Improving the Estimation of the Risk of School Suspension Using Continuous-Time Survival Analysis: A Case Study in the Public Middle Schools in One Metropolitan Region

Stephen L. Hoffman

Harvard Graduate School of Education

September 3, 2014

Qualifying Paper

stephen_hoffman@mail.harvard.edu

Improving the Estimation of the Risk of School Suspension

Using Continuous-Time Survival Analysis:

A Case Study in the Public Middle Schools in One Metropolitan Region

For 10 years I was a middle-school principal and a high-school assistant principal. On most days of the school year, I assigned formal discipline to the handful of students whose misbehavior was brought to my attention. Many times, I suspended students out-of-school, following district procedures and the approach to discipline advocated by my superiors. Furthermore, since my first day as an administrator, I was made acutely aware of the racial dynamics in schools, particularly regarding the disproportionality in school discipline for Black and Hispanic students that has also been documented extensively in recent scholarship (Beck & Muschkin, 2012; Gregory, Cornell, & Fan, 2011; Kinsler, 2011; Petras, Masyn, Buckley, Ialongo, & Kellam, 2011; Skiba et al., 2011) .

My experiences as a school principal and my subsequent reading of the scholarly literature have motivated me to scrutinize disciplinary policies in schools more formally. Thus, here in my qualifying paper (QP), I have conducted a case study to explore the out-of-school suspension of students in the public schools in one metropolitan region. In particular, I bring an innovative analytic framework—survival analysis—to bear, to improve the estimation of suspension rates and describe the timing of suspension using a finer lens than others have done. Broadly speaking, I extend prior work on school discipline by investigating *whether* middle-school students are suspended from school, and if so, *when, during the school year,* these school suspensions occurred.

I begin by analyzing when, during the course of the 2009-10 academic year, students were at the greatest risk of being first suspended from school (that is, given that they had not been suspended at an earlier point in the academic year). I explain my use of the term risk more formally, both in terms of *who* is at risk, and also *when* students are at relatively higher or lower rates of occurrence of first suspension. Given that I have worked with data in which time to suspension was measured continuously, I fitted a series of Cox-regression models (Cox, 1972) to first investigate whether there were differences in this risk, by student grade level. Then, I document the continued pattern of disproportionality in the risk of school suspensions by race.

For the last 40 years, researchers in fields like criminology (e.g. Dugan, Lafree, & Piquero, 2005) medicine (e.g. Heikinheimo et al., 2013), and sociology (e.g. Zheng & Thomas, 2013) have used the strategy of Cox (1972) regression analysis to study analogous topics. Cox-regression analysis is certainly the most popular technique for modeling rates of event occurrence in continuous-time, because it is both elegant and computationally practical (Cleves, Gould, Gutierrez, & Stata Corporation., 2010). In fact, as of March 7, 2014, Cox's seminal paper detailing this method had been cited 27,942 times on the Web of Science! However, I am unaware of any other researcher who has applied Cox-regression analysis to the investigation of the phenomenon of out-of-school suspensions, as I have here.

I have organized the QP into five major sections, following this introduction. First, in the *Background and Context* section, I review the literature to provide a rationale for my investigation of *whether* students have been suspended, historically, from school. I focus on disparities by race and ethnicity generally, and for middle-school students in particular. In this section, I argue that survival analysis is a plausible strategy for addressing the "when" question, in estimating the risk of suspension during the school year. I end this section by stating my specific research questions. Second, in my *Research Design* section, I describe the research site, the features of the dataset, the sample of students, and the procedures that I use to address my specific research questions. Third, I present my *Findings* from this analysis. Fourth, I discuss the limitations of my analysis in a section on *Threats to Validity.* Fifth, I synthesize and review the meaning and significance of the findings in a *Discussion* section.

Background and Context

Out-of-School Suspension as a Disciplinary Tool

The use of out-of-school suspension of students as a disciplinary tool has been a common practice by public-school administrators for decades (Edelman, Beck, & Smith, 1975; Wu, Pink, Crain, & Moles, 1982; Arum, 2003). However, prior to the 1970s, the use of school exclusion—such as out-of school suspension—was almost entirely informal and discretionary. Typically, school principals were not subject to formal oversight of their use of suspensions, even by school superintendents (Edelman et al., 1975). Then, during the 1960s and 1970s, the issue of school suspensions came to the forefront of public debate, as a result, in part, of the discussion of racial issues spearheaded by the Civil-Rights movement (Arum, 2003). This led to substantial new policy decisions about the use of school suspension, ordered by the courts and sanctioned by the federal government

(Edelman et al., 1975; Ornstein, 1982). In the Goss v. Lopez (1975) decision, the U.S. Supreme Court settled whether students might be suspended out-of-school, due to their behavior.

Goss v. Lopez. The Goss v. Lopez (1975) decision remains a signature legal holding about the use of out-of-school suspensions (Arum & Preiss, 2009). In this ruling, the U.S. Supreme Court described how school staff should enforce standards of conduct for students because some “modicum of discipline and order is essential if the educational function is to be performed” (Goss v. Lopez, p. 580). Noting that disciplinary events in schools were frequent occurrences, requiring “immediate, effective action,” the court explained that the use of out-of-school suspension was considered a necessary tool to maintain order (Goss v. Lopez, p. 580). By reacting to the social and political issues of the times, the Supreme Court outlined a broad framework for the use of school exclusion that still guides most public-school policies about out-of-school suspensions (Arum & Preiss, 2009).

But in addition to sanctioning the use of out-of-school suspension—the rights of school officials— the Supreme Court also outlined rights for students. Thus, the Goss v. Lopez (1975) ruling affirmed that students also possess rudimentary due-process rights regarding school exclusion. Even when a student is suspended out-of-school for violating school rules, school administrators must not act arbitrarily or capriciously when doling out the punishment. As Justice White explained in the ruling, students who are suspended from school “have interests qualifying for protection of the Due-Process Clause” (p. 581). Consequently, school officials, upon suspending a student for 10 days or less following a disciplinary incident, must explain explicitly *why* a student is to be suspended. School staff must specify what evidence the school used to make the decision to suspend. And, students must be afforded an opportunity to tell their side of the story.

Federal Responsibilities/States' Responsibilities. Despite political arguments about the federal role in public education in the United States—e.g. recent debates about *No Child Left Behind* (2002), *Race to the Top* (2011), or the implementation of the *Common Core Standards* (2010)—education remains primarily a function and responsibility of the states. All 50 states in the U.S. maintain a public-school system. Even after the turmoil in the wake of the Brown v. Board of Education (1954) ruling ordering school desegregation, state governments did not seriously question whether states would continue to support public education (Butts, 1955). While states like Indiana—an early adopter of the *Common Core Standards*—are now withdrawing their support for implementing them (e.g. Martin, 2014, reporting for the New York Times) the broader framework of federal and state responsibilities remains.

State-government policies stipulate the rules and regulations regarding school discipline, including policies about the use of out-of-school suspensions. In Goss v. Lopez (1975) the Court ruled specifically that the state of Ohio (as this case dealt specifically with a difficult disturbance at a high school in the Columbus, Ohio, public-school system) is "not constitutionally obligated to establish and maintain a public school system" (p. 574). But since each has chosen to extend this right of a public education, states must follow the Supreme Court's broad guidelines concerning due process when using out-of-school suspension as a disciplinary tool.

Thus, any study of school discipline broadly—and of the use of out-of-school suspensions in particular—must consider the context present in the home state of the schools. For example, in the current study, I investigated the application of student suspension in a non-random sample of schools, in one particular state. The relevant state law regarding the use of out-of-school suspension asserts that school officials may suspend a pupil for not more than five school days, unless a notice-of-expulsion hearing has been sent. Thus, principals, charged with maintaining order and discipline, *may* use out-of-school suspension for students violating school rules. However, questions about w*hether* students may be suspended for violating school rules are left frequently to the judgments of school principals—a situation similar to schools described by Edelman, et al (1975) 40 years ago.

***Whether* students were suspended from school?** As Justice White maintained, out-of-school suspension was "considered not only to be a necessary tool to maintain order, but a valuable educational device" (Goss v. Lopez, p. 580). Consequently, large numbers of students are suspended from the public schools in the U.S. every year, although summaries of the rates at which these suspensions are administered depend on the methodology used to estimate them. For example, the federal government authorizes the mandatory collection and summarization of data on school suspensions through the United States Department Office of Civil Rights (2013). These guidelines define out-of-school suspension as "excluding a student from school for disciplinary reasons for one school day or longer" (p. 35). A recent summary provided by the National Center for Education Statistics (NCES) estimated that 3,325,000 students were suspended out-of-school at least once during 2006

(U.S. Department of Education, 2012). Thus, about 7% of all public elementary- and secondary-school students were suspended in that year. And the NCES summary noted that students were counted only once in the estimates, even if students were suspended multiple times from the same school during the same school year. Furthermore, these NCES estimates of suspension rates differ considerably by state. In North Dakota, at one extreme, approximately 2.2% K-12 public-school students of were suspended out-of school in 2006. At the other extreme, in South Carolina, the percentage of students suspended at least once was 11.9%.

***Why* were students suspended from school?** In addition to the question of *whether* students are subject to out-of-school suspensions, formal school discipline is administered for a large range of kinds of disobedience. *Why* a student was suspended from school—for events that are moderately disruptive, like insubordination or disruption; to more serious infractions, such as fighting or using drugs or alcohol—has been the subject of scholarly interest for decades (Theriot & Dupper, 2010; Raffaele Mendez & Knoff, 2003; Skiba, Peterson, & Williams, 1997; Wu et al., 1982; Edelman et al., 1975). While in the current paper, I focus on questions of *whether* and *when* students are suspended, the reasons *why* a student was suspended from school will also become an important aspect of my future work, particularly when questions about disproportionality in school discipline by race come to the fore.

In the wake of the Civil-Rights movement, the *Office of Civil Rights* first began collecting data about the nature and use of school suspension in the United States (Edelman et al., 1975). Thus, immediately following the Goss v. Lopez (1975) ruling, cited above, advocates like those at the *Children's Defense Fund* attempted to make sense of the myriad reasons that students were being subject to out-of-school suspensions. For instance, in their report, Edelman, Beck and Smith (1975) found that most suspensions "were for nondangerous offenses," such as "insulting" the teacher or tardiness to class—and they recommended that schools reduce the use of out-of-school suspension for these behaviors (Edelman, Beck & Smith, 1975, pp. 37-38). They concluded that patterns of suspension were largely a consequence of differences among school administrators and suspension policies, rather than different behavioral patterns among students. This was confirmed empirically by Wu, et al (1982) in a subsequent national study of 641 public secondary schools conducted in 1976. Wu and his colleagues concluded that student misbehavior was only one factor resulting in an out-of-school suspension. Principals' decisions to suspend students were mediated by teachers' perceptions and beliefs, schools' administrative structures, and institutional biases (Wu et al., 1982).

Several recent studies have also described the types of infractions that have led to out-of-school suspension for students. In a study of disciplinary interventions in middle schools, for instance, Skiba, Peterson, and Williams (1997) noted that the reasons why students were suspended out-of-school differed considerably among principals. While fighting seemed to be a straightforward case for the use of out-of-school suspension, rates of suspension for insubordination or disruption differed considerably by school. In another study, Raffaele Mendez and Knoff (2003) also studied out-of-school suspensions in a large school district in Florida and found that about four-times as many suspensions were handed out for "disobedience," than for "violence against persons." Very few suspensions were administered for the possession and use of alcohol, drugs, or weapons—particularly in middle schools. Similarly, Theriot and Dupper (2010) studied discipline problems during the transition from elementary to middle school in a medium-sized school district. They found that less than half of the out-of-school suspensions were due to infractions like fighting, threatening behavior, or theft. More commonly, out-of-school suspensions were imposed for reasons like "conduct prejudicial to good order" (p. 213). Thus, while issues of drugs, weapons, and gang activity presented serious issues that school principals must address, disruptive and insubordinate conduct were behaviors that were addressed as often as daily.

At what age are students suspended from school? Adolescence can be a particularly trying time for children, as well as for those who teach or otherwise staff a middle school. Few studies have investigated how rates of out-of-school suspension differ by grade level. Studies have indicated that suspension rates for middle-school students exceed those of elementary-school students substantially; and, in some studies, rates of suspension for middle-school students have even exceeded rates for high-school students (Theriot & Dupper, 2010; Losen & Skiba, 2010; Raffaele Mendez & Knoff, 2003). Losen and Skiba (2010) explained that, as of 2010, no nationally representative reports based on federally collected discipline data have been disaggregated by grade level. However, Losen and Skiba (2010) reported—in their study of suspensions in 18 large urban school districts—that the average suspension rate in middle schools in these districts was 11.2%, in 2006. Similarly, in their study of students in a public-school district in Florida, Raffaele

Mendez and Knoff (2003) reported that the percentage of students who experienced at least one suspension during the 1996-97 year was less than 4% for elementary-school students, 24% for middle-school students, and 18% for high-school students. And, as Theriot and Dupper (2010) noted in their study of the transition from elementary school to middle school, rates of suspension for students in 6th grade (at the beginning of middle school) were substantially higher than the rates of suspension for the same students during the previous school year (when the students were still in elementary school).

Refining Annual Suspension Rates: *When* Are Students Suspended During the School Year?

In his study of school principals, Lortie (2009) noted that the day-to-day work of principals is marked by both urgency and randomness. "Outbursts of student misbehavior can be sudden and unpredictable" (Lortie, 2009, p. 123). Thus, one might hypothesize that suspensions must be distributed essentially at random over the school year, and be administered uniformly. However, typically, the question of *when* suspensions occurred—that is, at what point during the school year—has been left unanswered in investigations of school suspensions. Are 6th-grade students at greatest risk of experiencing suspensions at the start of the school year, when, perhaps, 6th-grade students must acclimate immediately to the rigors of middle school? Or, are the first suspensions of middle-school students more evenly spaced out during the course of the year? Do 8th-grade students tend to become unruly towards the end of the academic year, experiencing substantial increases in the risk of suspension? In their concluding remarks, for instance, Theriot and

Dupper (2010) speculated that the rate of suspensions of 6th-grade students would decrease after students had time to adjust to middle school, although this hypothesis remained unaddressed in their research.

While the estimation of simple annual rates of suspension, as carried out in many studies, provided some information about *whether,* and *why,* students are suspended from school (Raffaele Mendez & Knoff, 2003; Theriot & Dupper, 2010), information about *when* during the school year suspensions have occurred is notably absent from most research. One innovative approach to investigating *when* students are suspended from school is by the use of the methods of survival analysis. Although indirectly, survival analysis focuses on, and summarizes, the *time* that elapses from an "origin" (such as the beginning of a school year) until the occurrence of a "failure" (such as an event like a school suspension). Thus, using survival analysis, I can examine both *whether*, but also, *when*—during the school year—a student was at greatest risk of being suspended from school (Singer & Willett, 2003). Modeling the day-by-day risk of school suspension allows me to determine when the occurrence of school suspension is at its greatest (conditional) risk of occurring (Singer & Willett, 2003).

As a result, I am able to answer interesting questions about school suspension. For example, I can ask: Do suspensions occur earlier in the school year, tending to set boundaries on student behaviors, as Theriot and Dupper (2010) hypothesized for 6th-grade students? Or, are suspensions more akin to random incidences that occur uniformly throughout the school year, with students socialized into complying with school norms more gradually, as Lortie (2009) explained?

Losen and Skiba (2010) described rule-breaking as normal behavior—particularly for middle-school students, who are accustomed to challenging authority both at home and at school, as part of their normal adolescent social development. So, while some behavioral norms may be established early in the school year, perhaps other developmental factors continue to play a substantial role throughout middle school.

Does survival analysis describe all incidents of event occurrence? Many studies, such as the NCES report (U.S. Department of Education, 2012), cited earlier, state that students were counted only once in the estimates, even if students were suspended multiple times from the same school during the same school year. However, in one recent exception, Sullivan et al (2013) reported that the proportion of students who were suspended *exactly* once during the school year was slightly larger than the proportion of students who were suspended multiple times.

One strategy to overcome this blurring of single and multiple suspensions is to focus on the *first* suspension of the school year. For example, Petras, Masyn, Buckley, Ialongo, and Kellam (2011), in a study of who is most at risk of school removal, focused on the *first* school suspension (in a student's school career), because they hypothesized that the timing of the first suspension was predictive of later school problems. Thus, they used discrete-time survival analysis to investigate the risk of suspension from school *for the first time*. However, Petras and colleagues focused on *whether* and *in which grade* a student was first suspended, during an entire student's elementary-school career (estimating a *yearly* probability of out-of-school suspension from grades 1 through 7). In contrast, I focus on the *daily* risk of students being suspended for the first time, out-of-school, during the course of one school year, for three cohorts of students in their 6th, 7th, and 8th grades, respectively.

The problem of censoring. One of the critical problems facing scholars who wish to investigate the risk of out-of-school suspension is the ubiquitous problem of *censoring*. In this study, in which data on school suspensions were collected and analyzed for three cohorts (6th, 7th, and 8th grade) of students over one school year, the students fall into two categories. The first category includes those who were suspended for the first time while they were observed during the school year. They provide direct insight into the question of *"whether"*, and the answer for them is "yes." The rest of the students in the three cohorts fall into the category of *censored observations* (Singer & Willett, 2003). For these latter students, I do not observe a first suspension during the school year. The challenge that I face in estimating the risk of suspension is to incorporate both uncensored and censored students legitimately into the same analysis in order to obtain unbiased estimates of suspension rates, by group.

The *risk* of first suspension. Assuming that all middle-school students are at eligible to be suspended from school, using survival analysis allows me to investigate when*,* during the school year*,* the risk of school suspension is the highest. In such research, the fundamental quantity of interest is called the *hazard*—the risk at any time during the school year that a student might be suspended, given that the student has not yet experienced a first suspension. Furthermore, in this study, where I have chosen to model this phenomenon in continuous time (because the student records are available to the nearest day), I estimate and report the *hazard rate*—the conditional probability of first suspension, *per day* in school (Singer & Willett, 2003).

Differences in Suspension From School By Race or Ethnicity

Historically, public schools in the United States had served, primarily, the needs of White students, both male and female. This was particularly the case for students enrolled in public education beyond primary school. Indeed, prior to the Civil War, more than 99% of students aged 12 years old in the nation's public schools were White students.[1] As Goldin and Katz (2008) explained, free public schooling has been provided for White youth of both genders and all social classes since the late 1800s. Average levels of educational attainment rose throughout the first two-thirds of the 20th century, with the proportions of males and females attaining a high-school diploma in school rising throughout that time at roughly comparable rates.

While public schools served the vast majority of Whites in the U.S., educational opportunities for non-Whites were considerably different, traditionally. For example, Anderson (1988) noted that for at least 70 years following the Civil War, a substantial portion of the White population resisted efforts to provide even basic education to Black youth, or endorsed education primarily for preparing Blacks to work in a perpetually lower-class status. Similarly, Lomawaima (1999) described a long, sordid history of providing educational "opportunities" for Native-American children—including a 450-year history of boarding schools and other

[1] Author's estimation using U.S. Census data, archived by the Minnesota Population Center at the University of Minnesota (Ruggles et al., 2010).

American institutions teaching a "special pedagogy" formulated by Whites. Thus, for both the historically enslaved Black population and the indigenous Native-American population, public educational opportunities lagged considerably behind those for Whites.

The presence of the two other major racial/ethnic minorities, Asian and Hispanic, was quite small until recent demographic shifts. Tamura (2001) noted that Asian Americans have lived in the U.S. for over 150 years. But the proportion of Asians Americans in the U.S. population remained less than 1% before the end of World War II. And beginning in 1970, the U.S. Census first separated out questions about "race" from questions about whether a person was Hispanic, or not (U.S. Census, 2002). This disaggregation of "Spanish/Hispanic origin or descent" (1980 wording) from country of origin is also mandated by the U.S. Office of Management and Budget (1997), and is now used in the recent decennial censuses and the American Community Survey.

However, the proportions of these two demographic groups in the population have grown substantially in recent years. Between 1980 and 2005, the public-school population of Asians/Pacific Islanders (hereafter, "Asian") has grown by 260 percent (KewalRamani, Gilbertson, Fox, & Provasnik; 2007). And, the part of the U.S. population of Hispanic ethnicity is now the largest minority population—surpassing the Black population—as of the turn of the 21st century (KewalRamani et al., 2007). Thus, in summary, as of 2012, the percentage of White public middle-school students in the U.S. was 55%, while 5% were Asian, 16% Black, and 23% Hispanic.[2]

Racial and ethnic identities and the differences among them are unstable concepts, constantly under dispute and transformation(Omi & Winant, 1994). During the first half of the 1800s, all people were counted by age, sex, and "color" (the choices were White, Black, or Mulatto) in the U.S. census (U.S. Census, 2002). And, as explained above, the modern conception of "Hispanic" was formulated in the middle of the 20th century as an ethnicity, rather than a race.

Similarly, since 1977 the federal government has also collected data systematically about the ethnicity and race of students. And, since 2007, these federal guidelines specifically ask students (or their parents) a two-part question about students. First the respondent (a parent or guardian, or the child herself/himself) is asked to identify the student's ethnicity as either Hispanic or not. Second, the respondent is asked to identify his or her race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander) or races (U.S. Department of Education, 2008).

During the last four decades, the demographic composition of public schools has changed substantially. In 1970, approximately 82% of all public-school middle-school students were White. About 13% were Black students, with less than 4% Hispanic. Asian and Native-American students comprised less than 1% of public-school middle-school students. However, 40 years later, in 2010, approximately

---

[2] Author's estimation using U.S. Census data, archived by the Minnesota Population Center at the University of Minnesota (Ruggles et al., 2010).

55% of all public-school middle-school students were White. About 16% were Black, with more than 23% Hispanic, and nearly 5% Asian.

Racial/ethnic disparities in school suspension. Decades of research have also documented important racial/ethnic disparities in the rate at which students are suspended out-of school (Edelman et al., 1975; McCarthy & Hoge, 1987; Skiba et al., 1997; Townsend, 2000; Losen & Skiba, 2010; Kinsler, 2011; Petras et al., 2011; Skiba et al., 2011; Sullivan et al., 2013). As Skiba et al (2011) commented, in a study of Black and Hispanic disproportionality in school discipline, "race is not neutral." By analyzing school disciplinary records in 364 elementary and middle schools, the authors, using logistic regression analysis, concluded that the probability that either Black or Hispanic students would be suspended from school was greater than for White students, for the same or similar misbehavior. Skiba and his colleagues also point out, in their study, that they drew a sample of schools that implemented one particular data-collection system, but they provided evidence that the sample represented the U.S. population of public-school students along dimensions of gender, race/ethnicity, and special education.

In another study also focused primarily on disproportionality in school suspension by race/ethnicity, Wallace, Goodkind, Wallace, and Bachman (2008) analyzed data collected from a nationally representative survey of high-school students. Using logistic regression analysis, they estimated that the odds that Black students reported being suspended from school were more than three times the odds for White students. Comparing Native-American students (a category not explored by Skiba et al) to White students, the odds were twice the odds for Whites, while the odds-ratio comparing Hispanic students to White students was 1.7. Wallace et al (2008) and other researchers (Morris, 2005; Sullivan et al., 2013) have found that Asian students, as a demographic group, have lower odds for reporting suspension, compared to White students. While these two studies differ in data-collection methodology—with Skiba et al (2011) utilizing administrative records, and Wallace et al (2008) using individual survey data—both studies highlight the continued pervasiveness of disproportionate risk of school suspension by student race. This disproportionality has changed little since the research of Edelman et al (1975) four decades ago. And so, in the current research, I have focused particularly on differences in the risk of school suspension by race/ethnicity.

But, why are Black students, as a demographic group, at the greatest risk of being suspended (Skiba et al., 1997; Townsend, 2000; Raffaele Mendez & Knoff, 2003; Nichols, 2004; Hinojosa, 2008; Kinsler, 2011; Sullivan et al., 2013) ? Kinsler (2011) attributed much of the differences in suspension rates between Black and White students to differences in the schools they attend. Most other researchers (Ferguson, 2000; Gregory et al., 2011; Hinojosa, 2008; Payne & Welch, 2010; Skiba et al., 2011; Sullivan et al., 2013) concluded that differences in the rate of suspension, by race, were attributed to pervasive differences in the school experiences of Black and White students. For example, Gregory, Cornell, and Fan (2011) combined their analysis of Virginia state records of school suspensions with the results from a survey of 9th grade students, sampled from more than 90% of the high schools in Virginia. They concluded that suspension rates for Black students and White students, which differ substantially, were not attributable to differences in either school size or SES. Rather, their findings are more consistent with those of Ferguson (2000) and Payne and Welsh (2010)—that the school experience for Black students is decidedly different, with Black students being subject to more punitive discipline than other students who committed the same offenses.

Less research has been conducted concerning the punishment of Hispanic students in public schools, in the United States.[3] However, several studies have also found that the rate of suspension for Hispanic students is higher than that for White students. For example, in addition to the differences in the rates of suspension they found between Black students and White students, Skiba et al (2011) also found that the odds of suspension for middle-school Hispanic students were higher than for White students. In another study of middle-school suspensions, conducted in the primarily Hispanic and Black schools in Miami, Florida, Arcia (2007) found that rates of suspension for Black students were higher than for Hispanic students. Finally, Sullivan Klingbeil, and Van Norman (2013) concluded that the odds of school suspension for Hispanic students, while lower than for Black students, was higher than the odds of school suspension for White students.

Conclusion: Specific Research Questions

In the above discussion of the background and context of my research, my predominant focus has been on studies that report estimates of the probability ("whether") a student will be suspended in a given academic year—known as the "suspension rate." The NCES (2012) report, described above, for example, estimates

[3] For the results in this paper, I use the term Hispanic, following the common demographic shorthand descriptor used by many. Other researchers prefer the term Latino/a.

explicitly the probability that students were suspended out-of-school at least one time ("whether"). And, several studies have noted high rates of suspension for middle-school students, as compared to other public-school levels (Theriot & Dupper, 2010; Losen & Skiba, 2010; Raffaele Mendez & Knoff, 2003). Consequently, the results of these studies, as well as my own professional experience as a middle-school principal, incentivize me to focus on similar probabilities—that is, *whether* middle-school students have been suspended from school, in this research.

However, as Lortie (2009) remarked, students are "only gradually socialized into complying with the norms of orderly behavior" (p. 123). Therefore, I believe that it is also important also to analyze *when*, during the school year, these suspensions occur. For school personnel faced with the ever-changing landscape of how the experiences of students in school change over time, understanding who is subject to school suspension, and *when* during the school year these occur, seems to me to be fundamental. Furthermore, I hypothesize that the risk of suspension likely differs by student race/ethnicity, and thereby potentially influences decisions made by school principals about whether students *should* be suspended from school.

Therefore, in this research, I conduct a case-study in the public middle schools of one metropolitan region to describe *whether,* and if so, *when*, middle-school students were first suspended during the course of a single school year, and how the occurrence of first suspension differed by student race/ethnicity. My specific research questions are:

1. How does the risk of first suspension differ for middle-school students, over the course of a school year? In particular is the risk of suspension higher at the beginning of the school year than it is at the end?
2. Is the annual temporal profile of risk of first suspension higher, and of a different shape, for Black and Hispanic students than for White students?

I address these questions by capitalizing on a rich and extensive database provided to me by a state agency. As my database contains the daily history of suspensions over one academic year, for three cohorts of students (beginning the school year at the 6th, 7th, and 8th grade levels), when I address each research question, I also investigate implicitly whether the within-year profile of risk of first suspension differs by grade level.

Research Design

Site

For this case study, I focus on students enrolled in all public middle schools located in one metropolitan region surrounding a medium-sized city. I have concealed the name of this city to protect the confidentiality of the students whose middle-school careers are examined. I argue that this was an ideal site for my case study for several reasons. First, the sample was large enough to provide a sizeable number of incidents of out-of-school suspensions—a subsample of students comparable in size to other research on this topic, reviewed above (e.g. Raffaele Mendez & Knoff, 2003; Theoriot & Dupper, 2010; Sullivan et al, 2013). Second, I have access to information about the students' grade-level in school, allowing me to compare risk of first suspension among students in the 6th, 7th, and 8th grades, over one school year, in middle school. Third, my sample contains racial/ethnic diversity ample enough to support the investigation of any disproportionate risk of suspension for Black and Hispanic students, compared to White students.

This metropolitan region contained approximately a half-million residents. Approximately four-fifths of the residents were non-Hispanic White, while Asian, Black and Hispanic residents constituted roughly equal proportions of the population. The largest district served more than 20,000 students from kindergarten through 12th grade. Each of the other districts was smaller and served less than 6,000 students. Most of the school districts served 6th through 8th grade students in a traditional middle-school configuration, but a few of the districts served only 7th and 8th grade students in a two-year configuration of middle schools. Consequently, I excluded approximately 350 6th-grade students who attended "intermediate" schools in these districts.

Dataset

I have been provided confidential access to the educational records of all of the middle-school students enrolled in any middle school in this region at the beginning of the 2009-10 school year. My dataset contains information on the date each student was enrolled in school, was first suspended out-of-school (if suspended at all), left school for another reason (perhaps moving to another school), or completed the school year without formal discipline by school officials. I constructed this merged dataset by combining information from three linked datasets containing individual student records that I obtained, with permission, from a government agency.

The first dataset that I integrated into my larger data assembly consisted of student-enrollment information. It contained the date of enrollment and departure from each of the public middle schools in the region, for each student, identified by a unique student number. More than 15,000 students attended one of the middle schools in the region for at least a portion of the school year. Students enrolled and left schools throughout their school career. However, in this sample, 97% of the students who were enrolled in one of these middle schools during the first three weeks of school remained enrolled until the end of the school year. Consequently, in my analysis, I focus on the records of the 13,256 students who enrolled at one of the schools within the first three weeks of the school year.

The second dataset that I incorporated into my work contained demographic information. The contents of this dataset provide information on the two variables that I treat as time-invariant covariates in my analysis. In this dataset, student grade-level was recorded as each student's grade at the end of the school year. Additionally, schools recorded students' race/ethnicity when students were enrolled initially into the public school, and students—or their parents—identified their primary racial/ethnic classification.

The third dataset that I incorporated into my analysis recorded the date of each disciplinary incident that was reported to the state by the schools and districts in this sample, by student ID. Thus, it contains information about in-school suspensions, out-of school suspensions, and expulsions reported by the schools and districts. But, for the current investigation, I limit my investigation to the *first* occurrence of out-of-school suspension of the school year, for any student. I do this—in a similar vein to Petras, Masyn, Buckley, Ialongo, and Kellam (2011)—because the timing of suspension may predict subsequent school disciplinary issues. In my subsequent research, in my doctoral thesis, I plan to investigate the occurrence and timing of school suspension longitudinally over each student's entire middle-school career, incorporating information on the first and all subsequent suspensions in a multiple-spell survival analysis.

Sample

I analyze data on an analytic sample containing 13,256 middle-school students, who attended the 31 schools at my site, over the 2009-10 school year. Following Kinsler (2011), I included students at all public middle schools that served students in grades 6th through 8th, or grades 7th and 8th, exclusively.

Measures

As is required for continuous-time survival analysis, I formatted my combined dataset as a person-level dataset (Singer & Willett, 2003) with each sampled student contributing one row of data. In each row, in addition to ID codes that identified the individual student, school, and district, I coded the values of the following variables:

Outcome:

Conceptually, in Cox-regression analysis, the analytic outcome is an expression of the risk of suspension, called the hazard rate. It is defined as the conditional probability that an individual will experience the event of interest—in this case, a first suspension from school—per unit of time (Singer & Willett, 2003) .

It is constructed implicitly during the analysis from the values of the following two outcome variables:

TIME (t) is a continuous variable that records the passage of time, measured in school days, from the origin (the student's entry date into the school), until a student experienced a first suspension from school during the 2009-10 school year, or was censored. Note that, in my sample, 92.3% of the students were *censored*—that is, they did not experience a first suspension while observed—either because they moved away on a specific day during the school year before any record of suspension, or because they completed the school year without being suspended from school at all.

FIRSTSUS is a dichotomous variable, coded as 1 if a student experienced a first suspension from school during that year, and 0 if not suspended (the latter being the *censored* students).

Question Predictors:

GRADE is a vector of three time-invariant dichotomous predictors (GRADE6, GRADE7, GRADE 8) to record the grade in which each student was enrolled during the year of the investigation. Each variable is coded 1, if the student was enrolled in the respective grade, 0 otherwise. I omit the 6th grade category from my statistical models to provide a reference category.

RACE is a vector of time-invariant dichotomous variables that record the primary race/ethnicity of each student, using definitions provided by the state. The vector includes the variables ASIAN, BLACK, HISPANIC, NATIVE, and WHITE, each coded 1 or 0, to identify the student's race/ethnicity. I omit the predictor WHITE from my statistical models to define a reference category.

Data-Analytic Plan

In my dataset, the occurrence of out-of-school suspensions was recorded by the state to the nearest day, guiding me to use continuous-time survival analysis to address my research questions. While parametric approaches to continuous-time survival analysis (e.g. Weibull, 1951) might also shed light on the issue, for this paper, I elected to apply a popular semi-parametric method—Cox-regression analysis (Cox, 1972)—as a strategy for describing differences in the risk of suspension among students from different groups. The Cox-regression strategy is a popular technique but may seem to be a puzzling analytic choice for this study, because it assumes no particular parameterization of the relationship between hazard rate and time. In fact, if the baseline hazard rate were left un-estimated, Cox-regression would provide no information about *when* the risk is highest. However, once I include the two question predictors—*GRADE* and *RACE*—into the model, comparing the estimated of risk of first suspension between groups provides a sensible method to analyze the phenomenon. In addition, the temporal dependence of the hazard rate on time can be estimated post-hoc in the data using conventional methods of Kaplan-Meier (1958) estimation. The obtained sample risk profile—combined with the findings of grade and race-dependence from fitted Cox-regression models—provide answers to my research questions.

In each Cox-regression model that I specify and fit below, my outcome was the (log) hazard rate. The right-hand side of each hypothesized model then contained an unspecified (log) baseline hazard function (that was not estimated directly during the model fitting, but was recovered from the sample data post-hoc), plus a weighted linear combination of the effects of the hypothesized predictor variables, each accompanied by its respective slope parameter. In each model, I accounted for the clustering of students in schools by adding a random effect of school, or "frailty," (Vaupel, Manton, & Stallard, 1979) that acknowledges differences in estimated risk attributable to what school a student attended. Finally, I used the Efron (1977) method for handling tied outcomes, when two or more students have identical estimated times to the event.

RQ1: How does the risk of first suspension differ for middle-school students, over the course of a school year? In particular, is the risk of suspension higher at the beginning of the school year than at the end?

As my sample of public-school students contains middle schools with students at three grade-levels, I first began by investigating whether the within-year profile of risk of first suspension differed *in level* among 6th, 7th, and 8th grade students. To describe the distribution of the risk of first suspension over the school year, I fitted the following Model A, in the full sample:

$$\ln h\left(t_{ijk}\right) = \ln h_0\left(t_{ij}\right) + \boldsymbol{\alpha}'\boldsymbol{GRADE}_{ijk} + v_k \qquad (1)$$

where $h\left(t_{ijk}\right)$ represents the hazard rate describing the risk of first suspension of student *i*, on day *j*, in school *k*. Parameter v_k represents the random effect of school *k*, and was assumed to be gamma-distributed.

Parameter vector $\boldsymbol{\alpha}'$ represents the population fixed effects of student grade on the log-hazard-rate of first suspension, thereby permitting the *level* of first-suspension risk to differ by grade. If an estimate of one of the elements of this vector is positive and statistically significant, then I know that the hazard rate describing first suspension is elevated in the corresponding grade. I have treated 6th grade as a reference category. Thus, if an estimate of α_1 was positive and statistically significant, then I concluded that the risk of first suspension was higher for 7th grade students than for 6th grade students.

Note that, in this model and in the standard Cox-regression analysis, the effect of *GRADE* is specified as a main effect only. Then, anti-logging the associated parameter estimates, the several fitted hazard rate profiles in each grade, defined by parameter vector $\boldsymbol{\alpha}'$, have similar shapes and are *proportional* to each other. To facilitate interpretation of these latter model parameters, in my tables of fitted models, I have also included a column containing the anti-logged parameter estimates that can then be interpreted as *hazard ratios.* Each hazard ratio (the anti-log of α_1, say, corresponding to 7th grade) multiplies the *risk* of first suspension in the reference 6th grade to obtain the *risk* of first suspension for 7th grade students. A similar interpretation comparing 8th graders to 6th graders applies for α_2.

Finally, to this baseline proportional-hazards model, I tested the addition of an interaction between *GRADE* and linear *TIME,* to investigate whether the proportional-hazards assumption was violated and therefore that any pattern of yearly risk differed by grade, in *shape* as well as *level*. After exploring various parameterizations, I found that the following non-proportional hazards model, that included an interaction between each *GRADE* and linear time, was the most appropriate to fit the data:

$$\ln h\left(t_{ijk}\right) = \ln h_0\left(t_{ij}\right) + \boldsymbol{\alpha}'\boldsymbol{GRADE}_{ijk} + \boldsymbol{\gamma}'(\boldsymbol{GRADE}_{ijk} \times TIME) + \upsilon_k \qquad (2)$$

and I obtained the corresponding estimated parameters and fit statistics.

In the semi-parametric Cox approach to continuous-time survival analysis, the baseline hazard rate profile, contained as an “intercept” in the model, requires no particular parameterization and is not estimated directly during model fitting. However, after model fitting, it can be recovered from the data, using the Kaplan-Meier (1958) estimator, and plots of the fitted risk over time can be displayed, appropriately shifted according to the magnitudes of the estimated elements of parameter vector $\boldsymbol{\alpha}'$. Thus, after fitting this non-proportional hazards model, I was able to obtain and present plots of the fitted risk of the *TIME* to first suspension, by *GRADE,* throughout the school year, to summarize the occurrence and timing of first suspension.

RQ2: Is the annual temporal profile of risk of first suspension higher, and of a different shape, for Black and Hispanic students than for White students?

To address my second research question, I added my principal question predictor—describing the annual temporal profile of risk for students of different racial/ethnic categories—to the model fitted above, as follows:

$$\ln h\left(t_{ijk}\right) = \ln h_0\left(t_{ij}\right) + \boldsymbol{\alpha}'\boldsymbol{GRADE}_{ijk} + \boldsymbol{\gamma}'(\boldsymbol{GRADE}_{ijk} \times TIME) + \boldsymbol{\beta}'\boldsymbol{RACE}_{ijk} + \nu_k \quad (3)$$

where parameter vector $\boldsymbol{\beta}'$ represents the main effect of race on the log-hazard-rate of first suspension, thereby permitting the *level* of first-suspension risk to differ by *RACE,* with White students as the reference category. I also tested whether the profile of risk for students of different race/ethnicity differed in shape, by including interactions between race/ethnicity and time, and I retained these terms if they were required. Additionally, I included interactions between grade-level and race/ethnicity, and I retained these terms if they were required.

Findings

On virtually every school day of the academic year, in the geographic region that I studied, a relatively small number of middle-school students were suspended from school for the first time of the school year. In fact, in these sample schools, only five school days were entirely free of first suspensions over the course of the academic year: the first two days of school, the last day of school, and two other days throughout the 2009-10 school year. On average, over the entire year, the number of first suspensions was approximately 5 per school day. But, the number of *first* suspensions per day in the region ran from a minimum of zero, on five school days, to a high of 18 suspensions, on one particularly “hazardous” day for students to experience their first suspension, in mid-September.

In Table 1, I present the results of fitting Cox-regression models describing how the occurrence of first suspension depends on the passage of time (in school days), student grade level, and student race/ethnicity. For each model, I include parameter estimates, associated standard errors, the corresponding estimated hazard ratios, and approximate p-values. To account for the clustering of students within schools, I provide an estimate of the between-school correlation, θ, in row 11. So, all of the hazard ratios for the parameters in each model are conditional on the estimated value of ν_k, which is identical for each student in school k. In row 13, I provide the log-likelihood goodness-of-fit statistic for each model. And, in rows 14 and 15, I provide the log-likelihood fit statistic for each model, and the results of a general linear hypothesis (likelihood-ratio) test in which I compare the fit of each model with that of the previous model.

[Insert Table 1 about here]

RQ1: How does the risk of first suspension differ for middle-school students, over the course of a school year?

I first fitted equation 1 as Model A to estimate differences in hazard rate among grade levels, while retaining an assumption of proportionality among the fitted hazard-rate profiles. In this model, I estimate that the hazard rate for 8th grade students was 23% higher than for 6th grade students, and that the difference was statistically significant. The hazard rate for 7th grade students was 16% higher than for 6th grade students, and of marginal statistical significance. (Notice that when interpreting these anti-logged hazard ratios, I assigned a value of "1" to the reference category—6th grade.)

Furthermore, all models included the random effect of school, υ_k. Thus, in Model A, I estimated θ, the inter-school variation in hazard rates (as assumed to be gamma-distributed) was 1.13. I note then that the overall suspension rate in this sample varied substantially by school. Three middle schools recorded no suspensions for the entire year. While at the other extreme, three schools recorded overall suspension rates (*whether* students were suspended, or not) of more than 17% of the student population.

After fitting Model A, I noted a significant violation of the modeling assumption that hazard-rate profiles must be *proportional* among grades. In row 16, I supply the results of a diagnostic test of the assumption that the listed hazard rates for grades 6th through 8th grade (represented by included predictors GRADE7 and GRADE8) are proportional to each other, by inspecting the slope of the values of the Schoenfeld (1982) residuals versus time (χ^2=8.41, *df*=2, *p*=0.02), and detected a violation of the proportional-hazards assumption built into Model A. So I addressed this issue by including the two-way interaction of *GRADE* with *TIME*, specified in equation 2, and I list the fitted model as Model B. Including the two $GRADE \times TIME$ interaction terms alleviated violations of the proportional-hazard assumption (χ^2=4.02, *df*=4, *p*=0.40) and provided an interesting insight into the effect of *TIME* on the differences in hazard rate among students in different grades.

In Model B, the population profile of risk of first suspension differed substantially among the prototypical 6th, 7th and 8th grade students. On average, the *initial* hazard rate of first suspension of 7th grade students was 44% higher than the comparable risk of suspension for 6th grade students. And, the initial hazard rate first suspension of 8th grade students was 81% higher than comparable risk for 6th grade students. Both of these estimated hazard rates, by *GRADE*, are estimates of the risk of first suspension *at the beginning of the school year.* However, I also note that the two hazard ratios for 7th and 8th grade students, compared to 6th grade students, decreased throughout the school year.[4] For example, I estimate that, on each school day, the fitted hazard rate for 8th grade students was approximately 99.5% of the hazard rate for these students on the previous day. Consequently, while the hazard rate for 8th grade was 81% higher at the beginning of the school year, compared to 6th grade students, during the school year this difference diminished as a function of *TIME*. But by the end of the school year, the risk of first suspension was 20% *lower* for 8th grade students, compared to 6th grade students.

In Figure 1, I present a plot of the estimated hazard rates versus time (in school days), by grade, incorporating both the slope estimates associated with grade and the baseline hazard rate profile obtained using the Kaplan-Meier (1958) strategy, and employing a kernel smoother. Thus, this plot incorporates, visually, the effect of both the level of risk (the hazard ratio) and the interaction of risk with *TIME* for 7th and 8th grade students. Each of the three fitted profiles plotted in the figure reiterate the basic shape of the underlying baseline risk, but they differ in their orientation, by grade. Notice that the obtained empirical hazard rate—and the corresponding risk of first suspension—is *not* constant over the school year. Indeed, the variability in hazard rate for all three grades appears somewhat cyclic over time.

[Insert Figure 1 about here]

The risk of first suspension was highest near the beginning of the school year for 8th grade students, and also, to a lesser extent, for 7th grade students. In particular, during the month of October—school day number 22 through school day

[4] Although the hazard rate for the interaction between 7th grade students and *TIME* was not statistically significant, the rate fell consistently between 6th and 8th grade in both level and in the direction of the interaction with time. Consequently, I retained this interaction term for the entire analysis.

number 41—a particularly high rate of first suspension was evident. On average, 7.6 middle-school students in the sample experienced their first suspension of the school year each school day during October, including 38 6th grade students, 45 7th grade students, and 69 8th grade students. This average hazard rate for October was substantially higher than during any other month of the school year.

Interestingly, the period of lowest risk of first suspension occurred immediately after the end of the first academic quarter. The substantial "trough" in the hazard rate coincided, roughly, with the 45 school days from early November through the last week of January. During those three months, about 4.3 first suspensions occurred during each of those days—substantially lower than the total average daily rate of 5 suspensions per day. The second peak, which occurred around school day 100 (in early February), and the second trough, which occurred around school day 120 (in early March), described hazard rates that were closer to the average daily risk. An average of 6.2 first suspensions occurred in February, and 4.2 first suspensions occurred in March.

At the bottom of Figure 1, I include a risk table, displaying the number of students who were first suspended (and the number of students at risk), by academic quarter. I have disaggregated the risk table by grade level. Note that the sample of 8th Grade students had the most pronounced risk of first suspension at the beginning of the year, and then substantially diminished. Meanwhile, in contrast, the risk of first suspension for 6th grade students was highest during the 1st and 4th academic quarters, while at students were at lower risk during the 2nd and 3rd quarters. Also, notice that the quarterly summaries document that the number of first suspensions of 8th grade students during the 4th quarter was indeed lower than the number of first suspensions for 6th grade students.

Overall, I concluded that the timing of first suspensions for much of the latter part of the school year was essentially *random,* regardless of grade, as Lortie (2009) described. In fact, from the middle of the school year until its end, the hazard estimate was relatively constant, within each grade. A larger number of students in the sample who were suspended from school at all, were suspended during the first quarter—and the risk of first suspension was substantially lower during the second quarter than during the first quarter. Still, the pattern of risk for the third and fourth quarters was steadier.[5] Also, note that the estimates of risk by grade level cross late in the school year. At the end of the school year, the risk of first suspension for 8th-grade students was *lower* than the risk for 6th-grade students.

RQ2: Is the annual temporal profile of risk of first suspension higher, and of a different shape, for Black and Hispanic students than for White students?

In my sample, overall differences by race/ethnicity, in *whether* students were first suspended or not are stunning: 34% of Black students in the sample were suspended, compared to 2% of Asian students, 6% Hispanic, 6% Native American, and 3% White, during the period of time in which each student was at risk. In Model C, I added the vector *RACE* to the previous model, as specified in equation 3. My results, again reported in Table 1, confirm the enormous disparity in the risk of

[5] A keen observer might notice that there was a discrepancy between the number of students at risk at the beginning of each quarter, less the number suspended, compared to risk set for the next quarter. Recall that a few students in each quarter were *censored* at various points throughout the year, likely because they moved to another school.

suspension by student race/ethnicity. I found that Black students were at more than 10 times the risk of first suspension, and Hispanic students were at nearly 60% greater risk, compared to the reference category—White students. Asian students, as a demographic group, were at substantially lower risk than any other group. Compared to White students, I found that the risk of first suspension for Asian students was less than half the risk for White students. My estimate of the risk for Native American students was not significantly different than for White students, likely because the sample of Native students attending public schools was quite small, and consequently I had little statistical power to detect the difference.

In Model C, I conducted a global test of the assumption of proportional hazards. I found no statistically significant evidence that this assumption was violated in Model C (χ^2 =11.42, *df*=8, *p*=0.179). I found that the parameter estimates for *GRADE* were somewhat higher, and that parameter estimates associated with the effect of the $GRADE \times TIME$ interaction remained statistically significant, following the introduction of the main effects of *RACE.* I also tested whether the main effect of *RACE* interacted with *GRADE.* I found no statistically significant evidence to support this more complex model (χ^2 =9.23, *df*=8, *p*=0.323). Controlling for *RACE,* I concluded that 8th grade students were at more than twice the risk of first suspension than were 6th grade students at the beginning of the school year, and that the risk for 7th grade students was about 70% higher than 6th grade students.

Finally, regarding estimates of school-level differences in students' risk of first suspension, I note, in Model C, that my estimate of inter-school variance in hazard rate, θ, was 0.64. This estimate was about half the estimate of θ for Models A and B. So, differences in student *RACE* accounted for a substantial portion, but not all, of the inter-school variation, suggesting that school segregation by race/ethnicity is an important explanation of the risk of first suspension.

In Figure 2, based on the parameter estimates from fitting Model C, I present fitted survival plots for four prototypical students, identified by their race/ethnicity. (Due to the small number of Native American students in my sample, I did not present the corresponding fitted survival profile for these students in the figure.) Each of the prototypical students is of "average" grade level—a combination of 6th, 7th, and 8th grade. And each prototypical student attended an "average" school, with v=1. At the beginning of the school year, all of the students are at risk of experiencing their first suspension, yet (obviously) none of the students have experienced their first suspension of the school year prior to the beginning of the school year.

[Insert Figure 2 about here]

Notice, first, in Figure 2, that most Asian, White, and Hispanic students "survived" the year without being suspended from school. The low level of incidence of first suspension for these students is quite striking. Consequently, it is difficult to discern the "seasonal" patterns of risk described in Figure 1. For example, for White students, at the end of the first academic quarter (school day #45), the fitted survivor function for a prototypical White student is still more than 99.1%—even after surviving the relatively "hazardous" month of October.

However, what I highlight most vividly in Figure 2 is that my estimate of the risk of first suspension for Black students is more than 10 times that for White students. Consequently, the fitted survivor plot for Black students provides a stark contrast with the fitted survivor plots for Asian, Hispanic, and White students. At the end of the first academic quarter, only 92% of the Black students in the sample have "survived" the quarter. 8% of Black students have already experienced their first suspension of the school year by early November. Even the statistically significant difference in hazard ratios between Hispanic students and White students are dwarfed in Figure 2 by the enormous disparities in rates of first suspension for Black students. And this disparity, by race/ethnicity—particularly for Black students—continued throughout the school year.

Threats to Validity

My study highlights two main stories: that the risk of first suspensions is higher at the beginning of the year, but only for students who are in 8th grade, and to a lesser extent to 7th grade; and that the risk for Black students is substantially higher than for White students. However, in this study, I describe patterns of first suspension in only one region of the country. Therefore, I cannot generalize my findings to public schools in other locations. There are likely aspects of the demographic and regional characteristics of site of this case study that influence the findings directly. Thus, I am cautious about making claims regarding the external validity of this research. In addition, in this study, I describe the risk of suspension for middle-school students during the first years of the Obama administration, before contentious debates about collective bargaining for public-school teachers began in earnest. Patterns in school suspension may be different now.

In this research, I was forced by limitations in my data to focus on the risk of first suspension over one complete year of middle school, within each of three grades. My dataset did not include any information about students' previous suspension record. For instance, I did not know, and could not incorporate into my analyses, whether a student documented as being "first suspended" in 8th grade had actually been suspended earlier, in the 6th and 7th grades. Consequently, my results may be substantially different than results about the risk of first suspension, if I had been able to incorporate the discipline history for students attending the same school during previous years into account. To reconstruct a profile that describes the risk of first suspension across an *entire* middle-school career would require three years of longitudinal data, as I would need to follow each student from their entry in 6th grade through the end of middle school. Then, in addition, I could account for multiple occasions of suspension. I plan to adopt this approach in my future thesis research, with new longitudinal data.

Another threat to the validity of the current analysis was that the students in this sample attended middle schools with substantially different suspension rates, by school. As I noted earlier, these frailty effects make a statistically significant contribution in all of my fitted models. But does this imply that a few schools with very high suspension rates drove my results? In order to test the sensitivity of my findings to this clustering, I compared the results of fitting Model C with those from fitting the same model, but having dropped the three schools with the highest estimated value of the parameter modeling the school-level frailty, υ. The results of this sensitivity analysis were very similar with those obtained in the full sample. I also refitted Model C after dropping the three schools with the lowest frailty (the three schools recording no suspensions for the year). But again, my parameter estimates were remarkably robust, and well within the confidence intervals of the full model.[6]

Interestingly, perhaps a more basic question about my case study asks *why* students transition from elementary school to middle school at all?[7] Schwerdt and West (2013) posed this question in a study in which they estimated the impact student achievement, student absences, and grade 10 dropout rate of attending public schools with different grade configurations. Similar to the case in my sample, Schwerdt and West (2013) noted that about 88% of the public-school students in Florida attended 6th through 8th grade middle schools. They concluded that entering middle school was associated with a drop in student achievement, an increase in student absences, and a subsequent higher rate of dropout rate by grade 10 (Schwerdt & West, 2013).

In light of their findings, I conducted a sensitivity analysis to investigate first suspension in the five middle schools in the sample (in five different small school districts) that were configured as 7th & 8th grade middle schools, rather than the

---

[6] One of the schools in the sample likely failed to document how student discipline was enforced. While no students were suspended during 2009-10, during the 2008-09 school year 29 students (6%) were suspended. And, during the 2010-11 school year 35 students (7%) were suspended.

[7] For many years, the public schools in Cambridge, Massachusetts were configured as K-8 schools feeding into one large high school: Cambridge Rindge and Latin. The new configuration of twelve K-5 schools feeding into five 6-8 middle schools was implemented in the fall of 2012 http://www3.cpsd.us/ia2/ia. Conversely, in Vermont (where I presently reside), the majority of public elementary schools are configured as K-6 schools. And, many of these elementary schools feed into a single 7th -12th grade school in each "Supervisory Union."

more common 6^{th}-8^{th} grade middle schools. Four of the five 7^{th} & 8^{th} grade middle schools had very low suspension numbers. Thus, I am reluctant to draw firm conclusions, because for these four schools, each recorded less than 5 suspensions per grade level for the year. The other 7^{th} & 8^{th} grade school, apparently, had a substantial “incident”. Nine 7^{th} grade students were all suspended for their first time of the school year on the same day in April 2010. This incident alone accounted for more than a third of the total number of first suspensions for the school year at that grade. Perhaps these 7^{th} graders in 7-8 schools behaved more like 6^{th} grade students in the other 6-8 middle schools—being relatively complacent for most of the school year only to wreak havoc once spring came along? This anomaly about behavior in 7-8 schools also warrants further investigation using multi-year data. And perhaps a larger sample, including several metropolitan regions with more junior high schools, might shed some light on this issue.

Finally, an important threat to the validity of my findings concerns how the structure of my dataset itself. Note that in Model B, I included an interaction between *GRADE* and *TIME,* and produced Figure 1—which did not account for the effect of *RACE.* However, the “seasonal” variation in risk revealed in this figure was driven primarily by the high rates of first suspension among Black students, compared to other students in the sample. If I had chosen to answer the second research question first, and then refined my results by subsequently including the effect of *GRADE,* I would have chosen to highlight the effect of *RACE* first—and noted that the risk of first suspension for Black students, compared to White students, was highest at the beginning of the school year. Would I have concluded that the differential effect of *TIME* on different groups pointed to an interaction between *RACE* and *TIME*, rather than between *GRADE* and *TIME?*

In order to test this rival hypothesis, I reversed the order of predictor inclusion and refitted the critical models. Thus, I first fitted a model, in which the profile of risk was hypothesized to depend only on the main effects of *RACE:*

$$\ln h(t_{ijk}) = \ln h_0(t_{ij}) + \boldsymbol{\beta}'\boldsymbol{RACE}_{ijk} + \nu_k \quad (4)$$

I tested whether the proportional-hazards assumption was violated, and I concluded that it was not (χ^2=5.92, *df*=4, *p*=0.20), and that the parameter estimates for the four *RACE* categories were nearly identical.

Then, I added the effect of *GRADE* to this model, in which the profile of risk was hypothesized to depend on only the main effects of *RACE* and *GRADE:*

$$\ln h(t_{ijk}) = \ln h_0(t_{ij}) + \boldsymbol{\beta}'\boldsymbol{RACE}_{ijk} + \boldsymbol{\alpha}'\boldsymbol{GRADE}_{ijk} + \nu_k \quad (5)$$

In this model, I concluded that there is *marginally* statistically significant evidence of a violation of the assumption of proportional hazards in this main effects model (χ^2=12.28, *df*=6, *p*=0.056). My detailed analysis of this test revealed *two* explanations. First, I noted that the parameter α_2, which described the impact of a student being in 8th grade, compared to 6th grade, was responsible for a large portion of the global test statistic (χ^2=6.86, *df*=1, *p*=0.01). However, the parameter β_2, describing the impact of a student being Black rather than White *also* warrants further investigation (χ^2=5.48, *df*=1, *p*=0.02), in a subsequent study, with a larger sample.

I also tested whether including an interaction between *RACE* and *GRADE*, added to the above model, specified in equation 5, might prove informative. But when I conducted a likelihood-ratio test comparing this model with the previous, the additional *eight* interaction terms (four comparison *RACE* parameters, multiplied by two comparison *GRADE* parameters) did not improve the model statistically (χ^2=10.11, *df*=8, *p*=0.26).

As I explained in my introductory comments on **Differences in Suspension From School By Race or Ethnicity**, the demographic composition of public schools is rapidly changing, with most schools enrolling increasing numbers of Asian and Hispanic students. Consequently, I argue that simple comparisons in risk between Black students and White students are likely inadequate for future studies. And comparing the risk of suspension for Black students to non-Black students or "minority" students with Whites will probably obscure more differences than such a study might reveal. Thus, I am indeed wary about making broad claims about my research. Modeling the risk with even two categorical variables forces me, and other researchers, to make choices about the primacy of the variables—particularly when ethnicity and race are the subject.

Discussion

As I noted in my introductory comments, the work of school principals is unpredictable in many respects. Urgent demands—including vigilant attention to order and discipline—require school principals to weigh both the facts of the situation *and* the policies and procedures that serve as guides to suspending students. In this context, using survival analysis to simply describe the rates of out-of-school suspensions may seem rather banal, compared to other pressing matters. Still, I believe it somewhat puzzling that no research in the scholarly literature has tested whether or not the risk of school suspension was highest at the beginning of the school year.

My results suggest that the risk of first suspension of the school year was indeed higher at the beginning of the school year, *for 8th and 7th grade students*. For example, in Figure 1, I demonstrate that, once the school year was fully underway (after several weeks), principals were now seemingly fully engaged in maintaining order and discipline. As the hazard rate peaked during the month of October, the use of the "valuable educational device" that Justice White championed seemed to play a substantial role in preserving order and discipline in schools. Once the behavioral expectations were set during the opening days of the school year, it apparently was time for principals to "bring the hammer down"—dealing with disciplinary issues in a more decisive and formal way.

Still, *randomness* in the risk of first suspensions played a larger role than I expected. The risk of first suspensions, after an initial flurry, seemed to settle into a constant hazard rate. Thus, this case study provided evidence that the *first* out-of-school suspension for many students was imposed rather randomly throughout the school year. No formal hypotheses about why the risk of first suspension was lower in November and December seemed apparent to me. Nor, why, perhaps, did the risk of first suspension apparently rise again in February? Consequently, using Cox (1972) regression—with a baseline hazard rate that undulated with no recognizable pattern—seemed like a justifiable modeling strategy to me.

The risk of first suspension estimated at the beginning of the school year was far more pronounced for 8th grade students than for 6th grade students. As I noted in my previous **Threats to Validity** section, this case study highlights the first suspension for students in the sample during *one school year*. It may have been that many students—particularly 7th and 8th grade students—who experienced their first suspension of the 2009-10 school year, had also been suspended the previous year. Indeed, many of these 7th and 8th grade students were likely suspended multiple times, or suspended late the previous year, perhaps continuing a pattern of misbehavior that had not been attenuated by prior administrative actions. I could not test such hypotheses in these data, but I hope to in my future thesis using longitudinal data on students over the entire three-year period of middle school.

Also, middle-school principals were working with different teams of teachers, in different grades. Theriot and Dupper (2010) highlighted the substantial difference in rates between 5th and 6th grade in their study, which focused exclusively in a school district where all of the students transitioned from elementary to middle school at that age. But in addition, their research also noted the high degree of subjectivity in "deciding" most disciplinary "cases." This subjectivity might contribute to the differences in the *profile* of risk, by grade level. For example, in this case study, many 6th grade teachers might have chosen to teach 6th grade rather than older middle-school students. And certainly these teachers also played some part in deciding which students were sent to the principal for discipline, based on teachers' own attitudes about student behavior. Individual teachers likely contributed their own take on how 6th graders should be disciplined. Consequently, differences in staff attitudes about younger students, then, seem to reinforce Lortie's (2009) assertion that students are only gradually socialized into school behaviors. So, in my study, the risk of first suspension for 6th grade students remained relatively constant until the end of the school year.

As my study confirms, large differences in patterns of suspension are demonstrated by differences among school principals and district suspension policies, rather than by individual student behaviors—findings that have been reported for decades (Edelman et al., 1975; Skiba et al., 1997; Wu et al., 1982). Furthermore, although including the main effect of *RACE* in Model C substantially attenuated the inter-school variability, it did not eliminate it. Teacher perceptions, administrative structure, and institutional procedures and biases (Wu et al., 1982) likely still played a substantial role in *whether* students were suspended from school.

The "full" fitted model, presented as Model C and Figure 2, highlighted what I believe was the most provocative insight—that the risk of first suspension for Black students was substantially higher than any other racial/ethnic category—a full order of magnitude higher in risk, compared to White students. My research adds to a large body of studies that has researched the gap in disciplinary outcomes between Black and White students. This racial disparity has been on the agenda of researchers since Edelman (1975) and her colleagues brought this issue to the fore 40 years ago.

Future Research

Skiba et al (2011) noted that "despite widespread beliefs to the contrary, there is no previous evidence that the overrepresentation of African American or Latino students in school disciplinary outcomes can be fully explained by individual or community economic disadvantage" (p. 103). In this analysis, I present only a description of the *problem*: that Black students—and Hispanic students to a much lesser extent—are overrepresented in school suspensions. With a hazard rate for Black middle-school students of more than *10 times* the rate for White students, I believe it seems unlikely to be full explained by poverty. Thus, in my future research, I propose to parse the mediation of *RACE* by measures of student poverty like free or reduced-price lunch status.

Differences in rates of school suspension by student gender are also a potential focus of my future research. Consistently, since the 1970s, studies that have examined gender differences in school punishment have reported a substantially higher rate of school suspension for male students, compared to female students. A survey conducted by the Children's Defense Fund reported that national suspension rates were 5.4% for boys and 3.4% for girls (Edelman et al., 1975). Raffaelle and Knoff (2003) investigated the rate of out-of-school suspensions for students attending a large Florida district in 1996-97, reporting that 32% of male middle-school students were suspended at least once that year, compared to 16% of females. And, in a recent analysis of discipline data for the 2009-10 school year from a school district in Wisconsin, Sullivan, Klingbeil, and Van Norman (2013) reported that the odds of out-of-school suspension for male students were more than twice the odds of out-of-school suspension for female students. Consequently, student gender will be an important topic when I refine and expand this analysis.

Furthermore, few other studies have explored the interaction between gender and race/ethnicity on school discipline. Ferguson (2000) investigated how

Black *male* students are particularly vulnerable to strict administration of discipline, like school suspensions, compared to White male students. And Morris (2005) observed, in a case study of a Texas school, that Black *female* students were likely to be disciplined at a rate similar to Black *male* students. Thus, in my further research, I plan to explore how the effect of *RACE* may itself be moderated by gender.

Conclusion

Using continuous-time survival analysis, I concluded that the risk of first suspension for middle-school students was higher at the beginning of the school year for 8th and 7th grade students. However, other than this, the risk of first suspension was relatively flat, holding steady at about five incidents of first suspensions per day, among the students in the 31 schools during the second half of the school year. Employing Cox-regression analysis, I also found that the risk differed by grade in school. At the beginning of the school year, the risk of first suspension for 8th grade students was more than double the risk for 6th grade students. But this difference in risk between grades diminished over time. Additionally, the risk of first suspension for Black students was substantially higher—more than 10 times the risk of first suspension for White students.

References

Anderson, J. D., (1988). *The education of Blacks in the south, 1860-1935*. Chapel Hill: University of North Carolina Press.

Arcia, E. (2007). A comparison of elementary/K-8 and middle schools' suspension rates. *Urban Education, 42*(5), 456-469. doi:10.1177/0042085907304879

Arum, R. (2003). *Judging school discipline: The crisis of moral authority*. Cambridge, Mass.: Harvard University Press.

Arum, R., & Preiss, D. (2009). Law and disorder in the classroom. *Education Next, 9*(4)

Beck, A. N., & Muschkin, C. G. (2012). The enduring impact of race: Understanding disparities in student disciplinary infractions and achievement. *Sociological Perspectives, 55*(4), 637-662. doi:10.1525/sop.2012.55.4.637

Butts, R. (1955). Our tradition of states' rights and education. *History of Education Journal, 6*(3), 211-228.

Cleves, M. A., Gould, W. W., Gutierrez, R. G. (2010). *An introduction to survival analysis using Stata* (3rd. ed.). College Station, Tex.: Stata Press.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the American Statistical Association, Series B, 34*, 187-220.

Dugan, L., Lafree, G., & Piquero, A. R. (2005). Testing a rational choice model of airline hijackings. *Criminology, 43*(4), 1031-1066.

Edelman, M. W., Beck, R., & Smith, P. V. (1975). *School suspensions--are they helping children?* Cambridge, Mass.: Children's Defense Fund.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, *72*(359), 557-565.

Ferguson, A. A. (2000). *Bad boys: Public schools in the making of Black masculinity*. Ann Arbor, MI: The University of Michigan Press.

Goldin, C. D., & Katz, L. F. (2008). *The race between education and technology*. Cambridge, Mass.: Belknap Press of Harvard University Press.

Goss v. Lopez, 419 U.S. 565, Justia US Supreme Court Center (U.S. Supreme Court 1975).

Gregory, A., Cornell, D., & Fan, X. (2011). The relationship of school structure and support to suspension rates for black and white high school students. *American Educational Research Journal, 48*(4), 904-934.

Heikinheimo, T., Broman, J., Haapaniemi, E., Kaste, M., Tatlisumak, T., & Putaala, J. (2013). Preceding and poststroke infections in young adults with first-ever ischemic stroke effect on short-term and long-term outcomes. *Stroke, 44*(12), 3331-3337. doi:10.1161/STROKEAHA.113.002108

Hinojosa, M. S. (2008). Black-white differences in school suspension: Effect of student beliefs about teachers. *Sociological Spectrum, 28*(2), 175-193. doi:10.1080/02732170701796429

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*(282), 457-481.

KewalRamani, A., Gilbertson, L., Fox, M. A., Provasnik, S. (2007). *Status and trends in the education of racial and ethnic minorities. NCES 2007-039.* (No. NCES 2007-039). National Center for Education Statistics.

Kinsler, J. (2011). Understanding the black-white school discipline gap. *Economics of Education Review, 30*(6), 1370-1383. doi:10.1016/j.econedurev.2011.07.004

Lomawaima, K. T. (1999). *The unnatural history of American Indian education.*

Lortie, D. C. (2009). *School principal: Managing in public*. Chicago: University of Chicago Press.

Losen, D., & Skiba, R. J. (2010). *Suspended education: Urban middle schools in crisis.* Southern Poverty Law Center.

Martin, J. (2014, April 20 2014). As G.O.P wedge, the common core cuts both ways. *New York Times*, pp. A1.

McCarthy, J. D., & Hoge, D. R. (1987). The social construction of school punishment: Racial disadvantage out of universalistic process. *Social Forces, 65*(4), 1101-1101-1120.

Morris, E. W. (2005). "Tuck in that shirt!" Race, class, gender, and discipline in an urban school. *Sociological Perspectives, 48*(1), 25-48.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards*. Washington, DC:

Nichols, J. D. (2004). An exploration of discipline and suspension data. *The Journal of Negro Education, 73*(4), 408-408-423.

No child left behind (NCLB) act of 2001, Pub. L. No. 107-110, § 115, Stat.1425, (2002).

Office of Management and Budget. (1997). Revisions to the standards for the classification of federal data on race and ethnicity. Retrieved February 21, 2014, Retrieved from www.whitehouse.gov/omb/fedreg/1997standards.html

Omi, M., & Winant, H. (1994). *Racial formation in the United States: From the 1960s to the 1990s* (2nd ed.). New York: Routledge.

Ornstein, A. (1982). Student disruptions and student rights: An overview. *The Urban Review, 14*(2), 83-91.

Payne, A. A., & Welch, K. (2010). Modeling the effects of racial threat on punitive and restorative school discipline practices. *Criminology, 48*(4), 1019-1062. doi:http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1111/j.1745-9125.2010.00211.x

Petras, H., Masyn, K. E., Buckley, J. A., Ialongo, N. S., & Kellam, S. (2011). Who is most at risk for school removal? A multilevel discrete-time survival analysis of individual- and context-level influences. *Journal of Educational Psychology, 103*(1), 223-237. doi:10.1037/a0021545

Raffaele Mendez, L. M., & Knoff, H. M. (2003). Who gets suspended from school and why: A demographic analysis of schools and disciplinary infractions in a large school district. *Education and Treatment of Children, 26*(1), 30-30-51.

Race to the top act of 2011, H.R. 1532--112th Congress (2011).

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika, 69*(1), 239-241.

Schwerdt, G., & West, M. R. (2013). The impact of alternative grade configurations on student outcomes through middle and high school. *Journal of Public Economics, 97*(0), 308-326. doi:http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1016/j.jpubeco.2012.10.002

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford ;New York: Oxford University Press.

Skiba, R. J., Horner, R. H., Choong-Geun Chung, Rausch, M. K., May, S. L., & Tobin, T. (2011). Race is not neutral: A national investigation of african american and latino disproportionality in school discipline. *School Psychology Review, 40*(1), 85-107.

Skiba, R. J., Peterson, R. L., & Williams, T. (1997). Office referrals and suspension: Disciplinary intervention. *Education & Treatment of Children (ETC), 20*(3), 295.

Sullivan, A. L., Klingbeil, D. A., & Van Norman, E. R. (2013). Beyond behavior: Multilevel analysis of the influence of sociodemographics and school characteristics on students' risk of suspension. *School Psychology Review, 42*(1), 99-114.

Tamura, E. H. (2001). Historiographical essay. *History of Education Quarterly, 41*(1), 58.

Theriot, M. T., & Dupper, D. R. (2010). Student discipline problems and the transition from elementary to middle school. *Education and Urban Society, 42*(2), 205-222. doi:10.1177/0013124509349583

Townsend, B. L. (2000). The disproportionate discipline of African American learners: Reducing school suspensions and expulsions. *Exceptional Children, 66*(3), 381-381-391.

U.S. Census. (2002). *Measuring America: The decennial censuses from 1790 to 2000.* (No. POL/02-MA(RV)).U.S. Department of Commerce.

U.S. Department of Education. (2008). New race and ethnicity guidance for the collection of federal education data. Retrieved January 18, 2014 Retrieved from http://www2.ed.gov/policy/rschstat/guid/raceethnicity/index.html

U.S. Department of Education. (2012). *Digest of educational statistics*. Washington DC: U.S. Dept. of Health, Education, and Welfare, Office of Education.

United States Department of Education Office for Civil Rights. (2013). *Civil rights data collection* www2.ed.gov/about/offices/list/ocr/docs/crdc-2011-12-p1-p2.doc;.

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography, 16*(3), 439-454.

Wallace, J.,John M., Goodkind, S., Wallace, C. M., & Bachman, J. G. (2008). Racial, ethnic, and gender differences in school discipline among U.S. high school students: 1991-2005. *Negro Educational Review, 59*(1), 47-62.

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics,* (September), 293-297.

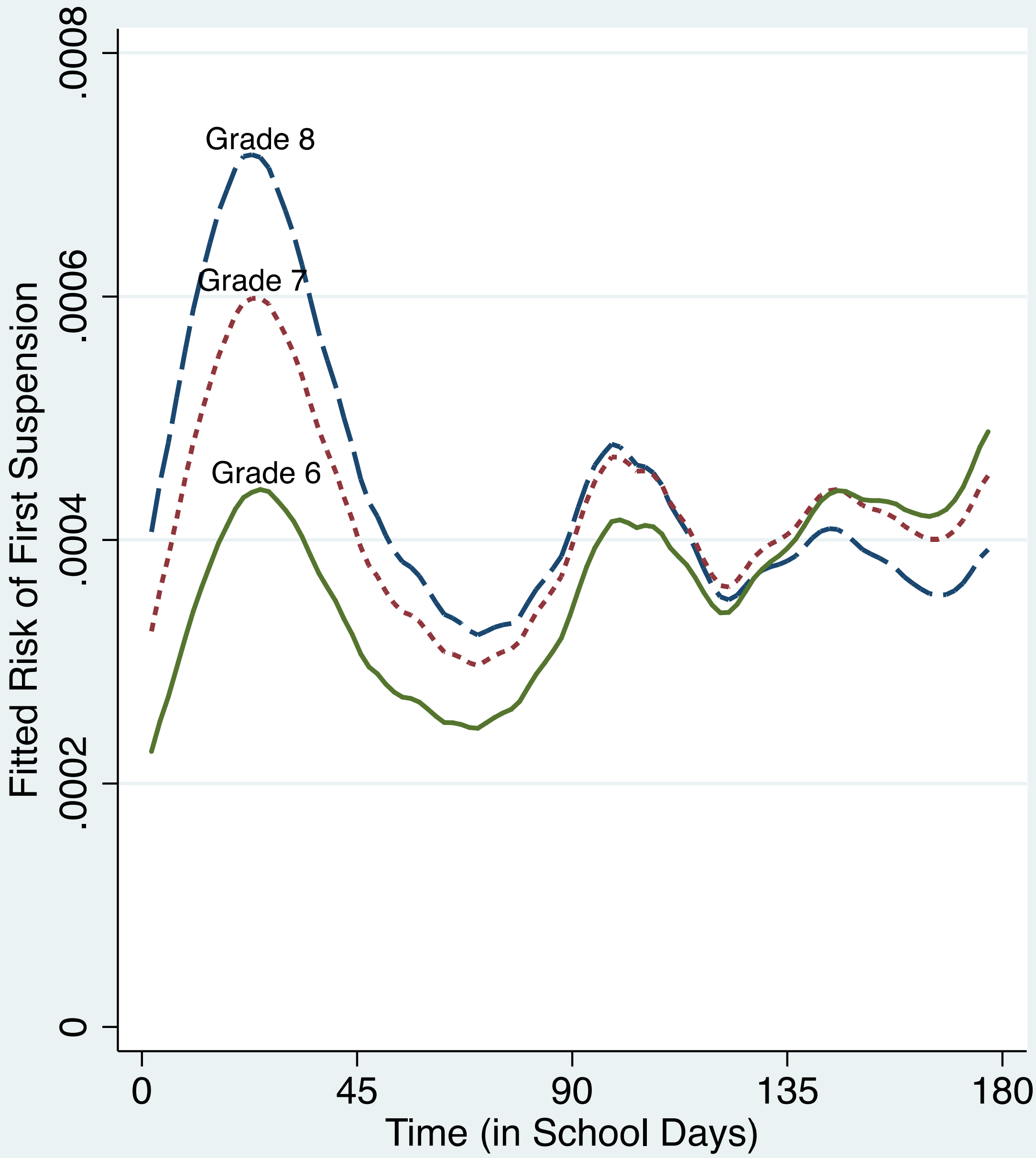
Wu, S., Pink, W., Crain, R., & Moles, O. (1982). Student suspension: A critical reappraisal. *The Urban Review, 14*(4), 245-303. doi:10.1007/BF02171974

Zheng, H., & Thomas, P. A. (2013). Marital status, self-rated health, and mortality: Overestimation of health or diminishing protection of marriage? *Journal of Health and Social Behavior, 54*(1), 128-143. doi:10.1177/0022146512470564

Table 1. *Fitted Cox-regression models, with estimated slope parameters, standard errors, estimated hazard ratios and approximate p-values, describing how the occurrence of first suspension depends on time and race/ethnicity of the student ($n_{students}$=13,256; $n_{suspensions}$=897; $n_{schools}$=31).*

	Model A			Model B			Model C		
	Parameter	*s.e.*	Hazard Ratio	Parameter	*s.e*	Hazard Ratio	Parameter	*s.e*	Hazard Ratio
1. GRADE6 *(reference)*	-	-	1	-	-	1	-	-	1
2. GRADE7	0.1475	*0.0854*	1.1590~	0.3668	*0.1662*	1.4431*	0.5323	*0.1664*	1.7030**
3 .GRADE8	0.2054	*0.0850*	1.2281*	0.5952	*0.1625*	1.8134***	0.7433	*0.1626*	2.1028***
4. GRADE7xTIME				-0.0025	*0.0016*	0.9975	-0.0022	*0.0016*	0.9978
5. GRADE8xTIME				-0.0046	*0.0016*	0.9954**	-0.0039	*0.0016*	0.9961*
6. ASIAN							-0.8086	*0.2758*	0.4455**
7. BLACK							2.3106	*0.0821*	10.0809***
8. HISPANIC							0.4579	*0.1447*	1.5808**
9. NATIVE AMERICAN							0.5022	*0.5040*	1.6524
10. WHITE *(reference)*							-	-	1
11. θ	1.1283	*0.2912*		1.1280	*0.2912*		0.6360	*0.1870*	
12. LR test of $\theta = 0$	646.00	*p<.001*		645.99	*p<.001*		252.71	*p<.001*	
13. Log-likelihood	-8152.322			-8148.226			-7669.877		
14. LR Test of Model B vs. Model A				8.192	*p=0.02*				
15. LR Test of Model C vs. Model B							956.698	*p<.001*	
16. χ^2 test of proportional hazards	8.41 *(2)*	*p=0.02*		4.02 *(4)*	*p=0.40*		11.42 *(8)*	*p=0.18*	

17. ~ *p<0.10; * p<0.05 ** p<0.01 *** p<0.001*



Number First Suspended Number (# at Risk)

	First Quarter	Second Quarter	Third Quarter	Fourth Quarter
Grade 8	114 (4,572)	79 (4,476)	72 (4,376)	64 (4,288)
Grade 7	99 (4,678)	63 (4,538)	82 (4,498)	73 (4,399)
Grade 6	73 (3,924)	41 (3,865)	64 (3,805)	73 (3,720)

Figure 1. Estimated risk of first suspension versus time (school days), by student grade, from fitted Model B *($n_{students}$=13,256, $n_{schools}$=31, $n_{first\ suspensions}$=897).*

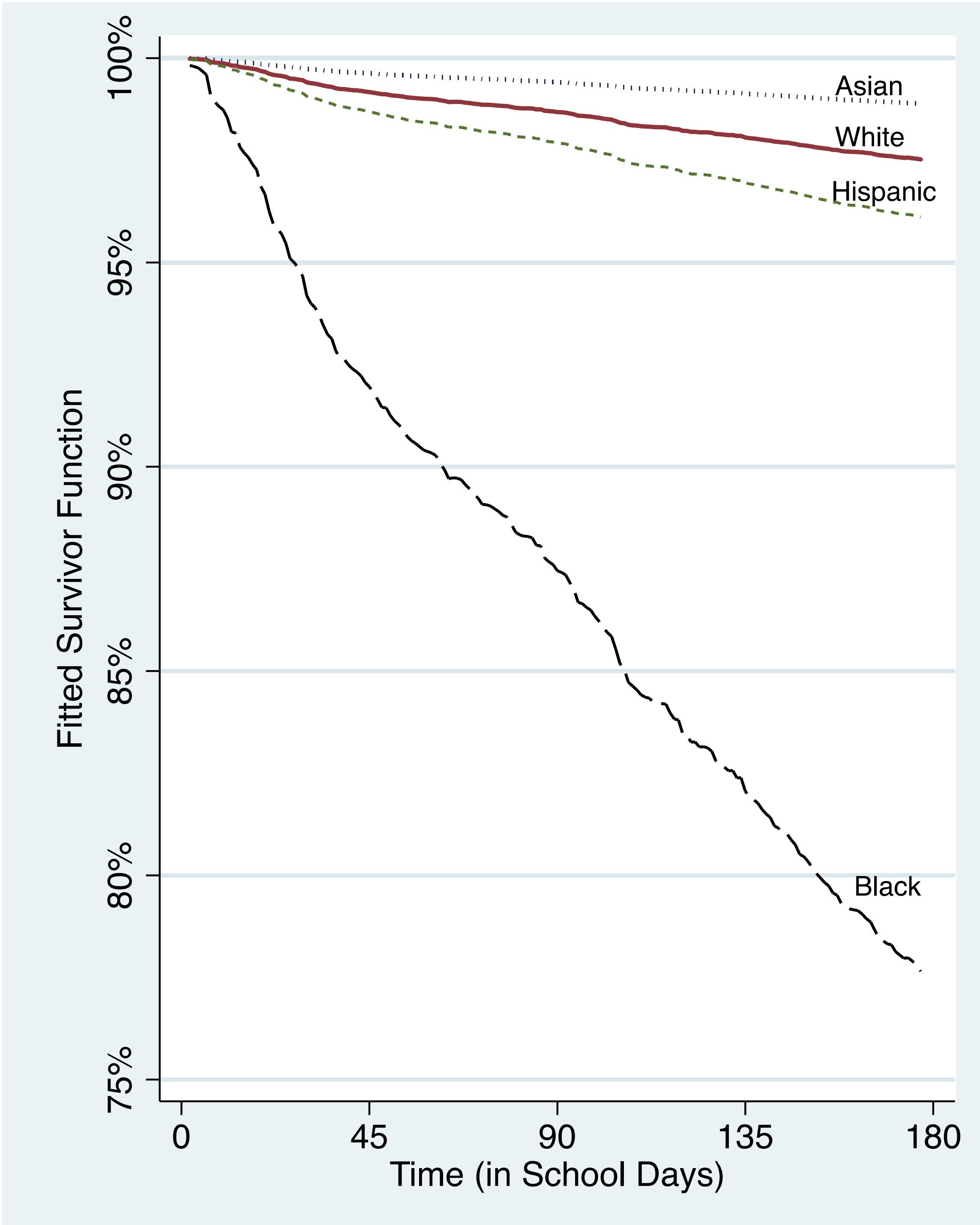


Figure 2. Estimated risk of first suspension versus time (school days) for four prototypical middle-school students, at an average (θ = 1) sample school, by student race/ethnicity, from fitted Model C *($n_{students}$=13,256, $n_{schools}$=31, $n_{first\ suspensions}$=897).*