Educational and criminal justice outcomes 12 years after school suspension

Janet Rosenbaum, Ph.D.

Department of Epidemiology and Biostatistics, School of Public Health,
SUNY Downstate Medical Center, Brooklyn, NY 11203

Corresponding author: Janet Rosenbaum, Department of Epidemiology and Biostatistics, School of Public Health, SUNY Downstate Medical Center, 450 Clarkson Ave, MS 43, Brooklyn, NY 11203. Tel: 347-557-1112, Fax: 718-270-2533, janet@post.harvard.edu.

Acknowledgements
The author thanks Constantine Frangakis, Fran Goldscheider, Sandra Hofferth, Clea McNeely, Stephen Raudenbush, James Rosenbaum, and anonymous reviewers for helpful conversations. This research was funded by the Spencer Foundation (grant #201000138), the Eunice Kennedy Shriver National Center for Child Health and Human Development grant R24-HD041041 (Maryland Population Research Center), and the SUNY Downstate School of Public Health. The funding sources did not have any role in any aspect of the study design, implementation, writing, or manuscript submission.

Conflict of Interest: The author declares that she has no conflict of interest.

Contributors' Statement:
JR conceptualized and designed the study, carried out the analyses, wrote the manuscript, and approved the final manuscript as submitted.

Abstract

A third of US students are suspended over a K-12 school career. Suspended youth have worse adult outcomes than non-suspended students, but the relationship may be due to which students are selected to be suspended. This study examines whether first-time suspension in 1995 leads to different educational and criminal justice outcomes for suspended versus matched non-suspended youth from the National Longitudinal Study of Adolescent and Adult Health (n=480 suspended, n=1193 non-suspended). The suspended and non-suspended youth were matched on 60 pre-suspension socioeconomic, educational, and health variables to minimize the possibility that outcomes are attributable to these variables or unobserved variables associated with the 60 variables. In 2008, at ages 25-32, suspended youth were less likely than matched non-suspended to have earned bachelors degrees, and more likely to have been arrested and on probation, which suggests that suspension rather than selection explains these outcomes. Causal mediation analysis in the matched sample suggests that expulsion at 5 years mediates the effect of suspension on high school graduation at 12 years, especially for Black youth, consistent with the secondary deviance hypothesis that the social consequences of initial deviance results in more severe deviance.

Keywords: adolescence; deviance; suspension; educational attainment; school discipline.

Abbreviations:

GED general equivalency degree

BA bachelor of arts (or other 4-year college degree)

AA associate of arts (or other 2-year college degree)

SES socioeconomic status

STI Sexually Transmitted Infection

School suspension is used widely, affects many students, and begins as early as preschool (Gilliam and Shahar, 2006). Over a school career from kindergarten to 12th grade, 35% of students are suspended; among Black students, 67% of males and 45% of females are suspended at least once during a K–12 school career (Shollenberger, 2015).

School discipline policies have utilitarian goals, including obtaining better behavior from the punished student and maintaining school norms. Most evidence suggests that suspension has no effect on students' behavior other than by temporarily removing disruptive students from schools (Kinsler, 2013; Cook et al., 2010), although one study of North Carolina middle school students suggests that suspended students are more likely to comply with school rules in the school year that they were suspended (Kinsler, 2013). Suspension may also improve school climate and by reducing peer influences to engage in deviant behavior (Zimmerman, 2014).

Despite these positive effects, suspension predicts greater deviance, more involvement with the criminal justice system, and lower rates of school completion in both the short and long-term. Youth are more likely to be arrested both during the month of suspension (Monahan et al., 2014) and within a year of suspension (Hemphill et al., 2006a). Within a year of suspension, suspended youth are also more likely to engage in antisocial behavior (Hemphill et al., 2006b, 2013) and use marijuana (Evans-Whipp et al., 2015) and tobacco (Hemphill et al., 2012). In a 13-year national longitudinal survey, youth suspended for at least 10 days were less likely to graduate high school and more likely to be arrested and incarcerated by the end of the study (ages 26–31) (Shollenberger, 2015). In a 7–8 year longitudinal study in Florida, youth suspended in 9th grade were less likely to graduate high school, graduate on time, and enroll in post-secondary education, and more suspensions predicted worse outcomes (Balfanz et al., 2015).

School suspension is characterized by racial disparities. The white-black disparity has declined for achievement but increased for suspensions: from 1972 to 2012, the proportion of all students suspended for at least one day increased from 3% to 5% for white students and from 6% to 16% for black students (Wald and Losen, 2003; Losen et al., 2015). Among secondary students, the gap is larger and has widened more over the same period: from 6% to 7% among whites and from 12% to 23% among Blacks (Losen et al., 2015). Racial disparities may stem from discrimination by teachers and administrators, rather than differences in students' behavior. Psychological experiments using vignettes that manipulate a hypothetical student's race find that teachers punish black students more harshly than white students for the same infraction (Okonofua and Eberhardt, 2015), and students seem to perceive such racial disparities in suspension (Ruck and Wortley, 2002). A cross-sectional study in North Carolina suggests that apparent racial disparities in suspension are attributable to stricter school suspension policy and that teachers and administrator race does not predict student suspension (Kinsler, 2011). A nationally representative study found that teacher reports of misbehavior in kindergarten, first, and third grade accounts for the racial gap in school suspension in 8th grade (Wright et al., 2014). Racial disparities in suspension are problematic in themselves but also predict racial disparities in school completion. One nationally representative study found that school suspension makes the largest

contribution to the widening of the black-white high school drop-out gap between 1979 and 1997 (Suh et al., 2014).

For over a decade, the American Psychological Association, American Association of Pediatrics, and American Bar Association have criticized zero tolerance school suspension policies for potentially reducing educational attainment, harming employment prospects, increasing risk behavior, and increasing criminal justice involvement (Reynolds et al., 2008; of Pediatrics Committee on School Health, 2003; Lamont et al., 2013) and for creating injustice through mandatory minimum sentences that do not permit judicial discretion (ABA, 2001). In place of out-of school-suspension, the American Association of Pediatrics recommends Positive Behavioral Interventions and Supports (Lamont et al., 2013) because initial studies suggest that schools can replace suspension with positive reinforcement and a larger range of consequences for misbehavior (Cook et al., 2010). More recently, large cities have made isolated attempts to reduce school suspension at the local level (Barnhart et al., 2008). Evidence of the negative impacts of school suspension (Fabelo et al., 2011) prompted the creation of federal and state statements about excessive school suspension and initiatives to promote alternatives, such as the federal Supportive School Discipline Initiative in 2011. The federal government has urged reducing school suspension in preschool (HHS, 2014), and the New York State government has called for eliminating school suspension in preschool beginning in 2017–18 (DeLorenzo and Rider, 2015). Legal scholars have also criticized zero tolerance school suspension for depriving students of the right of equal access to education (Bitner, 2015).

Despite these critiques, school suspension continues to be common nationwide (Losen et al., 2015). Most school districts have discipline codes, and school boards set punishments for poor discipline. Educators justify suspension as a short-term punishment. Many suspensions are under 2 weeks long because suspensions longer than 2 weeks are subject to stricter legal protections (Arum, 2003). Schools treat suspension as an appropriate response to "insubordination, habitual indolence, [or] disorderly conduct," and hearings may be unnecessary or informal (Garson, 2010; Arum, 2003; Skiba et al., 1997). A suspension informs students that certain behaviors are not tolerated. School administrators may report no adverse reactions, but administrators' observations are often short-term, and long-term implications are hard to see.

Hypotheses

School suspension may result in short-term compliance. Administrators may view suspended students' short-term compliance after suspension as the first step towards long-term socialization and internalization of ethical and social norms. Suspended students may be outwardly compliant with rules in the short term, but suspension may also create delayed mechanisms that increase long-term deviance, as observed in the literature. One possible mechanism is "secondary deviance," a process in which initial deviance induces further deviance due to social ramifications of the suspension, such as labeling, stigma, limiting options, or creating separation, thus magnifying the impact of the initial deviance. Suspension for initial deviance can lead to further and more lasting deviance that may not have occurred otherwise, even if the initial deviance were minor,

not premeditated, or a one-time experiment (Becker, 1963; Lemert, 1967; Paternoster and Iovanni, 1989). Recent research finds that youth who are arrested or stopped by police are more likely to engage in secondary deviance as a result of labeling (Wiley et al., 2013; Liberman et al., 2014). School suspension may create a similar process, which is supported by studies that describe a "school-to-prison pipeline" (Nicholson-Crotty et al., 2009). Suspended students and their parents/caregivers suggest that they feel more disengaged from school after a suspension, and students report that they did not improve their behavior after a suspension (Michail, 2012; Gibson and Haight, 2013), but student reports may be distorted.

Alternatively, a problem in much available research is that negative effects observed among suspended youth may be attributable to selection into suspension, rather than the suspension itself. This possibility is supported by an Australian longitudinal study that did not find differences in educational attainment two years after suspension. The study attributed the association between suspension and lower educational attainment in other studies to selection bias (Cobb-Clark et al., 2015). According to this selection bias hypothesis, suspended youth differ from non-suspended youth in pre-suspension risk-taking and low socioeconomic status. Teens with these risk factors may have lower educational attainment and greater chances of criminal justice involvement than teens without the risk factors, whether or not they are suspended. Unfortunately, existing research is inadequate for reducing selection bias because statistical adjustment methods such as regression may not address confounding adequately (Rubin, 1997; Berk, 2010). The matched sampling methods used in this study can minimize confounding on matched variables and on unobserved variables associated with the matched variables.

This study compares the outcomes 5 and 12 years after a first suspension for youth suspended for the first time in 1995-96 with those of comparable youth not suspended in that time interval or before. The study uses matched sampling methods so that suspended and non-suspended youth are comparable prior to suspension to minimize confounding on observed factors. Minimizing confounding in this way allows the study to discriminate between the secondary deviance and selection bias hypotheses. To clarify temporal ordering of events, the study is distinctive in focusing on students who had never been suspended at baseline, which avoids bias from unobserved suspension history (Kinsler, 2011), allows the study to observe first suspensions, and preserves temporal ordering between control variables and suspension. Including students with previous suspensions in the analyses would prevent matching on pre-suspension factors because these students' pre-suspension deviance behaviors of students and other processes related to selection into suspension would be unknown. Such analysis also could not exclude the possibility that deviant behavior of students with a previous suspension was due to the previous suspension.

This study finds that suspended youth have worse education and criminal justice outcomes than matched non-suspended youth. The outcomes for employment are mixed, which suggest a more complex process. The greater likelihood of assigning suspension to Blacks, and the different factors that predict assignment into suspension, such as greater chances of suspension for tall Black males, concurs with findings of racial discrimination in psychology experiments (Okonofua and Eberhardt, 2015).

Methods

Data

The National Longitudinal Study of Adolescent and Adult Health (Add Health) comprises a nationally representative sample of adolescents who were attending public and private high schools and their feeder middle schools in 1994–95. Adolescents not attending school were excluded from the sample, including home-schooled adolescents. Adolescents with disabilities, Blacks with college-educated parents, among other groups, were oversampled at wave 1 (Tourangeau and Shin, 1999).

The data came from the Add Health surveys given to adolescent respondents in 1995 (wave 1, response rate 79.0%), 1996 (wave 2, response rate 88.6%), 2001 (wave 3, response rate 77.4%), and 2008 (wave 4, response rate 80.3%), their parents (93% female parents) in 1995 (response rate 82.5%), and school administrators in 1995 (response rate 97.7%.) (National Longitudinal Study of Adolescent Health, 2015)

Respondents were a 9593 person subsample who participated in the first two surveys, at least one of the subsequent surveys (wave 3 and/or 4), and reported that they had never received an out-of-school suspension or been expelled from school at baseline (Item wording: "Have you ever received an out-of-school suspension from school?" and "Have you ever been expelled from school?") Limiting the analyses to never-suspended students avoids bias from unobserved suspension history (Kinsler, 2011), allows the study to observe incident suspensions, and preserves temporal ordering between control variables and suspension.

Because some evidence suggests that Blacks are treated differently, the same analysis was repeated in a subsample of 1719 Black youth who had never been expelled or suspended from school at baseline. The sample selection and sample sizes are shown in Figure 1.

The survey weights were not used because the Add Health survey weights were developed for the entire sample, based on both probability of selection and probability of response (i.e., post-stratification). The weights include post-stratification weights, so using survey weights with a subsample will cause the variance to change in unpredictable ways; the Add Health survey advises researchers not to use the survey weights with sub-samples (Chantala and Tabor, 2010). This study uses a highly constrained sub-sample: respondents who were never suspended at baseline and matched on dozens of variables. Using the survey weights would cause inaccurate estimation of standard errors and reduce sample size because only 74% of the Add Health sample has available survey weights.

Predictor

The predictor of interest is self-reported suspension between 1995 and 1996, based on the wave 2 question, "During this school year (during the 1995–96 school year) did you receive an out-of-school suspension from school?"

Control variables

The control variables were potential confounders of the relationship between suspension and educational/criminal justice outcomes that may be associated with both the predictor (suspension) and the outcomes. These potential confounders were identified from past research about suspension (Shollenberger, 2015; Losen et al., 2015) and educational attainment (Bowen et al., 2011), and arrest (Ou et al., 2007), as well as from experts and peer reviewers of this paper.

The control variables were all measured at baseline, except for father ever in prison, which was not measured until 2001. The 2001 father in prison measurement was used as a control variable because it was not likely to be a consequence of a child's being suspended from school. The father could have gone to prison after the child's school suspension, but it is reasonable to assume that the father's likelihood of going to prison existed prior to the child's school suspension.

For ease of reading, the control variables have been organized into categories: demographics, socioeconomic status, educational achievement, parents' risk behavior, substance use, personality, delinquency and adverse experiences, appearance, relationship with parent, physical and mental health, and environmental context.

The demographics category comprises 11 variables: gender, male-black interaction term, age, Latino ethnicity, Asian and Black race, nativity, whether the respondent's primary home language is English, and region of country.

Socioeconomic status (SES) includes 6 variables: parent is high school grad, college grad, parent-reported household income (log scale), parent-reported enough money to pay bills, parent receives public assistance, and parent is currently employed.

Educational achievement includes 9 variables: standardized test score (Add Health Peabody Vocabulary Test), expectations to attend college, whether the respondent attends a private school, grade point average (average of 4 self-reported grades, alpha=0.72), school is strict on substance use (top quartile of administrator-reported school discipline policy for alcohol, drugs, and smoking, 8 items, alpha=0.97), school is strict on civil order (top quartile of administrative-reported school discipline policy for offenses such as stealing school property and verbally abusing a teacher, 7 items, alpha=0.73), positive expectations for the future (aggregate variable of 5 items such as will not be killed by age 21, will live to age 35, alpha=0.61), and school attachment (aggregate variable of 9 items including feeling safe at school, problems with teachers, problems completing homework, alpha=0.78).

The parents' risk behavior category includes 4 variables: parent-reported parent smoking, household member smokes, binge drinking, and one item from 2001: whether the respondent's father was ever in prison.

Substance use includes 4 factors: lifetime marijuana use, lifetime cocaine use, regular smoking status, and friends' substance use (number of friends who drink alcohol monthly, use marijuana monthly, smoke daily, 3 items, alpha=0.72).

Personality includes 8 variables: self-esteem (aggregate of 11 factors modified from Rosenberg's scale, alpha=0.88), conscientiousness (aggregate of 5 items describing systematic approach to solving problems, alpha=0.78), systematic versus gut-feeling decision-making was measured by the Likert item, "When making decisions, you usually go with your 'gut feeling' without thinking too much about the consequences of each alternative." where higher means more systematic decision-making style, emotional stability (aggregate of 6 items including have a lot of good qualities, a lot to be proud of, alpha=0.87) (Young and Beaujean, 2011), agreeableness (sum of 3 items: never argue with anyone, never get sad, never criticize other people, alpha=0.63), personal control was measured by the Likert-scale item "When you get what you want, its usually because you worked hard for it.", problem avoidance was measured by the Likert-scale item "You usually go out of your way to avoid having to deal with problems in your life.", and "Difficult problems make you very upset.".

The delinquency and adverse experiences category includes 4 variables: delinquency was the sum of 15 binary items including running away, hurting someone so badly that they needed medical care, participating in a group fight, lying to parents, and stealing <$50 and ≥$50 (alpha=0.80); experiences with violence was the sum of 8 binary variable such as saw shooting, was shot, shot another (alpha=0.75.); number of truant days in past year; and never truant in the past year.

Appearance includes 4 factors: having a permanent tattoo, height, height-male interaction term, and interviewer assessment of appearance (attractive, personality attractive, well-groomed, 3 items, alpha=0.74).

Relationship with parent includes 4 factors: parent's assessment of relationship with child (how is child's life going, get along with child, trust child, child doesn't have a bad temper, 4 items, alpha=0.67), parental closeness (aggregate of 14 items such as perceived love and warmth, satisfaction with relationship, alpha=0.81), talk with mother (talk with mother about social, personal, school issues, 4 items, alpha=0.62), parental monitoring (parents let respondent make own decisions about weekday bedtime, weekend curfew, how much TV, 7 items, alpha=0.70).

The physical and mental health category includes 6 factors: modified Center for Epidemiologic Studies Depression screen score (19 items, alpha=0.86), history of sexually transmitted infections (STI), having ever been pregnant, having ever had sexual intercourse, sufficient sleep, and number of people they know who have had an STI.

Environmental context included 1 factor: neighborhood support (e.g., know most of the people in the neighborhood, average of 4 binary items, alpha=0.72).

Gottfredson and Hirschi's self-control theory of crime is relevant for studying school suspension (Gottfredson and Hirschi, 1990), but Add Health data lacks a set of items with clear face validity intended to measure self-control. At least three approaches have been used in the past 12 years (Perrone et al., 2004; Beaver, 2011; Wolfe and Hoffmann, 2016), but many other approaches to measuring self-control are possible (Wolfe and Hoffmann, 2016). This analysis measured self-control using six constructs:

personal control (1 item), gut-decision making style (1 item), school attachment (9 items), conscientiousness (5 items), agreeableness (3 items), and parent's assessment of child (4 items). All 21 measures that comprise the most recent psychometric derivation of the self-control construct in Add Health (Wolfe and Hoffmann, 2016) appear in these factors. These factors have clear face validity and address other potential confounders.

Outcomes

The outcomes are derived from the suspension research literature and the statement on school suspension of the American Association of Pediatrics (Lamont et al., 2013): educational attainment, criminal justice involvement, and job qualities. Educational attainment in 2001 included attainment of a high school diploma (not including equivalency degree (GED)) or bachelor's degree (BA) and having ever been expelled, and four variables in 2008: attainment of a high school diploma (not including GED), associates' degree (AA), BA, and graduate (post-BA) degree.

Criminal justice involvement measured in 2001 included having ever been arrested, having been arrested or convicted as a minor, having been arrested or convicted as an adult. Criminal justice involvement measured in 2008 included having been arrested once, having been arrested 2 or more times, having ever been in prison, and having ever been on probation. Job qualities were measured in 2008 and included perceived socioeconomic status, household income, and personal earnings (3 continuous variables) and whether the respondent is currently employed, has ever been fired from a job, and whether their current or most recent job has health, retirement, or vacation benefits, and the level of strenuousness (very strenuous, moderately, lightly, or a desk job).

Data analysis

We analyzed the data using the R 3.3.0 and Stata 11.2 statistical packages.

Factor analysis

We used standard factor analysis procedures to derive all multi-item measures, requiring that factor loadings be at least 0.4. To improve the quality of matching, factor analysis decisions avoided the overuse of data reduction.

Bivariate analysis

We identified factors where suspended and non-suspended youth differed most using standardized differences, a measure of effect size considered to be significant if they are greater than 0.2, medium if greater than 0.5, and large if greater than 0.8, and statistically insignificant if they are less than 0.1.

Matching sampling procedure

Matched sampling refers to a number of statistical methods for creating a comparison group of non-suspended youth similar to suspended youth using variables

prior to suspension. The specific matched sampling method is selected through trial and error by its ability to construct a similar comparison group, rather than derived according to criteria known in advance (Morgan and Winship, 2015).

Sixty potential confounders of the relationship between school suspension and each outcome were identified using literature review and expert feedback. The matched sampling method that achieved balance was 3:1 exact and nearest-neighbor Mahalanobis matching with replacement, within propensity score calipers of 0.25 standard deviations, using the R library (Ho et al., 2008). Matched sampling achieved balance on 60 variables plus the estimated propensity score in the general sample and 55 variables plus the propensity score in the Black subsample.

In the full sample, 3 non-suspended youth were matched to each suspended youth using the following procedure. For each suspended youth, exact matching reduced the set of eligible non-suspended youth by requiring that only non-suspended youth with the same daily smoking status and ever-marijuana status could be considered. Propensity calipers reduced the set of eligible non-suspended youth further to those within 0.25 standard deviations of the estimated propensity score. Finally, Mahalanobis matching identified the 3 closest youth according to a correlation-adjusted distance measure of age, grade point average, and delinquency scores.

In the full sample, propensity scores were estimated from a logistic regression predicting a first suspension from demographic factors (male gender, age, born in US, Latino, Asian, and Black race/ethnicity, home language is English), SES factors (mother high school graduate, mother college graduate, parent is currently employed, parent-reported household income, parent-reported enough money to pay bills, father ever in prison [2001]), health and risk behavior factors (experiences with violence, respondent smokes daily, household member smokes, mother smokes, mother binge drinks, respondent smokes daily, depression score, positive expectancies, respondent sleeps "enough"), educational factors (standardized test score, school attachment, expect to attend college, attend private vs. public school, school is strict on civil order), and personality factors (parent's assessment of their child, agreeableness, emotional stability, parental closeness, systematic vs. gut-feeling decision making).

The literature suggests that Black youth are suspended disproportionately, particularly for subjective offenses such as insubordination, so the assignment mechanism for Black youth is likely to be different. This analysis includes a separate matched sampling model for the subsample of Black youth. This subsample will have reduced power due to lower sample size, so any significant relationships are particularly noteworthy.

In the Black subsample, 8 non-suspended youth were matched to each suspended youth on age and grade point average. The propensity scores were estimated from a logistic regression predicting a first suspension from demographic factors (male gender, born in US), SES factors (parent-reported enough money to pay bills, father ever in prison [2001]), health and risk behavior factors (household member smokes, positive expectancies, overweight status, delinquency score, ever used marijuana), and

educational factors (school administrator's reported disciplinary policies are strict on substance use and strict on civil order, never truant from school, standardized test score, expect to attend college, attend private vs. public school).

Analysis within matched sample

After matching, the analysis estimated the relative risks of each outcome with a multivariate Poisson working model with robust standard errors within the matched sample, using the weights obtained from the matched sampling procedure. A Poisson model allows coefficients to be interpreted as relative risks, which are more easily interpreted than odds ratios from a logistic regression model and less subject to bias away from the null (McNutt et al., 2003; Zou, 2004; Austin and Laupacis, 2011).

Causal mediation analysis using the R mediation package evaluated whether outcomes at 5 years predicted outcomes at 12 years, and whether mediation was moderated by Black race (Tingley et al., 2013). Causal mediation analyses controlled for race, ethnicity, gender, and age.

Results

Among 9593 youth with no history of expulsion or suspension, 480 were suspended and 54 were expelled for the first time between 1995 and 1996. The one-year incidence of first suspension was 4.5% among non-Hispanic whites, 5.8% among non-black Hispanics, and 6.7% among Blacks, a significant difference (Kruskal-Wallis chi-squared p=0.0005, Pearson chi-squared p=0.004.) For first expulsion, the one-year incidence was 0.3% among non-Hispanic whites, 0.6% among non-Black Hispanics, and 1.3% among Blacks, a significant difference (Kruskal-Wallis and Pearson chi-squared $p < 0.0001$.) The model matched 1193 never-suspended youth to the suspended youth, 30 of whom (2.5%) had weights over 2. The matching model balanced on 60 variables plus the estimated propensity score (Figure 2).

Among 1719 Black youth never suspended or expelled at baseline, 116 were suspended and 23 were expelled between 1995 and 1996. The model matched 568 never-suspended Black youth to the suspended Black youth, 33 of whom (5.3%) had weights over 2. The matching model balanced on 56 variables including the propensity score (Figure 3).

Factors predicting suspension

The factors most strongly associated with suspension included lower grade-point averages, a gut-feeling decision style, parent's low assessment of their child, lower school attachment, lower expectations of college attendance, lower positive expectancies, more daily smoking, more likely to have ever smoked marijuana, more experiences with violence, and higher delinquency scores (Table 1, Figure 2). After matching, suspended and never-suspended youth had similar values of all 60 factors plus the propensity score (Table 1, Figure 2).

Among Black youth, the factors most strongly associated with suspension were

younger age, gut-feeling decision style, lower grade-point averages, and more experiences with violence (Table 2, Figure 3). Factors that predict suspension among Blacks but not the general population include higher agreeableness, strict school substance use policy, and being a tall male. After matching, suspended and never-suspended youth were similar on all 55 factors plus the propensity score (Table 2, Figure 3.)

Outcomes five years after suspension

Comparing outcomes five years after suspension, in 2001, youth suspended for the first time between 1995 and 1996 were 8% less likely to have earned a high school diploma than similar youth who had never been suspended by 1996 (non-suspended youth) and 2.7 times as likely to have been expelled. Among Black youth, suspended youth were 94% less likely to have earned a BA than similar youth who had never been suspended by 1996, and 2.8 times as likely to have been expelled.

Suspended youth were 40% more likely to have been arrested, 94% more likely to have been arrested as a minor, and 3.8 times as likely to have been convicted as a minor than similar non-suspended youth.

Outcomes 12 years after suspension

Comparing outcomes 12 years after suspension, in 2008, youth suspended for the first time between 1995 and 1996 were 6% less likely to have earned a high school diploma and 24% less likely to have earned a BA than similar non-suspended youth.

Suspended youth were 30% more likely to have been arrested once, 51% more likely to have been arrested two or more times, 23% more likely to have been in prison, and 49% more likely to have been on probation than similar non-suspended youth. Among Black youth, suspended youth were 58% more likely to have been arrested once than similar non-suspended Black youth.

Suspended youth were 10% less likely to have a job with retirement benefits, but they did not differ in perceived SES, household income, or personal earnings, or employment history. Among Black youth, suspended youth were 20% less likely to have a job with retirement benefits and 15% less likely to have a job with vacation benefits than similar non-suspended Black youth.

Expulsion, arrest, and conviction as minor in 2001 mediate high school graduation in 2008, and expulsion and high school graduation in 2001 mediate bachelors degree in 2008 (Table 4). Expulsion reported in 2001 mediates 27% of the effect of suspension on high school graduation by 2008, and the effect is modified by Black race. For Blacks, expulsion explains 62% of the effect of suspension on high school graduation, and for non-Blacks, expulsion explains 15% (not shown). Expulsion reported in 2001 mediates 24% of the effect of suspension on ever going to prison by 2008, but the effect is not modified by Black race: for Blacks, expulsion explains 21% of the effect of suspension on ever going to prison, and for non-Blacks, expulsion explains 17% (not shown).

Discussion

Suspension removes youth from school in an attempt to impose a short-term, minor sanction on youth and to create a more orderly school temporarily. This temporary removal may have long-term consequences for those removed that cannot be explained by selection effects. Even within these closely matched groups, suspended youth go on to have lower educational attainment and more involvement with criminal justice system 5 and 12 years later — lower likelihood of high school diploma and BA and higher likelihood of arrest and probation — compared with youth with similar pre-suspension characteristics. Some of the effect of suspension on not graduating high school is attributable to subsequent expulsion for Black but not non-Black students, but other negative outcomes such as arrest and probation are not attributable to subsequent expulsion. The supposedly minor penalty of suspension has a large impact on students' outcomes, even if they are not subsequently expelled.

Suspended and non-suspended youth were matched on 60 pre-suspension characteristics that cover a wide spectrum of background. The observed outcome differences cannot be attributed to selection bias on these characteristics, but also on any characteristics that are correlated with these 60 characteristics.

This evidence is consistent with the secondary deviance hypothesis that suspension for initial deviance results in additional deviance. Sometimes subsequent deviance results in expulsion, but even suspended youth who are not expelled have lower educational attainment and higher involvement in the criminal justice system. The observed effects are too large to be attributed to missing days of school, especially because most suspensions are less than two weeks long for procedural reasons (Arum, 2003). This evidence suggests that school suspension may function similarly to stops and arrest in labeling youth as deviant so that the youth are likely to engage in further deviance, as recent studies suggest (Liberman et al., 2014; Wiley et al., 2013).

Black youth are more likely to suspended for reasons that are not true for the general population. Tall male youth are not more likely to be suspended in the general population, but they are in the Black sub-sample. Teachers' fears may explain why the likelihood of suspension increases with height for Black males but not for Black females. Similarly, strict school discipline policies do not predict greater suspension risk in the general population, but strict policies predict greater suspension risk in the Black sub-sample. Strict suspension policies are justified because it is thought that uniformity reduces subjectivity, but in practice strict policies appear to magnify racial disparities in suspension. These findings that males' height and strict school policies predict suspension in Blacks but not in non-Blacks are consistent with the experimental psychology results suggesting that teachers choose stricter punishments for Blacks than for non-Blacks in identical vignettes (Okonofua and Eberhardt, 2015).

Black youth also seem to have worse effects from school suspension, which could be explained by a greater secondary deviance effect. Suspended Black youth may correctly perceive their suspensions as related to school staffs' over-reactions, rather than their behavior, as is supported by experimental evidence (Okonofua and Eberhardt,

2015). Black youth who are aware that they are treated more harshly by teachers and administrators than non-Black classmates may be more likely to engage in secondary deviance because they perceive that the educational system is racially biased.

Suspended youth have lower educational attainment and are substantially more likely to become involved with the criminal justice system, which is consistent with the contention that suspension facilitates the school-to-prison pipeline (Nicholson-Crotty et al., 2009). Suspension is an important concern for policy makers concerned about the growth of mass incarceration, especially of minorities. Current discussions focus on police-youth interactions, but these findings about suspension suggest that the dynamics of police-youth interactions may be shaped by earlier school-youth interactions and need to be defused earlier in schools.

Twelve years after their first suspension, adults who were suspended have a greater risk of one or 2+ arrests, prison, and probation, and reduced educational attainment, compared with those with similar pre-suspension demographics, socioeconomic status, personality, and grades. Contrary to the stereotype that few individuals engage in deviant behavior, many adolescents experiment with risky or deviant behavior, but this behavior is usually temporary. School disciplinary policy makes normal adolescent experimentation with risky behavior have potentially long-term impacts into adulthood. Suspension could induce secondary deviance: suspended youth may acquire a stigma that leads them to engage in further deviant behavior that is consistent with their new deviant identity. During suspension, suspended youth may meet and socialize with more deviant peers, and begin to engage in further deviant behavior as a result of these peers. Research should examine what activities students do and whom they meet during suspension.

These effects are similar to those in the two published studies of long-term effects of suspension. Balfanz and colleagues found that the likelihood of drop-out increased with the number of suspensions in 9th grade, ranging from 32% for one suspension to 53% for 4 or more suspensions; each suspension reduced the odds of high school graduation by 20% and reduced the odds of post-secondary enrollment by 12% in logistic regressions controlling for attendance, demographics, and grades in administrative data in Florida (n=181,897) (Balfanz et al., 2015). Shollenberger found that suspended White boys are 23 percentage points less likely to have a high school diploma, 31 percentage points less likely to have attended any college, 29 percentage points less likely to have a BA, and 38 percentage points more likely to have ever been arrested, with similar gaps for Black and Hispanic boys in her analysis of the National Longitudinal Study of Youth 1997 (Shollenberger, 2015).

Job outcomes are mixed. In their 20s and early 30s, adults who were suspended during school do not have lower earnings or employment rates, but they are less likely to have jobs that offer retirement benefits. Jobs with retirement benefits may suggest that the jobs have better futures and greater potential for advancement. Suspended youth may improve earnings and compensate for lower educational attainment by taking lower-quality jobs, suggesting that the earnings parity with matched never-suspended adults could erode with age. This finding is similar to Mulhern's finding that suspended youth

compensate for lower educations by taking riskier jobs, as demonstrated in analysis of suspended youths outcomes from the National Longitudinal Study of Youth 1979 cohort (Mulhern, 2014)

Greater delinquency and lower grades predict greater chances of suspension in both the full sample (which is 17.9% Black) and the Black subsample. Black youth have unique risk factors for school suspension. In the Black subsample, taller males were more likely to be suspended, which suggests that some Black youth are suspended because school personnel feel threatened by them. In the full sample, height does not predict suspension for males or females.

Suspended youth are less likely to earn college and graduate degrees, and even somewhat less likely to earn high school degrees, compared with youth with similar levels of pre-suspension deviance. Among Black youth, the effects are at least as large as in the full sample. Despite having lower educational attainment, suspended youth do not earn less on average. These negative long-term outcomes for suspended youth are particularly disappointing because school suspension is intended to help youth correct their behavior before adulthood.

Strengths and Limitations

This study uses matched sampling, which identifies a comparable group of non-suspended youth according to dozens of factors. This method minimizes potential confounding on the matched pre-suspension factors and factors correlated with them. Both matching and regression yield associative rather than causal inference, but matching yields more valid results for 3 reasons. First, regression models rely on dubious parametric assumptions about linear or log-linear relationships between variables. Regression models cannot adjust, even on average, for large differences between affected groups. Suspended youth differ dramatically from non-suspended youth in factors such as grades, delinquency, and school attachment, so matched sampling is particularly appropriate for studying school suspension. Second, in contrast with traditional regression methods, this matched sampling model computed outcome differences only after verification that the matched suspended group is similar to the non-suspended group. This separation ensures that the model is selected independently of the study's results; with regression, it is impossible to verify model correctness without seeing the results. Third, matching allows adjustment for more variables than does regression: in this case, matching balanced on 60 factors, including some composite variables based on many survey items.

Matching adjusts for observed characteristics and unobserved characteristics to the extent that they are associated with the observed characteristics. A factor not associated with the matched variables could partially explain some of the observed effects. Residual confounding may remain after matching because factors not included in the matching model — both observed and unobserved — could partially explain outcome differences between suspended and non-suspended youth. Large effect sizes are unlikely to be explained completely by unobserved characteristics, however. It is unlikely that a factor that is not associated with the 60 matched variables completely confounds the

relationship between suspension and the outcomes (Rosenbaum, 2002). For example, some of the differences between suspended and non-suspended youth may be attributable to an unobserved characteristic, such as darker skin color. Youth with darker skin of any race or ethnicity may be more likely both to be suspended and to have lower educational attainment and higher arrest likelihoods in adult life due to colorism (Landor et al., 2013; Sweet et al., 2007). Colorism may have a pervasive impact on youth, so skin color is unlikely to be unrelated to any of the matched variables. Similar arguments may apply to other unobserved potential confounders.

Any study with multiple outcomes risks false significance due to multiple comparisons. We restricted the study to outcomes predicted by the suspension literature and consistent with the theory of secondary deviance. This study reported all investigated outcomes related to educational, criminal justice involvement, and job conditions.

Survey weights are designed for use with the full sample and can yield biased estimates when used with subsamples, so survey weights should generally not be used with matched sampling analysis. Also, Add Health has survey weights for only three-quarters of all participants, so the analysis did not use the Add Health survey weights to preserve sample size and power. The Add Health survey over-sampled Black youth with college-educated parents. College-graduate parents have more authoritative parenting styles and more resources to help youth (Dornbusch et al., 1987), so their children may have better outcomes after suspension. This study found worse results for Blacks than for the entire population, but suspension may have even worse results in the general Black population, and this study may underestimate the effects of suspension.

Suspension and expulsion are self-reported, so students may misclassify long suspensions as expulsions, resulting in under-estimates of the effects of suspensions. Most suspensions are for periods shorter than 2 weeks to avoid the need for due process hearings (Arum, 2003), but students may be suspended from school for months or a year. Students may report these suspensions as expulsions rather than suspensions; although they are administratively wrong, a month or year suspension resembles expulsion.

Conclusions

Twelve years later, adults who were suspended as youth have lower educational attainment and greater criminal justice involvement than a matched comparison group with similar pre-suspension characteristics, both in the general population and within the Black subsample. Suspension does not predict reduced deviance. Instead, suspension predicts worse adult outcomes, consistent with the hypothesis of secondary deviance. Strict school disciplinary policies intended to be temporary deterrents may turn temporary deviance into permanent consequences. These results are especially concerning because they do not subside over time, remaining strong in adulthood. Even though administrators may see little negative impact in the short-term, their sanctions may be starting social processes that have long-term implications for these individuals and for society. Schools that use Positive Behavioral Interventions and Supports may avoid this negative cycle. Further research can identify programs to help previously suspended youth to become positively engaged with their schools.

References

ABA (2001). Zero tolerance policy report. American Bar Association Juvenile Justice Committee.

Arum, R. (2003). Judging school discipline. Harvard University Press, Cambridge.

Austin, P. C. and Laupacis, A. (2011). A tutorial on methods to estimating clinically and policy-meaningful measures of treatment effects in prospective observational studies: A review. International Journal of Biostatistics, 7(1):Article 6.

Balfanz, R., Byrnes, V., and Fox, J. (2015). Closing the School Discipline Gap: Equitable Remedies for Excessive Exclusion, chapter Sent home and put off track: The antecedents, disproportionalities, and consequences of being suspended in 9th grade, pages 17–30. Teachers College Press.

Barnhart, M. K., Franklin, N. J., and Alleman, J. R. (2008). Lessons learned and strategies used in reducing the frequency of out-of-school suspensions. Journal of Special Education Leadership, 21(2):75–83.

Beaver, K. M. (2011). The Effects of Genetics, the Environment, and Low Self-Control on Perceived Maternal and Paternal Socialization: Results from a Longitudinal Sample of Twins. Journal of Quantitative Criminology, 27(1):85–105.

Becker, H. (1963). Outsiders: Studies of the Sociology of Deviance. Free Press of Glencoe, Illinois, Glencoe, Illinois.

Berk, R. (2010). What You Can and Can't Properly Do with Regression. Journal of Quantitative Criminology, 26(4):481–487.

Bitner, R. K. (2015). Exiled from Education: Plyler V. Doe's Impact on the Constitutionality of Long-Term Suspensions and Expulsions. Virginia Law Review, 101(3):763–805.

Bowen, W. G., Chingos, M. M., and McPherson, M. S. (2011). Crossing the Finish Line: Completing College at America's Public Universities. Princeton University Press, Princeton, N.J..

Chantala, K. and Tabor, J. (2010). Strategies to perform a design-based analysis using the add health data. Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Cobb-Clark, D. A., Kassenboehmer, S. C., Le, T., McVicar, D., and Zhang, R. (2015). Is there an educational penalty for being suspended from school? Education Economics, 23(4):376–395.

Cook, P. J., Gottfredson, D. C., and Na, C. (2010). School Crime Control and

Prevention. Crime and Justice, 39(1):313–440.

DeLorenzo, J. P. and Rider, R. (2015). Field advisory: Suspension and expulsion of preschool children. Office of Special Education and Office of Student Support Services, Albany, NY.

Dornbusch, S. M., Ritter, P. L., Leiderman, P. H., Roberts, D. F., and Fraleigh, M. J. (1987). The Relation of Parenting Style to Adolescent School Performance. Child Development, 58(5):1244–57.

Evans-Whipp, T. J., Plenty, S. M., Catalano, R. F., Herrenkohl, T. I., and Toumbourou, J. W. (2015). Longitudinal effects of school drug policies on student marijuana use in Washington State and Victoria, Australia. American Journal Of Public Health, 105(5):994–1000.

Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., III, M. P. M., and Booth, E. A. (2011). Breaking schools' rules: A statewide study on how school discipline relates to students' success and juvenile justice involvement. Center for State Governments Justice Policy Center.

Garson, J. (2010). Overview of fourteen Southern states' school suspension laws. Center for Child and Family Policy, Duke University.

Gibson, P. A. and Haight, W. (2013). Caregivers' Moral Narratives of Their African American Children's Out-of-School Suspensions: Implications for Effective Family-School Collaborations. Social Work, 58(3):263–272.

Gilliam, W. S. and Shahar, G. (2006). Preschool and child care expulsion and suspension: Rates and predictors in one state. Infants and Young Children, 19(3):228–245.

Gottfredson, M. R. and Hirschi, T. (1990). A General Theory of Crime. Stanford University Press, Palo Alto, CA.

Hemphill, S., Toumbourou, J., Herrenkohl, T., McMorris, B., and Catalano, R. (2006a). The effect of school suspensions and arrests on subsequent adolescent antisocial behavior in Australia and the United States. J Adolesc Health, 39(5):736–44.

Hemphill, S. A., Heerde, J. A., Herrenkohl, T. I., Toumbourou, J. W., and Catalano, R. F. (2012). The Impact of School Suspension on Student Tobacco Use: A Longitudinal Study in Victoria, Australia, and Washington State, United States. Health Education & Behavior, 39(1):45–56.

Hemphill, S. A., Kotevski, A., Herrenkohl, T. I., Smith, R., Toumbourou, J. W., and Catalano, R. F. (2013). Does school suspension affect subsequent youth non-violent antisocial behaviour? A longitudinal study of students in Victoria, Australia and Washington State, United States. Australian Journal of Psychology, 65(4):236–249.

Hemphill, S. A., Toumbourou, J. W., Herrenkohl, T. I., McMorris, B. J., and Catalano, R. F. (2006b). The effect of school suspensions and arrests on subsequent adolescent antisocial behavior in Australia and the United States. The Journal of adolescent health, 39(5):736–744.

US Department of Health and Human Services and U.S. Department of Education (2014). Policy statement on expulsion and suspension policies in early childhood settings. US Department of Health and Human Services and U.S. Department of Education, Washington DC.

Ho, D., Imai, K., King, G., and Stuart, E. (2008). Matchit: Nonparametric preprocessing for parametric causal inference, software version 2.4-21. Harvard Institute for Quantitative Social Sciences. Cambridge, MA.

Kinsler, J. (2011). Understanding the black–white school discipline gap. Economics of Education Review, 30(6):1370–1383.

Kinsler, J. (2013). School Discipline: A Source or Salve for the Racial Achievement Gap? International Economic Review, 54(1):355–383.

Lamont, J. H., Devore, C. D., Allison, M., Ancona, R., Barnett, S. E., Gunther, R., Holmes, B., Lamont, J. H., Minier, M., Okamoto, J. K., Wheeler, L. S. M., and Young, T. (2013). Out-of-School Suspension and Expulsion. Pediatrics, 131(3):e1000–e1007.

Landor, A. M., Simons, L. G., Simons, R. L., Brody, G. H., Bryant, C. M., Gibbons, F. X., Granberg, E. M., and Melby, J. N. (2013). Exploring the impact of skin tone on family dynamics and race-related outcomes. Journal of family psychology: JFP: journal of the Division of Family Psychology of the American Psychological Association (Division 43), 27(5):817– 826.

Lemert, E. (1967). Human Deviance, Social Problems and Social Control. Prentice-Hall, Englewood Cliffs, NJ.

Liberman, A. M., Kirk, D. S., and Kim, K. (2014). Labeling Effects of First Juvenile Arrests: Secondary Deviance and Secondary Sanctioning. Criminology, 52(3):345–370.

Losen, D., Hodson, C., II, M. A. K., Morrison, K., and Belway, S. (2015). Are we closing the school discipline gap? Center for Civil Rights Remedies, UCLA Civil Rights Project.

McNutt, L., Wu, C., Xue, X., and Hafner, J. P. (2003). Estimating the relative risk in cohort studies and clinical trials of common outcomes. American Journal of Epidemiology, 157(10):940 –943.

Michail, S. (2012). "...becouse Suspension Dosent Teatch You Anything" [sic]: What Students with Challenging Behaviours Say about School Suspension. Report 1324-9320, Australian Association for Research in Education.

Monahan, K., VanDerhei, S., Bechtold, J., and Cauffman, E. (2014). From the School Yard to the Squad Car: School Discipline, Truancy, and Arrest. Journal of Youth & Adolescence, 43(7):1110–1122. 16

Morgan, S. L. and Winship, C. (2015). Counterfactuals and causal inference: Methods and principles for social research. Cambridge University Press.

Mulhern, C. (2014). High school suspension's impact on dangerous jobs and mid-life earnings. Unpublished manuscript.

Nicholson-Crotty, S., Birchmeier, Z., and Valentine, D. (2009). Exploring the Impact of School Discipline on Racial Disproportion in the Juvenile Justice System. Social Science Quarterly, 90(4):1003–1018.

National Longitudinal Study of Adolescent Health (2015). FAQ: Questions about field work. Carolina Population Center, University of North Carolina at Chapel Hill.

American Association of Pediatrics Committee on School Health, A. A. (2003). Policy statement: Out-of-school suspension and expulsion. Pediatrics, 112:1206--1209.

Okonofua, J. A. and Eberhardt, J. L. (2015). Two Strikes Race and the Disciplining of Young Students. Psychological Science, 26(5):617–624.

Ou, S., Mersky, J. P., Reynolds, A. J., and Kohler, K. M. (2007). Alterable Predictors of Educational Attainment, Income, and Crime: Findings from an Inner-City Cohort. Social Service Review, 81(1):85–128.

Paternoster, R. and Iovanni, L. (1989). The labeling perspective and delinquency: An elaboration of the theory and an assessment of the evidence. Justice Quarterly, 6(3):359–394.

Perrone, D., Sullivan, C. J., Pratt, T. C., and Margaryan, S. (2004). Parental efficacy, self-control, and delinquency: A test of a general theory of crime on a nationally representative sample of youth. International Journal of Offender Therapy and Comparative Criminology, 48(3):298–312.

Reynolds, C., Skiba, R., Graham, S., Sheras, P., Conoley, J., and Garcia-Vazquez, E. (2008). Are zero tolerance policies effective in the schools?: An evidentiary review and recommendations. The American Psychologist, 63(9):852–862.

Rosenbaum, P. R. (2002). Observational Studies. Springer-Verlag, New York, 2nd edition.

Rubin, D. (1997). Estimating causal effects from large data sets using propensity scores. Annals of Internal Medicine, 127(8 Pt 2):757–763.

Ruck, M. D. and Wortley, S. (2002). Racial and Ethnic Minority High School Students' Perceptions of School Disciplinary Practices: A Look at Some Canadian

Findings. Journal of Youth & Adolescence, 31(3):185–195.

Shollenberger, T. L. (2015). Closing the School Discipline Gap: Equitable Remedies for Excessive Exclusion, chapter Racial Disparities in school suspension and subsequent outcomes: Evidence from the National Longitudinal Study of Youth, pages 31–43. Teachers College Press.

Skiba, R. J., Peterson, R., and Williams, T. (1997). Office referrals and suspension: Disciplinary intervention in middle schools. Education and Treatment of Children, 20(3):295–315.

Suh, S., Malchow, A., and Suh, J. (2014). Why did the black-white dropout gap widen in the 2000s? Educational Research Quarterly, 37(4):19–40.

Sweet, E., McDade, T. W., Kiefe, C. I., and Liu, K. (2007). Relationships between skin color, income, and blood pressure among African Americans in the CARDIA Study. American Journal of Public Health, 97(12):2253–2259.

Tingley D, Yamamoto T, Hirose K, Keele L, Imai K (2013). mediation: R Package for Causal Mediation Analysis, version 4.4.5.

Tourangeau, R. and Shin, H. (1999). National longitudinal study of adolescent health: grand sample weight. Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Wald, J. and Losen, D. J. (2003). Defining and redirecting a school-to-prison pipeline. New directions for youth development, 99:9–15.

Wiley, S. A., Slocum, L. A., and Esbensen, F.A. (2013). The Unintended Consequences of Being Stopped or Arrested: An Exploration of the Labeling Mechanisms Through Which Police Contact Leads to Subsequent Delinquency. Criminology, 51(4):927–966.

Wolfe, S. E. and Hoffmann, J. P. (2016). On the measurement of low self-control in Add Health and NLSY79. Psychology, Crime, and Law.

Wright, J. P., Morgan, M. A., Coyne, M. A., Beaver, K. M., and Barnes, J. (2014). Prior problem behavior accounts for the racial gap in school suspensions. Journal of Criminal Justice, 42:257–266.

Young, J. K. and Beaujean, A. A. (2011). Measuring Personality in Wave I of the National Longitudinal Study of Adolescent Health. Frontiers in Psychology, 2.

Zimmerman, S. D. (2014). The Returns to College Admission for Academically Marginal Students. Journal of Labor Economics, 32(4):711–754.

Zou, G. (2004). A modified Poisson regression approach to prospective studies with binary data. American Journal of Epidemiology, 159(7):702 –706.

Figure 1: Construction of matched sample.

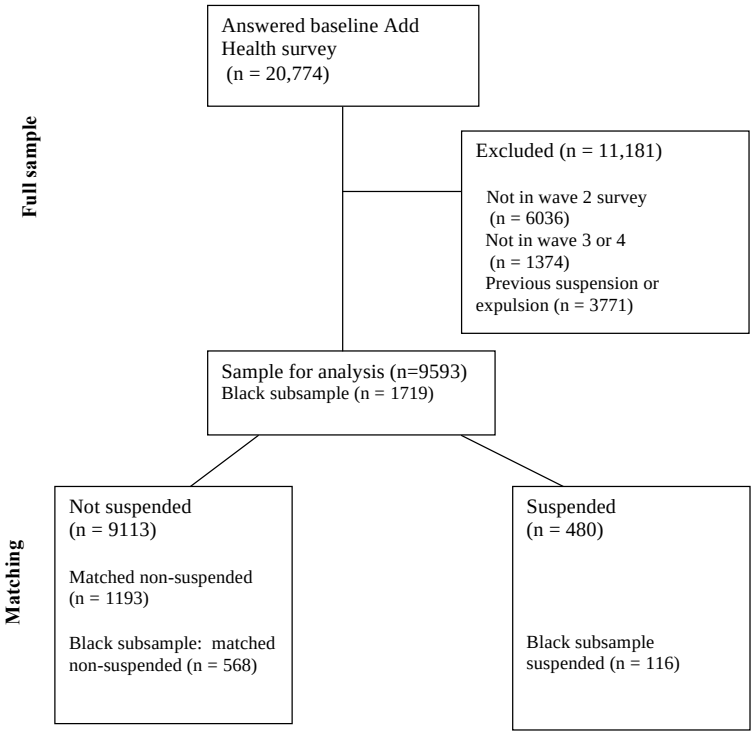

Figure 2: Comparison of standardized differences of baseline factors before and after matching, comparing students who were suspended for the first time in 1995-96 with students who had never been suspended as of 1996.

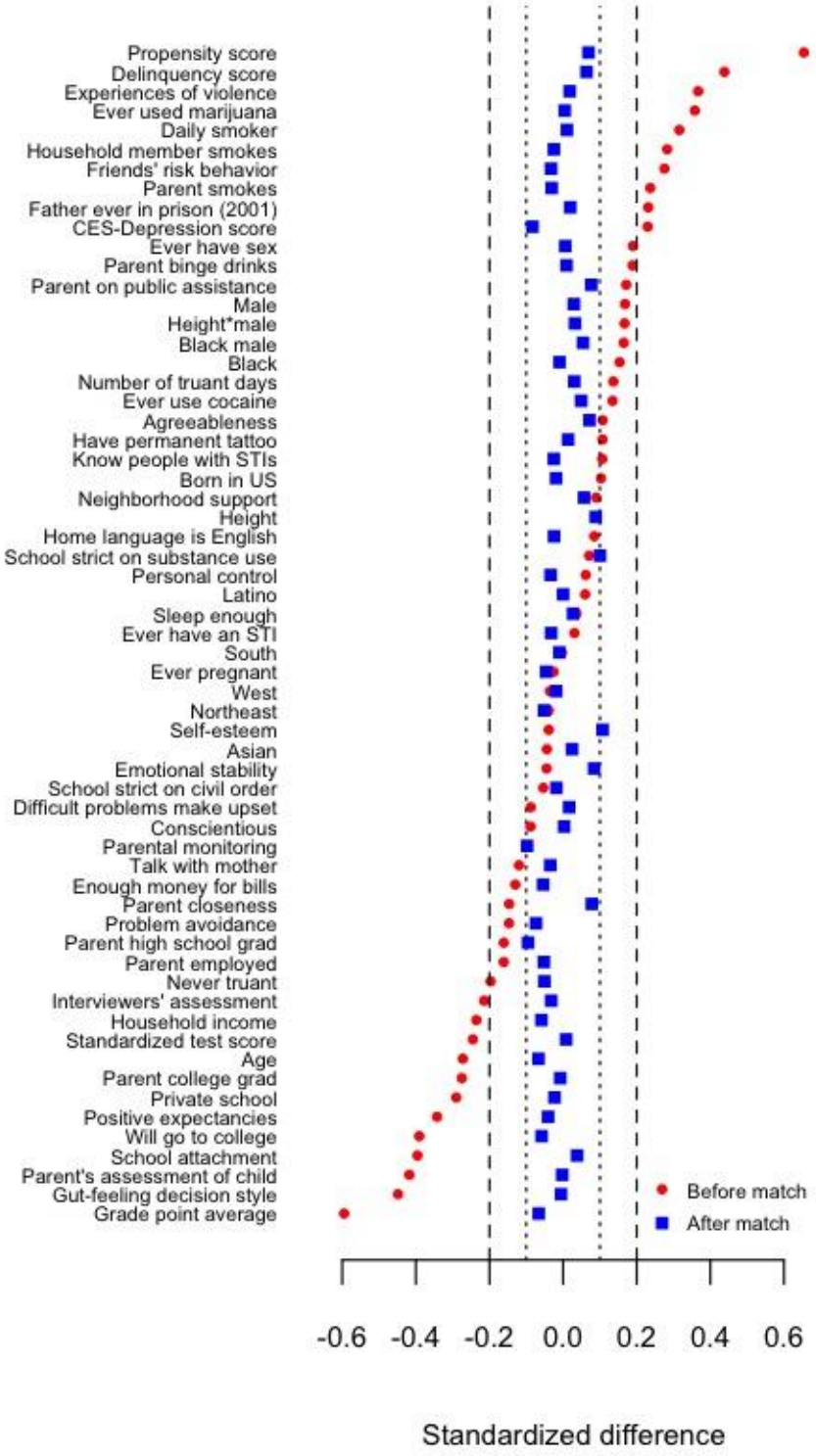

Figure 3: Comparison of standardized differences of baseline factors before and after matching, comparing Black students who were suspended for the first time in 1995-96 with students who had never been expelled or suspended as of 1996.

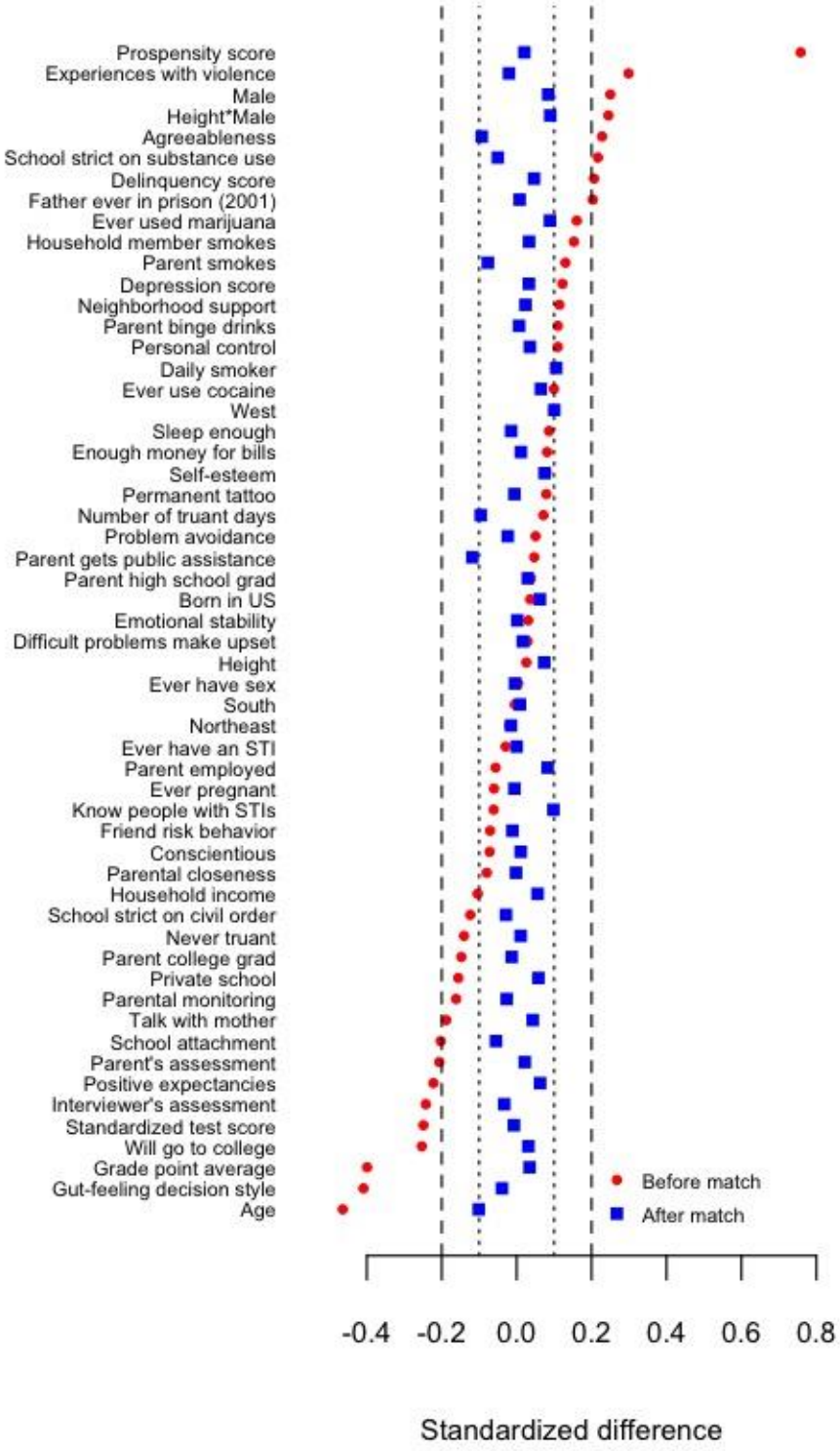

Table 1: Comparison of standardized differences of baseline factors before and after matching, comparing students who were suspended for the first time between 1995 and 1996 with students who had never been suspended as of 1996.

	Pre-matching			Post-matching		
	Susp.	Not susp.	Std. diff	Susp.	Not susp.	Std. diff
N	480	9113		480	1193	
Grade point average	2.6	3.0	-0.59	2.6	2.6	-0.07
Systematic (vs. gut-feeling) decision style	41.3	53.5	-0.45	41.3	41.5	-0.01
Parent's assessment of child	74.3	80.7	-0.42	74.3	74.3	0.00
School attachment	67.2	73.4	-0.40	67.2	66.6	0.04
Will go to college	71.7	83.4	-0.39	71.7	73.5	-0.06
Positive expectancies	81.2	86.2	-0.34	81.2	81.8	-0.04
Private school	3.3	8.5	-0.29	3.3	3.8	-0.02
Parent college grad	15.0	24.8	-0.28	15.0	15.3	-0.01
Age (years)	14.8	15.2	-0.27	14.8	14.9	-0.07
Standardized test score	75.5	78.0	-0.25	75.5	75.5	0.01
Household income ($1k)	31.7	37.8	-0.24	31.7	33.2	-0.06
Never truant	71.5	80.4	-0.20	71.5	73.8	-0.05
Parent employed	59.0	66.9	-0.16	59.0	61.5	-0.05
Parent high school grad	67.9	75.4	-0.16	67.9	72.4	-0.10
Problem avoidance	42.8	46.5	-0.15	42.8	44.6	-0.07
Parent closeness	77.7	79.7	-0.15	77.7	76.6	0.08
Enough money for bills	66.2	72.4	-0.13	66.2	68.8	-0.05
Talk with mother	45.9	49.6	-0.12	45.9	47.0	-0.03
Parental monitoring	68.7	71.1	-0.10	68.7	71.1	-0.10
Conscientiousness	68.8	70.2	-0.09	68.8	68.7	0.00
Difficult problems make upset	34.9	37.0	-0.09	34.9	34.5	0.02
School strict on civil order	22.9	25.2	-0.05	22.9	23.7	-0.02
Emotional stability	77.6	78.3	-0.04	77.6	76.4	0.08
Asian	7.1	8.2	-0.04	7.1	6.5	0.02
Self-esteem	76.1	76.6	-0.04	76.1	74.7	0.11
Northeast	12.9	14.2	-0.04	12.9	14.7	-0.05
West	21.2	22.7	-0.04	21.2	22.0	-0.02
Ever pregnant	1.9	2.2	-0.03	1.9	2.5	-0.05
South	35.6	35.8	0.00	35.6	36.1	-0.01
Ever have sexually transmitted infection	1.7	1.3	0.03	1.7	2.1	-0.03
Sleep enough	74.6	73.0	0.04	74.6	73.4	0.03
Latino	17.9	15.6	0.06	17.9	17.9	0.00
Personal control	28.4	27.0	0.06	28.4	29.1	-0.03
School strict on substance use	24.4	21.3	0.07	24.4	20.1	0.10
Home language is English	91.2	88.9	0.08	91.2	91.9	-0.02

Height (inches)	66.0	65.6	0.09	66.0	65.6	0.09
Neighborhood support	79.6	77.3	0.09	79.6	78.1	0.06
Born in US	74.4	69.9	0.10	74.4	75.2	-0.02
Know people with STIs	26.5	21.8	0.11	26.5	27.6	-0.03
Have permanent tattoo	4.4	2.2	0.11	4.4	4.1	0.01
Agreeableness	38.3	36.2	0.11	38.3	36.9	0.07
Ever use cocaine	5.4	2.4	0.13	5.4	4.3	0.05
Num. truant days	1.8	1.0	0.14	1.8	1.6	0.03
Black	24.2	17.6	0.15	24.2	24.6	-0.01
Black male	10.6	5.5	0.16	10.6	9.0	0.05
Height*male	30.4	24.8	0.17	30.4	29.3	0.03
Male	44.8	36.4	0.17	44.8	43.3	0.03
Parent on public assistance	12.1	6.5	0.17	12.1	9.6	0.08
Parent binge drinks	16.9	9.8	0.19	16.9	16.5	0.01
Ever have sex	34.2	25.2	0.19	34.2	33.9	0.01
Depression score	12.7	11.1	0.23	12.7	13.2	-0.08
Father ever in prison (2001)	19.2	10.2	0.23	19.2	18.4	0.02
Parent smokes	32.7	21.6	0.24	32.7	34.2	-0.03
Friends' risk behavior	46.9	36.0	0.28	46.9	48.2	-0.02
Household member smokes	49.4	35.2	0.28	49.4	50.6	-0.02
Daily smoker	33.1	18.2	0.32	33.1	32.6	0.01
Ever use marijuana	36.2	19.0	0.36	36.2	36.0	0.00
Experiences of violence	7.7	3.9	0.37	7.7	7.5	0.06
Delinquency score	2.3	1.5	0.44	2.3	2.2	0.07
Propensity score	10.6	4.7	0.61	10.5	9.9	0.07

Table 2: Comparison of standardized differences of baseline factors before and after matching, comparing Black students who were suspended for the first time between 1995 and 1996 with students who had never been suspended as of 1996.

	Pre-matching			Post-matching		
	Susp.	Not susp.	Std. diff	Susp.	Not susp.	Std. diff
N	116	1603		116	568	
Age	14.5	15.1	-0.46	14.5	14.6	-0.10
Systematic (vs. gut-feeling) decision style	42.2	53.5	-0.41	42.2	43.3	-0.04
Grade point average	2.6	2.8	-0.40	2.6	2.6	0.03
Will go to college	77.9	85.5	-0.25	77.9	77.0	0.03
Standardized test score	71.8	74.3	-0.25	71.8	71.9	-0.01
Interviewer's assessment	42.0	51.7	-0.24	42.0	43.3	-0.03
Positive expectancies	82.5	86.0	-0.22	82.5	81.5	0.06
Parent's assessment	78.3	81.2	-0.21	78.3	78.0	0.02
School attachment	69.7	72.9	-0.20	69.7	70.6	-0.06
Talk with mother	45.3	51.0	-0.19	45.3	43.9	0.04
Parental monitoring	65.5	70.0	-0.16	65.5	66.3	-0.03
Private school	4.3	7.5	-0.16	4.3	3.1	0.06
Parent college grad	20.7	26.7	-0.15	20.7	21.2	-0.01
Never truant	79.3	85.0	-0.14	79.3	78.9	0.01
School strict on civil order	28.4	34.1	-0.12	28.4	29.7	-0.03
Household income ($1k)	28.0	30.3	-0.10	28.0	26.8	0.06
Parental closeness	78.4	79.4	-0.08	78.4	78.4	0.00
Conscientiousness	72.1	73.2	-0.07	72.1	71.9	0.01
Friends' risk behaviors	27.6	29.8	-0.07	27.6	27.9	-0.01
Know people with STIs	30.2	33.0	-0.06	30.2	25.6	0.10
Ever pregnant	3.4	4.6	-0.06	3.4	3.6	-0.01
Parent employed	63.8	66.5	-0.06	63.8	59.8	0.08
Ever have sexually transmitted infection	2.6	3.1	-0.03	2.6	3.3	-0.05
Northeast	8.6	9.2	-0.02	8.6	9.1	-0.02
South	56.0	56.3	0.00	56.0	55.6	0.01
Ever have sex	37.9	37.7	0.00	37.9	38.1	0.00
Height (inches)	65.8	65.7	0.03	65.8	65.5	0.07
Difficult problems make upset	35.5	34.8	0.03	35.5	35.1	0.01
Emotional stability	81.8	81.4	0.03	81.8	81.8	0.00
Born in US	75.0	73.4	0.04	75.0	72.3	0.06
Parent high school grad	75.0	73.4	0.04	75.0	73.7	0.03
Parent gets public assistance	12.9	11.4	0.05	12.9	16.9	-0.12
Problem avoidance	42.2	40.9	0.05	42.2	42.8	-0.02
Num. truant days	0.8	0.6	0.07	0.8	1.0	-0.10
Enough money for bills	64.7	60.8	0.08	64.7	64.1	0.01

Sleep enough	75.0	71.2	0.09	75.0	75.6	-0.01
West	18.1	14.3	0.10	18.1	14.2	0.10
Ever use cocaine	3.4	1.6	0.10	3.4	2.3	0.06
Daily smoker	11.2	7.9	0.10	11.2	7.9	0.11
Personal control	28.5	26.1	0.11	28.5	27.7	0.04
Parent binge drinks	12.9	9.2	0.11	12.9	12.7	0.01
Neighborhood support	81.7	78.9	0.11	81.7	81.1	0.02
Depression score	12.4	11.6	0.12	12.4	12.2	0.03
Parent smokes	25.0	19.3	0.13	25.0	28.3	-0.08
Household member smokes	38.8	31.3	0.15	38.8	37.2	0.03
Ever use marijuana	23.3	16.5	0.16	23.3	19.5	0.09
Father ever in prison (2001)	21.6	13.2	0.20	21.6	21.2	0.01
Delinquency score	1.8	1.4	0.21	1.8	1.7	0.05
School strict on substance use	39.7	29.0	0.22	39.7	42.1	-0.05
Agreeableness	40.2	35.4	0.23	40.2	42.2	-0.09
Height*male	29.7	21.4	0.24	29.7	26.7	0.09
Male	44.0	31.5	0.25	44.0	39.8	0.08
Experiences with violence	8.1	5.1	0.30	8.1	8.3	-0.02
Propensity score	14.6	6.2	0.76	14.6	14.4	0.02

Table 3: Relative risks for primary outcomes in 2001 and 2008 associated with a first suspension between 1995 and 1996. Each entry is derived from a regression coefficient within the respective matched samples, of all students and Black students, controlling for demographics, socioeconomic status, educational achievement, parents' risk behaviors, substance use, personality, delinquency and adverse experiences, appearance, relationship with parent, physical and mental health, and neighborhood context. Dichotomous outcomes are from a Poisson working model and continuous outcomes are from a linear regression. P-values are listed if $P \leq 0.1$.

	All students (n=1673)			**Black students (n=684)**		
	%	**IRR (95% CI)**	**P**	**%**	**IRR (95% CI)**	**P**
Outcome in 2001 (dichotomous)						
Ever expelled	5.1	2.69 (1.70, 4.26)	<0.001	6.4	2.77 (1.44, 5.32)	0.002
High school diploma	68.2	0.92 (0.86, 0.99)	0.03	71.7	0.93 (0.81, 1.07)	
BA	3.6	0.78 (0.40, 1.55)		2.7	0.06 (0.005, 0.80)	0.03
Ever arrested	10.3	1.40 (1.05, 1.89)	0.02	7.1	1.38 (0.62, 3.08)	
Arrested as a minor	4.1	1.94 (1.15, 3.19)	0.01	2.9	2.90 (1.04, 8.07)	0.04
Convicted as a minor	1.6	3.75 (1.60, 8.79)	0.002	0.8	2.81 (0.60, 13.3)	
Arrested as adult	7.9	1.31 (0.79, 1.61)		5.1	0.47 (0.16, 1.38)	
Convicted as adult	3.9	1.22 (0.73, 2.05)		2.5	1.45 (0.32, 6.64)	
Outcome in 2008 (dichotomous)						
Arrested once	15.5	1.30 (1.01, 1.66)	0.04	13.3	1.58 (0.97, 2.59)	0.07
Arrested 2+ times	13.7	1.51 (1.17, 1.95)	0.002	11.8	1.15 (0.66, 1.98)	
Ever in prison	16.3	1.23 (0.97, 1.56)	0.09	15.8	1.75 (1.11, 2.75)	0.02
Ever on probation	14.0	1.49 (1.15, 1.92)	0.003	12.9	2.06 (1.25, 3.41)	0.005
High school diploma	74.8	0.94 (0.88, 1.00)	0.05	76.9	0.91 (0.83, 1.00	0.05
AA	11.2	1.04 (0.75, 1.44)		8.9	0.87 (0.42, 1.82)	
BA	16.3	0.76 (0.59, 0.97)	0.03	19.7	0.93 (0.63, 1.39)	
Graduate degree	3.5	0.88 (0.48, 1.61)		4.5	1.04 (0.46, 2.37)	
Currently employed	59.8	0.95 (0.87, 1.04)		59.6	0.94 (0.79, 1.13)	
Job has health benefits	62.9	0.98 (0.88, 1.04)		63.7	0.91 (0.77, 1.07)	
Job has retirement benefits	54.1	0.90 (0.82, 1.00)	0.05	58.0	0.80 (0.66, 0.98)	0.03
Job has vacation benefits	62.6	0.95 (0.87, 1.03)		68.3	0.85 (0.73, 0.99)	0.04
Work is physically strenuous	11.4	1.18 (0.89, 1.56)		11.3	1.10 (0.60, 2.01)	
Work is desk job or sedentary	25.6	0.85 (0.70, 1.05)		31.2	0.78 (0.54, 1.11)	
Outcome in 2008 (continuous)	**Mean**	**Coefficient (95 % CI)**	**P**	**Mean**	**Coefficient (95% CI)**	**P**
Personal earnings (log $1000)	2.41	-0.07 (-0.24, 0.11)		2.33	-0.16 (-0.49, 0.17)	
Household income (log $1000)	8.79	-0.21 (-0.62, 0.20)		8.40	-0.30 (-1.07, 0.47)	
Perceived SES (1-10)	4.2	-0.17 (-0.41, 0.06)		4.1	-0.30 (-0.73, 0.14)	

Table 4: Causal mediation analysis

Mediator (2001)	Outcome (2008)	% mediated (95% CI)	p-value
Expulsion	High school graduation	27.4 (12.1, 1.23)	<0.001
Ever arrested	High school graduation	5.4 (-0.4, 31.9)	0.12
Arrested as minor	High school graduation	6.4 (0.7, 22.6)	0.04
Convicted as minor	High school graduation	8.7 (0.9, 24.2)	0.02
Expulsion	Bachelors degree	10.4 (1.9, 33.6)	0.02
High school graduation	Bachelors degree	9.7 (0.8, 34.4)	0.02
Ever arrested	Bachelors degree	1.3 (-2.7, 9.0)	0.42
Convicted as minor	Bachelors degree	-3.5 (-25.4, 6.8)	0.32
Arrested as minor	Bachelors degree	0	0.98