2).

Students attending ZTP schools reported a higher incidence of repeating a grade (Figure 10), lower recent math grades (Figure 11), lower reporting of feeling part of their school (Figure 12), feeling safe at school (Figure 13). Students who attend non-ZTP schools self reported lower amounts of ever having sex, but reporting across quintiles did not reflect trending (Figures 14 and 15 respectively), while students attending ZTP schools reported a lower feeling that their teacher cares about them (Figure 16) .

Table 10

Within-Stratum Distribution of the Logit of Propensity and the Average Propensity Estimate for Expulsion in Subpopulation C Students Attending ZTP Schools and Non-ZTP Schools, Based on Empirical Model of q_1

	ZTP school students				Non-ZTP school students			
Stratum	n	Mean	SD	Propensity	n	Mean	SD	Propensity
$q_1=1$	76	-7.70	.46	.000	747	-7.82	.58	.000
$q_1=2$	96	-6.80	.19	.001	729	-6.80	.19	.001
$q_1=3$	82	-6.25	.13	.002	713	-6.25	.14	.002
$q_1=4$	83	-5.78	.12	.003	745	-5.80	.11	.003
$q_1=5$	100	-5.42	.11	.004	724	-5.42	.11	.004

Table 11

Balance of Covariates in Logit of q1 Estimate for Subpopulation C Students

Covariate	Significance level ME covariate	ME quintile	Interaction cov*quint
bio_sex:Gender	.25	.01	.11
h1gi8:Race	.52	.70	.42
h1da1: Does Home Chores	.362	.36	.60
h1da8: Hours of Television	.60	.41	.43
h1gh23j: Doesn't Eat Breakfast	.20	.05	.18
h1ed3: Ever Skipped a Grade	.92	.99	.99
h1ed5: Ever Repeated a Grade	.06	.10	.18
h1ed11:Recent English Grade	.35	.95	.70
h1ed12: Recent Math Grade	.00	.66	.55
h1ed20: Feels Part of School	.03	.05	.11
h1ed24: Feels Safe at School	.01	.06	.14
Rmomed: Mother Education	.09	.81	.27
h1wp6: Weeknight limits	.06	.39	.53
h1wp8: Dinner with Parent	.93	.30	.81
h1pf5: Relationship with Father	.46	.71	.70
h1pf27: Seldom Sick	.15	.52	.58
h1co1: Ever had Sex	.10	.50	.41
h1to1: Ever Smoked	.07	.36	.49
h1to29: Friends Drink Alcohol	.83	.09	.51
Smokemj: Ever Tried Marijuana	.47	.08	.30
h1to51: Home Access Alcohol	.74	.06	.34
h1ds15: Ever Rowdy in Public	.78	.61	.26
h1fv6: Ever Jumped	.62	.51	.60
h1pr1: Adults Care about Me	.54	.80	.82
h1pr2: Teachers Care about Me	.00	.000*	.000*
h1pr4: Friends Care about Me	.14	.55	.55
h1ee1: Wants to go to College	.07	.04	.03
h1ee4: Works Schoolweek	.76	.10	.16
pa55: Household Income	.07	.99	.98
pa10: Parent Marital Status	.57	.08	.09

*<.001

Figure 10

Difference in Estimated Means of Repeating a Grade by ZTP School Policy

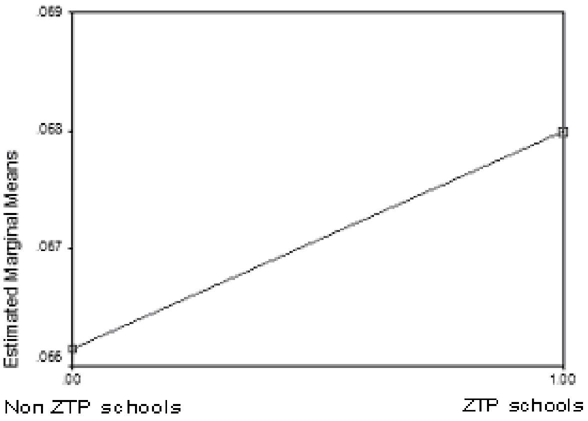


Figure 11

Difference in Estimated Means of Recent Math Grade by ZTP School Policy (1= A, 5=F)

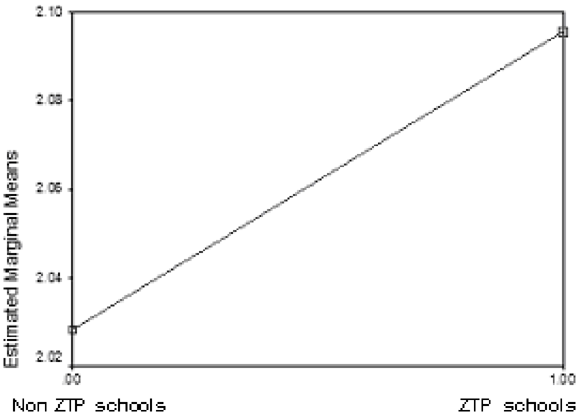


Figure 12

Difference in Estimated Means of Overall Feeling Part of Your School by ZTP School Policy (1= Strongly Agree; 4=Strongly Disagree)

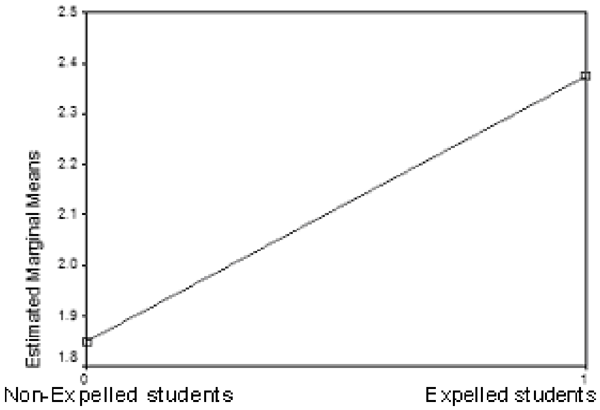


Figure 13

Difference in Estimated Means of Overall Feeling Safe at Your School by ZTP School Policy (1= Strongly Agree; 4=Strongly Disagree)

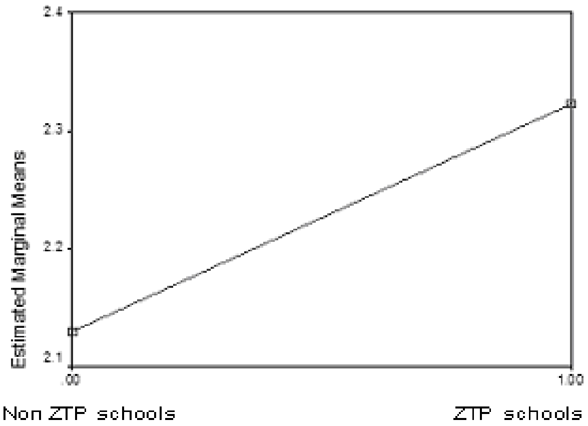


Figure 14

Difference in Estimated Means of Ever Had Sex by ZTP School Policy

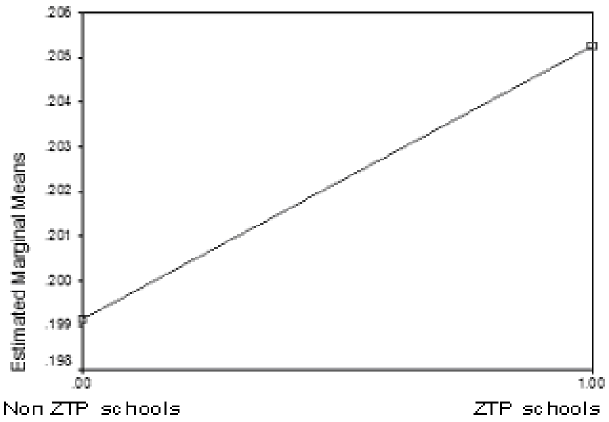


Figure 15

Difference in Estimated Means of Ever Had Sex by ZTP School Policy and by Quintile

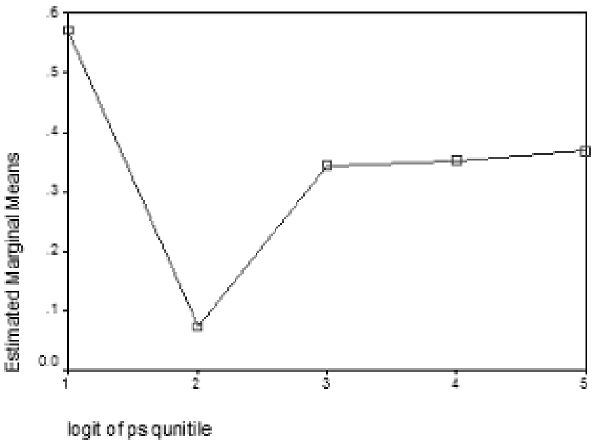
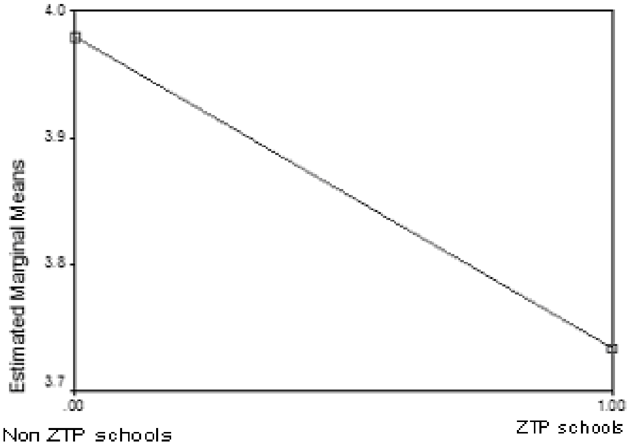


Figure 16

Difference in Estimated Means of Feeling Teachers Care About You by ZTP School Policy (1= Not At All; 4= Quite A Bit)



E. Zero Tolerance Policy Effect on Students in Subpopulation C

The raw difference in the average Add Health Picture Vocabulary Test score outcome was observed across strata, and an estimate of the overall differences in Add Health Picture Vocabulary Test was then calculated, with the results of the analysis being summarized in Table 12. The overall weighted effect on AH PVT scores for students attending ZTP schools is -.927. The difference in AH PVT scores between students who attend ZTP schools and non-ZTP schools is not statistically significant overall, or within strata, and varies across strata.

The causal estimand of interest was then obtained with a hierarchical two-level regression model, with students at level 1 clustered within school at level 2; the results of the analysis are summarized in Table 13.

The raw difference in AH PVT outcome of -.927 favoring non-ZTP schools was supported with the causal estimate of attending a ZTP school δ_{v0} obtained with the hierarchical linear model. The within strata effect of attending a ZTP school on AH PVT scores for students in subpopulation C was negative in all five strata. The causal estimand of attending a ZTP school, $\delta_{v0,\ obtained}$ was -.83with a standard error of 1.65 and a 95% confidence interval of (-4.01, 2.40). No statistically significant difference in the AH PVT outcomes within subpopulation C was estimated between ZTP school students and non-ZTP school students.

Table 12

Within-Stratum Distribution of the AH Peabody Picture Verbal Outcome of Students in Subpopulation C

	Attend ZTP school			Attend Non-ZTP school			Mean Difference
Stratum	n	Mean	SD	n	Mean	SD	
1	74	105.91	12.26	711	106.71	13.41	-.08
2	94	105.03	11.80	702	105.77	13.87	-.74
3	79	106.73	11.96	682	104.57	14.04	+2.16
4	84	100.99	13.59	712	104.15	13.61	-3.16
5	96	101.35	13.18	695	103.00	13.21	-1.65
Total	427	103.88	12.76	3502	104.85	13.68	-.93

Table 13

Causal Effect of Attending a ZTP School Versus Attending a NON ZTP School on Students AH PVT Cognitive Score Outcome Measure for Students in Subpopulation C, δ_{v0}

Fixed Effect	**Coefficient**			**SE**	**t**
NON ZTP students in Propensity stratum 1 intercept, γ_0	107.25			1.42	75.29
Effect of attending ZTP schools, δ_{v0}	-0.83			1.65	-0.50
Propensity stratum 2, γ_2	-4.65			1.91	-2.43
Propensity stratum 3, γ_3	-3.43			1.72	-2.00
Propensity stratum 4, γ_4	-2.85			1.72	-1.66
Propensity stratum 5, γ_5	-3.04			1.96	-1.55
Random Effect	**Variance**	**SD**	**df**	χ^2	**p value**
School specific effect, u_j	25.21	5.02	112	677.33	.000*
Student specific effect, e_{ij}	155.09	12.45			

*<.001

CHAPTER V: DISCUSSION

A. Overview

The purpose of this research was to obtain causal estimates of the impact of ZTP on individual students' cognitive outcomes with a nationally representative observational dataset. To obtain these estimates the distribution of covariates between the units of comparison at both school and student-levels, and the natural clustering of students within schools needed to be considered in the models.

At the school-level of this data, ZTP and non-ZTP schools estimated distribution of the logit for adopting a zero tolerance policy allowed for the causal estimate of interest (δ_{v0}) in Table 3. Schools adopt ZTP with a non-experimental assignment to treatment, but the results of the propensity score estimates allowed causal inference to be estimated with the Add Health data sample.

At the student-level of the sample, there were two propensity score estimates obtained empirically, the first being q_0 from children attending non-ZTP schools. The difference in the distribution of covariates within q_0 limited the causal inference estimates that could be obtained with this sample to expelled and non-expelled students with a high propensity to be expelled within non-ZTP schools only. This is a different population than all students who attend non-ZTP schools. This analysis did not make any causal estimate from $q_{0.}$

The second student-level estimate, $q_{1,}$ was obtained empirically from students who attended ZTP schools. Observations across the means of the covariates in the quintiles that contained both expelled and non-expelled students were analyzed. The q_1 empirical estimate was used to identify the students who attend non-ZTP schools who belonged to subpopulation C (students with low propensity for expulsion in ZPT schools).

After the Subpopulation C students were identified, the causal estimand of interest was obtained with a hierarchical model that accounted for school membership.

The methodology followed in this research produced an unbiased causal estimand of ZTP with an observational data sample if the assumptions for the modeling were plausible.

B. Implications for Zero Tolerance Policy

This is the first estimate of ZTP effects that took into account the non-experimental nature of educational data, and the natural grouping of students within schools, when modeling causal effects. The results did not support the hypothesis that ZTP has a positive effect on students at a low propensity for expulsion in ZTP and non-ZTP schools by creating learning environments that allow “good citizen students” to flourish by eliminating the disruptive students with mandatory expulsion (Ewing, 2000).

The majority of research thus far (Giroux, 2003, Skiba, et al., 2003, Skiba & Knesting, 2001, Skiba & Petersen, 1999) has attempted to produce estimates that support a hypothesis of the negative effect that ZTP has on a different subpopulation of students, those students at risk for expulsion in non-ZTP schools, and students at risk only in ZTP schools (subpopulations A and B). The results obtained with this research cannot be generalized to students in subpopulations A and B.

Previous research has been an attempt to measure both an individual-level treatment (expelled or not expelled) and the effects of zero tolerance policies on cognitive outcomes at the individual student-level, not accounting for the actual treatment assignment of interest, school adoption of zero tolerance policy. If the treatment variable of interest (zero tolerance policy adoption) is not measured at the level of assignment (school-level), accurate causal estimates will not be made. Also, if non-random treatment assignment at the group and individual level is not considered in the assumptions and modeling, the causal estimates produced may be biased. Lastly, the potential influence of peer treatment on individual students' outcomes needs to be considered within the model, as was done in this analysis.

The contribution of this analysis to ZTP research is not only the causal estimate, but also the methodology introduced in the study. The methodology described with this study provides the ZTP educational researcher a template for obtaining causal estimands with observational data.

C. Implications for Policy Applications of Causal Modeling Using HLM with PS Estimates

The lack of randomization of school policy adoption and of students' assignment to specific schools is an inferential limitation of educational research in observational settings. If assignment to treatment is not controlled by randomization, non-equivalent groups are usually produced. Educational research also needs to take into account the natural nesting of students within schools, and the non-random assignment to treatment of school-level policy adoption when estimating causal effects of any policy.

Propensity score estimates, at the appropriate level of treatment assignment within a hierarchical linear model, is a technique available to educational researchers that can produce unbiased estimates with observational data. These methods were described in detail in this study for further use by other researchers.

As with all modeling, the assumptions of the modeling techniques employed need to be considered. When propensity score estimation is used with observational data, the only assumption necessary to interpret the estimates causally is the assumption of ignorability of treatment assignment. Ignorability of treatment assignment can be assumed if all the covariates that affect the treatment assignment and the outcome variable have been accounted for. If ignorability holds, one can obtain unbiased treatment effect estimates. The Add Health data sample provided a rich list of covariates theoretically believed to be associated with both treatments (school adoption of ZTP, and student expulsion) and cognitive outcome measures. However, if a potentially confounding covariate was omitted from either model, the assumption of ignorability does not hold, and the estimates are potentially inaccurate. Ignorability is a largely untestable assumption, supported by theory (Foster, 2003).

D. Suggestions for Further Research

The Add Health data set had many of the student-level covariates considered to affect treatment assignment and outcome necessary for propensity score modeling at the student level, and also grouped students within a rich school-level data set that provided the school-level pretreatment covariates of interest, and treatment assignment mechanism adoption of zero tolerance policy. However, the data did not define suspension from school at a level that would have made the variable useful in modeling. An out of school suspension of over 15 days is considered an expulsion (Skiba, et al., 2003). The data did not contain length of suspension except a measure of greater than a few days; this left the data uninformative as to whether a suspension was indeed an expulsion by definition. The expulsion data was self-reported (possibly adding bias to the measure), and the data on the number of expulsions per student was not collected (failing to account for differences in dosage of treatment between expelled students) in this sample. ZTP research with a sample that contained more explicit student-level data and operational definitions regarding expulsion would produce valid estimates of ZTP on at-risk students (subpopulations A and B). There was missing data on many of the covariates theoretically believed to affect the treatment and outcome variables at both levels. The missing data may have left a sample of schools and students not representative of the initial populations of interest.

The definition of a zero tolerance policy school needs to be universally accepted within the educational literature. The U.S. Department of Education defines zero tolerance weapons policies in both Sec. 14601 of the Elementary and Secondary Schools Act (ESEA) – Gun–Free Requirements (the GFSA of 1994, from the Improving America's Schools Act of 1994) and the No Child Left Behind Act (NCLB) Sec. 4141 of 2001 (Potts et al., 2003). The GFSA specifies that each state receiving federal funds under this Act needs to have a law that requires schools to impose a minimum one year out of school expulsion on any student found to have brought a weapon to a school. Zero tolerance, as it relates to behavior and discipline, has been defined by Potts, Njie, Detch, & Walton (2003) as the policy of not tolerating undesirable behavior, such as violence or illegal drug use and possession, with the automatic imposition of severe penalties for first offenses.

The National Center for Education Statistics reports schools that have one or more zero tolerance policies in place (Heaviside, Rowand, Williams, & Farris, 1998). These are not zero tolerance policy schools; zero tolerance policy schools are schools with a zero tolerance to all drugs and weapons violations. If zero tolerance for one or more of any violation of school policy were used as the definition of a ZTP school, then every school in the Add Health sample would have been identified as a ZTP school, though the levels of ZTP vary greatly between these schools.

Estimating causal effects of ZTP policy needs to account for the non-random treatment assignment of school policy adoption, the non random assignment of students to expulsion, and the natural nesting of school data. Replication of the methods used in this analysis on a different sample would provide additional estimates of the impact of this policy adoption.

APPENDIX A Equation Notation Key

Symbol and Equation Key Estimates

z_i	individual's treatment assignment
V	the treatment assignments of all other students within an individual's school, except the individual's assignment
$\mathbf{s^*}$	the vector of school assignments observed in the sample
$\mathbf{X}$	vector of student-level covariates
$\mathbf{W}$	vector of school-level covariates.
q_1	student's probability of being expelled under a zero tolerance policy
q_0	student's probability of being expelled under a non-zero tolerance policy.
Q	probability of a school being assigned to a tolerance policy
δv_0	conditional effect of attending a zero tolerance policy school versus a non-zero tolerance policy school for students at no risk for expulsion in either school type
δz_1	conditional effect of expulsion under the zero tolerance policy for students actually attending zero tolerance policy schools.
Δz_0	conditional effect of expulsion under a non-zero tolerance policy for students actually attending non-zero tolerance policy schools.
logit $\hat{Q}j$	log-odds of adopting a zero tolerance policy for school j [the log-odds is defined as the natural logarithm of the probability of ever adopting a zero tolerance policy divided by 1 – (the probability of ever adopting a zero tolerance policy); logit $\hat{Q}$j = ln {p ($\hat{Q}$j)/[1- p ($\hat{Q}$j)]}
β_0	expected value of the log-odds for schools to be zero tolerance schools when all of the predictor variables equal zero.

β_{1}.... β_{n}	estimated change in the log- odds of school being a zero tolerance policy school for every unit change in the independent variables in model (school-level covariates and aggregated student covariates)
Logit $\hat{q}_{1ij}$	log-odds of being expelled for student *i* in zero tolerance policy school j [the log-odds is defined as the natural logarithm of the probability of ever being expelled in zero tolerance policy school j divided by 1- (the probability of every being expelled in zero tolerance policy school j); Logit $\hat{q}_{1ij}$ = ln{p($\hat{q}_{1ij}$)/]1- p($\hat{q}_{1ij}$)] }
β_{0j}	average log – odds for students of being expelled in zero tolerance policy school for x =0
β_{1j} .... β_{nij}	estimated change in the log- odds of a student being expelled in a zero tolerance policy school for each one unit change in the independent variable, while other variables held constant.
Logit $\hat{q}_{0ij}$	log-odds of being expelled for student *i* in non-zero tolerance policy school j [the log-odds is defined as the natural logarithm of the probability of ever being expelled in non-zero tolerance policy school j divided by 1- (the probability of every being expelled in non-zero tolerance policy school j); Logit $\hat{q}_{0ij}$ = ln{p($\hat{q}_{0ij}$)/[1- p($\hat{q}_{0ij}$)] }
β_{0j}	average log – odds for students of being expelled in non-zero tolerance policy school for x = 0
β_{1j} .... β_{nij}	estimated change in the log- odds of a student being expelled in a non-zero tolerance policy school for each one unit change in the independent variable, while other variables held constant.
γ_{0}	NON ZTP students in Propensity stratum 1 intercept,
δ_{v0}	Effect of attending ZTP schools

γ_2	Effect Propensity stratum 2
γ_3	Effect Propensity stratum 3
γ_4	Effect Propensity stratum 4
γ_5	Effect Propensity stratum 5
u_j	Random effect, School specific
e_{ij}	Random effect, Student specific

APPENDIX B School-Level Covariates

School-level Covariates Add Health data Wave I from School Administrator and Student and Parent Surveys

Covariate	
SCID	School ID
Dr_W_ZTP (ZTP)	ZTP School (1st time drug, weapon offenses=expulsion)
Catchme	School is a catchment area school (takes all students in geo area)
A7	Average class size
A12	Percentage of New Teachers
A13	Percentage of Teachers with Masters
Ln_mehhl	Average of the log of mean household income per school
Subdinc	Percentage of households with subsidized government income
Engprim	Percentage of households with English as primary language
Atype	Type of school: public or private
Rural	Dummy variable 1 location (urban, rural, suburban)
Suburb	Dummy variable 2 location (urban, rural, suburban)
Neast	Dummy variable 1 region (West, Neast, Midwest, South)
Midwest	Dummy variable 2 region (West, Neast, Midwest, South)
South	Dummy variable 3 region (West, Neast, Midwest, South)
Jhsandhs	Dummy variable 1 grade levels (HSonly, JHS and HS, JHonly)
Jhonly	Dummy variable 2 grade levels (HSonly, JHS and HS, JHonly)
Avgatt	Dummy variable 1 Daily attendance (Hiatt:>96%; Avgatt: 90-95%; Loatt:78-89%)
Loatt	Dummy variable 1 Daily attendance (Hiatt:>96%; Avgatt: 90-95%; Loatt:78-89%)
Hidropout	High dropout rate (>3% any year)

APPENDIX C Student-Level Covariates

Student-Level Covariates Add Health data Wave I, Wave III from Student and Parent Surveys

bio_sex	Biological sex
h1gi8	Race Single Category
h1da1	Do you work around the house?
h1da8	# of Hours per week watch television
h1gh23j	Does not eat breakfast
h1ed3	Has ever skipped a grade
h1ed5	Has ever repeated a grade
h1ed11	Past Semester Grade in English
h1ed12	Past Semester Grade in Math
h1ed20	I feel like a part of my school
h1ed24	I feel safe at my school
Rmomed	Residential mother completed years of education
H1wp6	I make own decisions weeknights how to spend time
H1wp8	I frequently eat dinner with at least one parent
H1pf5	I have good relationship with my father
H1pf27	I seldom get sick
H1co1	Have you ever had sex?
H1to1	Have you ever smoked a cigarette?
H1to29	I have 3 friends who drink alcohol more then once a month
Smokemj	Have you ever smoked marijuana?
H1to51	There is easy access to alcohol in my home
H1ds15	I have been loud and rowdy in public places
H1fv6	Were you ever jumped?
H1pr1	I feel adults care about me.
H1pr2	I feel teachers care about me.
H1pr4	I feel friends care about me.
H1ee1	I want to attend college
H1ee4	The # of hours worked per week during non-summer
Pa55	Total household income
Pa10	Parents marital status
Codgrexp (Wave I treatment variable)	Expelled or not in current school
AH_PVT (Wave III outcome variable)	Add Health PVT score

APPENDIX D

Empirical Logit q_1 Model Coefficients for Identification of Subpopulation C

Variable	B estimate	Exp of B estimate
BIO_SEX	.99	2.70
H1GI8	.21	1.24
H1DA1	-.07	.93
H1DA8	.02	1.02
H1GH23J	-1.04	.35
H1ED3	2.53	12.49
H1ED5	.80	2.23
H1ED11	.35	1.41
H1ED12	-.10	.91
H1ED20	.43	1.53
H1ED24	-.47	.63
RMOMED	-.50	.61
H1WP6	.38	1.40
H1WP8	-.27	.76
H1PF5	.05	1.06
H1PF27	-.04	.97
H1CO1	.15	1.16
H1TO1	-.31	.73
H1TO29	.26	1.30
SMOKEMJ	.99	2.70
H1TO51	-1.26	.29
H1DS15	.30	1.35
H1FV6	.72	2.05
H1PR1	.08	1.09
H1PR2	-.18	.83
H1PR4	.11	1.11
H1EE1	.18	1.20
H1EE4	-.02	.98
PA55	.00	1.00
PA10	.13	1.14
Constant	-5.85	.00

REFERENCES

Allen, Z. (2004, January 15). Zero tolerance school discipline derails Black & Latino students. *New York Amsterdam News*, *95*, 1-45.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Angrist, J. D. (1990). Lifetime earnings and the Vietnam era draft lottery: Evidence from Social Security administrative records, *American Economic Review, 80,* 313-335.

Cochran, W. G. (1968). The effectiveness of adjustment by sub classification in removing bias in observational studies. *Biometrics, 24,* 295-313.

Ewing, C. P. (2000, January/February). Sensible zero tolerance protects students. *Harvard Education Letter*. Retrieved August 10, 2009 from http://www.edlettr.org/past/issues/2000-jf/zero.shtml

Farrell, N. (2005). *Middle school students' experiences with discipline: A contextual look at neighborhood, adult, and school connectedness*. Unpublished doctoral dissertation, Loma Linda University.

Feda, D. (2008). *Written violence policies and assault deterrents in Minnesota schools: impact on educators' risk of physical assault*. Unpublished doctoral dissertation, University of Minnesota.

Fisher, R. A. (1925). *Statistical methods for research workers***.** Edinburgh, UK: Oliver and Boyd.

Foster, M. (2003). Propensity score matching an illustrative analysis of dose response. *Medical Care, 41,* 10, 1183-1192.

Garner, C., & Raudenbush, S. (1991, October). Neighborhood effects on educational attainment: a multilevel analysis. *Sociology of Education*, *64*, 251-262.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Giroux, H. (2003). Racial injustice and disposable youth in the age of zero tolerance. *International Journal of Qualitative Studies in Education (QSE), 16,* 553-565.

Giroux, H. (2003). Zero tolerance, domestic militarization, and the war against youth. *Social Justice*, *30,* 59-65.

Heaviside, S., Rowand, C., Williams, C., & Farris, E. (1998). *Violence and discipline problems in U.S. public schools: 1996-97*. (NCES 98-030). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Hill, J. (2008). Comment: Can Nonrandomized Experiments Yield Accurate Answers? A randomized experiment comparing random and nonrandom assignments. *Journal of American Statistical Association, 103,* 1346- 1350.

Hong, G., & Yu, B. (2008). Effects of kindergarten retention on children's social-emotional development: An application of propensity score method to multivariate, multilevel data. *Developmental Psychology, 44,* 407-421.

Hong, G., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multi-level observational data. *Journal of the American Statistical Association*, *101*, 901-910.

Imai, K, & Van Dyk, D.A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association, 99,* 854-866.

Love, T.E. (2004, October). *Using propensity score methods effectively.* Paper presented to the Cleveland Chapter of American Statistical Association, Cleveland, OH.

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

Pfefferman, D., Skinner, C. J., Homes, D. J., Goldstein, H., and Rasbash, J. (1998). Weighting for unequal selection models in multilevel models. *Journal of the Royal Statistical Society, 60*, 23–40.

Raudenbush, S. & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods.* (2nd Ed.). Sage Publications: London.

Raudenbush, S. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational Statistics, 18*, 321–349.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using sub classification on the propensity score. *Journal of the American Statistical Association, 79*, 516-552.

Rosenbaum, P.R. (2002). *Observational studies*. (2nd Ed.) New York: Springer.

Rubin, D.B. (1986). Comment: Which ifs have causal answers? *Journal of the American Statistical Association, 81*, 961–962.

Rubin, D.B. (1990). “[On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.] Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science, 5*, 472–480.

Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*, 688-701.

Rubin D. & Thomas N. (1992). Characterizing the effect of matching using linear propensity score methods with normal distributions. *Biometrika, 79,* 797–809.

Rubin, D. (2004). *Basic concepts of statistical inference for causal effects in experiments and observational studies.* Unpublished manuscript, Department of Statistics, Harvard University, Cambridge, MA.

Schoonover, B. J. (2007, October). *Customer satisfaction: Communication training and the help-desk* hot-*line*. Paper presented at the annual meeting of the National Conference on Safe Schools, Washington, D.C.

Shadish, W.R., Clark, M.H., & Steiner, P. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of American Statistical Association*, *103*, 1334-1343.

Skiba, R., & Knesting, K. (2001). Zero tolerance, zero evidence: An analysis of school disciplinary practice. *New Directions for Youth Development*, *92,* 17-43.

Skiba, R. J. & Peterson, R. L. (1999). The dark side of zero tolerance: Can punishment lead to safe schools? *Phi Delta Kappan, 80,* 372-382.

Skiba, R.J., Simmons, A.B, Staudinger, L.P., Rausch, M.K., Dow, G. & Feggins, L.R. (2003). *Consistent removal: Contributions of school discipline to the school-prison pipeline*. Paper presented at the Harvard Civil Rights Conference School-to-Prison Pipeline Conference, Cambridge, MA.

Winship, C., & Sobel, M. (2004). Causal inference in sociological studies. In M. Hardy, & A. Bryman (Eds.), *Handbook of data analysis* (pp. 481-504). Thousand Oaks, CA: Sage Publications.

Yanovitzky, I., Zanutto, E., & Hornik, R. (2005). Estimating causal effects of public health education campaigns using propensity score methodology. *Evaluation and Program Planning*, *28*, 209–220.