NORTHWESTERN UNIVERSITY

Stress, Sleep, and Cognitive Functioning:

Pathways from Stressors to Academic Outcomes

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Human Development and Social Policy

By

Jennifer Ann Heissel

EVANSTON, ILLINOIS

June 2017

Abstract

Achievement gaps between students from different SES and racial/ethnic backgrounds stubbornly persist despite decades of policy interventions. My dissertation focuses on several unexpected sources of stress and disruption that may be related to the achievement gap: sunlight exposure, a sister's teen childbearing, and nearby violent crime. I demonstrate that these may have important implications for academic performance, but each has received little attention in the academic literature.

In Study 1, I examine the change in academic performance associated with changes to the amount of sunlight before school. Though previous work has examined school start times, this is the first paper to my knowledge that explicitly accounts for the physiological influence of sunlight on sleep timing and its relationship to school start times over the pubertal transition. Sleep patterns are determined in part by sunrise times, which vary across time zones. Because school start times do not fully reflect this difference, I instrument for the hours of sunlight before school with the time zone boundary in Florida. I find that moving start times one hour later relative to sunrise increases test scores by 0.07 and 0.05 standard deviations for adolescents in math and reading, respectively. In math, the effect is larger for older children and co-varies with entry into an important pubertal stage. School districts can improve performance while maintaining the current distribution of start times by moving classes earlier for younger children and later for older children.

In Study 2, I examine another potentially disruptive event for students: the appearance of a baby belonging to a given student's teenage sister in the home. Despite the abundance of research on mothers' own outcomes, almost no research has examined the outcomes for the

siblings of teen mothers. Using annual longitudinal data, this study shows that both teen mothers and their siblings are on a downward trajectory well before the teen pregnancy begins. However, when compared to students on a similar trajectory in families without teenage childbearing, the siblings of teen mothers have worse test scores, higher high school dropout rates, lower college attendance, and lower college graduation following the birth. The change in test score outcomes only occurs after the baby is born, indicating that it is the appearance of the newborn that affects performance, rather than some unobserved occurrence that leads to both teen pregnancy and poor outcomes. Sisters of teen mothers have larger academic effects than brothers; brothers have larger juvenile justice effects. I demonstrate similar patterns for the teen mothers, though the divergence in scores begins in the year of the pregnancy, not the year of birth.

Finally, Study 3 examines the effect of unexpected nearby violent crime on students' sleep and cortisol patterns. The data combine objectively measured sleep and thrice-daily salivary cortisol collected from a four-day diary study in a large Midwestern city with location data on all violent crimes recorded during the same time period. The primary empirical strategy uses a within-person design to measure the change in sleep and cortisol from the person's typical pattern on the night/day immediately following a local violent crime. On the night following a violent crime, adolescents have later bedtimes. Adolescents also have disrupted cortisol patterns the following morning. Supplementary analyses using varying distances of the crime to the child's home address confirm more proximate crimes correspond to later bedtimes.

Acknowledgements

I am grateful to the friends, family, and colleagues who have supported me in graduate school. First, I cannot imagine a better dissertation committee: David Figlio, Emma Adam, and Jonathan Guryan. They not only provided useful feedback on my research, but their invaluable career and life advice has made me a better scholar and person. David's generosity and generativity set me up for the success that I have. I will deeply miss popping into Emma's office and disturbing the hallway near her lab.

My Northwestern colleagues have helped me through this process in more ways than they probably know. The HDSP quantitative group gave advice at the earliest stages of my research, and their words of wisdom changed the trajectory of some of my best papers. I will miss my cohort – Eric Brown, CC DuBois, Jiffy Lansing, Mollie McQuillan, and Emily Ross – as well as the residents of 626 Library Place and the first floor Annenberg Hallway. Olivia Healy's top-notch babysitting skills gave me the time to get a job market paper together. Sam Norris was a fantastic co-author on the journal version of the school start times paper (Heissel & Norris, 2016), and I'm still a little sad we had to let the Tennessee analysis go. I also thank Patrick Sharkey, Gerard Torrats-Espinosa, and Kathryn Grant, my co-authors, along with Emma, for the journal version of the violent crime paper (Heissel, Sharkey, Torrats-Espinosa, Grant, & Adam, 2017). I also thank seminar participants at the American Economic Association, Association for Public Policy Analysis and Management, Association for Education Finance and Policy, Southern Economic Association, Association for Psychological Sciences, Naval Postgraduate School, University of Alabama, University of Chicago (Crime Lab New York and Poverty Lab), University of Illinois, and University of Notre Dame; the following papers are better because of their comments over the years.

Abigail Wozniak at the University of Notre Dame first told me I should get a PhD before I really knew what that meant. I ignored her, and got a Master's in Public Policy instead – and realized I wanted a PhD while I was there. I should always listen to my advisors, and I thank Abbie for her career advice. While I was at Duke, Helen "Sunny" Ladd told me I should work with David Figlio at Northwestern. She was correct, again confirming that my advisors are always right. Sunny taught me to write compelling introductions and to use my research to improve policy.

I have benefited from generous financial support over the course of my graduate career. I am grateful to Duke University, the Joel L. Fleishman Endowment, Northwestern University, the Sybil N. Heide Fellowship Fund, the Institute for Policy Research, and the SESP Global Initiative Award for their support.

Abigail Durgan provided excellent research assistance on school start times and many other projects throughout her four years at Northwestern, and Komal Khoja was an excellent editor. I will miss the many undergraduate research assistants from the Adam Lab.

Most of all, I thank my family. There's a lot of chance in life, and I am very lucky. My parents – Linda and Norm Heissel – set me off on this path: giving me a love of education and supporting my crazy adventures. My sister, Emily Miller, is a teacher, who grounds my work. My brother, Ben Heissel, is the hardest worker I know. Brian Tracy, you are the love of my life and the only reason I was able to get through grad school. Thank you for your support, humor, and excellent cooking skills. Teddy, thank you for reminding me what's important.

Dedication

To my parents, who first emphasized the importance of education, and to Brian, my favorite editor.

Contents

Table of Tables

Study 1 Tables

Study 1 Appendix Tables

Study 2 Tables

Study 2 Appendix Tables

Study 3 Tables

Study 3 Appendix Tables

Table of Figures

Study 1 Figures

Study 1 Appendix Figures

Study 2 Figures

Study 3 Figures

Study 3 Appendix Figures

Introduction

The achievement gaps between students from different SES and racial/ethnic backgrounds stubbornly persist despite decades of policy interventions suggested by politicians, policymakers, and advocates (Barton & Coley, 2009, 2010; Reardon, 2011; Reardon & Robinson, 2007). Researchers have extensively studied many factors that contribute to this gap.[1] My dissertation instead focuses on several unexpected sources of stress and disruption that have received little attention in the academic literature: sunlight exposure, a sister's teen childbearing, and nearby violent crime.

Two of these studies are directly related to sleep and stress. Heissel, Levy, and Adam (2017) lay out a theoretical model for why stress and sleep may matter for academic achievement gaps. As a summary, stress exposure can affect the functioning of two stress-sensitive systems that are relevant to learning and test performance: the hypothalamic-pituitary-adrenal (HPA) axis and the sleep cycle. Alterations in the HPA axis and in sleep quality and quantity can affect students' learning and test-taking experience (Buckhalt, 2011; Joëls, Pu, Wiegert, Oitzl, & Krugers, 2006; Lupien et al., 2002; Maldonado et al., 2008; Schwabe & Wolf, 2010; Sharkey, 2010; Vedhara, Hyde, Gilchrist, Tytherleigh, & Plummer, 2000; Walker & Stickgold, 2006). To the extent that stress exposure is not randomly distributed across the population, these processes have policy implications for the interpretation of achievement gaps measured by standardized

[1] Children from low-SES and minority families face a variety of barriers to success. Teachers in high-poverty schools and those with many minority students have much lower qualifications than teachers in more affluent and White-majority schools (Ingersoll, 2001; Jackson, 2009; Jacob, 2007), and these differences systematically affect student achievement (Clotfelter, Ladd, & Vigdor, 2007, 2010). Race and SES also predict poor health outcomes (Adler & Rehkopf, 2008; Adler & Stewart, 2010; Matthews, Gallo, & Taylor, 2010; Myers, 2008; Orsi, Margellos-Anast, & Whitman, 2010), and risky family environments may exacerbate mental and physical health problems for these students (Flinn & England, 1995; Repetti, Taylor, & Seeman, 2002). In turn, poor health reduces academic achievement (Fletcher & Lehrer, 2009). Moreover, low SES children may disengage from school if they see it as unimportant for their future (Destin & Oyserman, 2009).

tests such as end-of-course exams, course placement exams, high school competency tests, and college entrance exams.

In Study 1, I examine the change in academic performance associated with changes to the amount of sunlight before school. Though previous work has examined school start times, this is the first paper to my knowledge that explicitly accounts for the physiological influence of sunlight on sleep timing and its relationship to school start times over the pubertal transition. Sleep patterns are determined in part by sunrise times, which vary across time zones. Because school start times do not fully reflect this difference, I instrument for the hours of sunlight before school with the time zone boundary in Florida. I find that moving start times one hour later relative to sunrise increases test scores by 0.07 and 0.05 standard deviations for adolescents in math and reading, respectively. In math, the effect is larger for older children and co-varies with entry into an important pubertal stage. School districts can improve performance while maintaining the current distribution of start times by moving classes earlier for younger children and later for older children.

In Study 2, I examine another potentially disruptive event for students: the appearance of a baby belonging to a given student's teenage sister in the home. Prior research has reached conflicting conclusions about whether teenage pregnancy causes poor outcomes for the teen mother herself or whether it is a symptom of prior trends. Teenage parenthood is popularly understood to be a negative outcome for the mother, including reduced education attainment and worse long-term economic prospects (Kane, Morgan, Harris, & Guilkey, 2013; A. R. Miller, 2009). However, many such studies do not account for the negative selection into pregnancy and, among those who get pregnant, the positive selection into abortion (Ashcraft, Fernández-Val, &

Lang, 2013; Fletcher, 2011; Geronimus & Korenman, 1992; Hotz, McElroy, & Sanders, 2005; Hotz, Mullin, & Sanders, 1997). Recent work has argued that miscarriage is correlated with community-level factors, and after accounting for this correlation teenage pregnancy in the U.S. does indeed reduce high school graduation rates and annual income (Fletcher & Wolfe, 2009).

Despite the abundance of research on mothers' own outcomes, almost no research has examined the outcomes for the *siblings* of teen mothers. Using annual longitudinal data, this study shows that both teen mothers and their siblings are on a downward trajectory well before pregnancy. I show that methods that compare students based on first-observed characteristics, rather than trajectories, systematically overestimate the effects and spillover of teen pregnancy because of these pre-pregnancy trends.

When compared to students on a similar trajectory in families without teenage childbearing, the siblings of teen mothers have worse test scores, higher high school dropout rates, lower college attendance, and lower college graduation. The change in test score outcomes only occurs after the baby is born, indicating that it is the appearance of the newborn that affects performance, rather than some unobserved occurrence that leads to both teen pregnancy and poor outcomes. Sisters of teen mothers have larger academic effects than brothers; brothers have larger juvenile justice effects. A similar analysis provides a maximum bound for the effects of teen pregnancy on mothers' outcomes, finding lower test scores, more grade repetition, higher high school dropout, less college-going, and less college graduation, relative to females who had been on a similar downward trajectory from families without teenage childbearing. For mothers, the divergence in scores begins in the year of the pregnancy, not the year of birth.

Finally, Study 3 examines the effect of unexpected nearby violent crime on students' sleep and cortisol patterns. The data combine objectively measured sleep and thrice-daily salivary cortisol collected from a four-day diary study in a large Midwestern city with location data on all violent crimes recorded during the same time period. The primary empirical strategy uses a within-person design to measure the change in sleep and cortisol from the person's typical pattern on the night/day immediately following a local violent crime. On the night following a violent crime, adolescents have later bedtimes. Adolescents also have disrupted cortisol patterns the following morning. Supplementary analyses using varying distances of the crime to the child's home address confirm more proximate crimes correspond to later bedtimes.

Each of these studies is related to the achievement gap. Black adolescents and lower-SES students have worse sleep quality and quantity than white adolescents and higher-SES adolescents, respectively (Adam, Snell, & Pendry, 2007; Hale et al., 2013; Heissel, Levy, et al., 2017). Study 1 provides evidence that sleep affects academic performance. Similarly, black and lower-SES families have higher rates of teen pregnancy than white and higher-SES families, respectively (Penman-Aguilar, Carter, Snead, & Kourtis, 2013; U.S. Department of Health and Human Services, 2015), and Study 2 provides evidence that a sister's teen motherhood affects her siblings. Finally, Study 3 indicates that acute violent crime affects both sleep and cortisol. Crime rates are higher in neighborhoods with higher proportions of black and lower-SES families than white and higher-SES families, respectively (U.S. Department of Housing and Urban Development, 2016). Both sleep and cortisol disruptions negatively test performance, so these sort of acute disruptions are more likely to affect the performance of black and lower-SES adolescents. Thus, all three studies provide potential pathways to black-white and SES achievement gaps.

Study 1. Good Day, Sunshine: The Effect of School Start Times on Academic Performance from Childhood through Puberty

American teenagers are chronically sleep-deprived (Eaton et al., 2010). As children enter puberty, physiological changes delay the onset of sleep and make it more difficult to wake up early in the morning. By the end of middle school there is a large disconnect between physiological sleep patterns and school schedules: Hansen, Janssen, Schiff, Zee, & Dubocovich (2005) find that students lose as much as 120 minutes of sleep per night after they start school in September, compared to the summer months when they can better control their own sleep schedules.

Sleep matters for learning and cognition. Important memory formation and consolidation processes occur overnight, as the brain replays patterns of brain activity exhibited during learning (Fogel & Smith, 2011; Sadeh, Gruber, & Raviv, 2003). Restricting sleep also reduces alertness and attention levels (Lufi, Tzischinsky, & Hadar, 2011; Sadeh et al., 2003), which likely affects students' ability to learn or take tests the next day. In light of these findings, the American Academy of Pediatrics recommends that adolescents wake up no earlier than 8:00 a.m. (2014). As of 2011, the median *start time* for American high schools was 8:00 a.m., suggesting that current policy may have cognitive costs for students.

Relatively little research has directly examined the effect of K-12 start times on academic performance. I study this question with a novel identification strategy that takes advantage of the biological effect of light on sleep patterns. Sleep timing is partially regulated by sunlight exposure; holding hours of darkness constant, more sunlight in the morning (and less at night) naturally moves bedtimes earlier and increases alertness in the morning (Crowley, Acebo, &

Carskadon, 2007). Sunlight before school – as opposed to clock start times – is therefore the correct measure of policy when comparing between schools.[2] Students exposed to more sunlight should then improve their academic performance, an effect that should be stronger for pubertal children because of their delayed sleep schedules (Carskadon, Acebo, Richardson, Tate, & Seifer, 1997). The empirical strategy in this paper leverages the discontinuous change in sunrise times at a time zone border, combined with the fact that school start times do not fully adjust for this difference. This rich administrative dataset of all public school students in Florida between 1999 and 2013 tracks children as they move across the Central-Eastern time zone boundary. Treating time zone as an instrument for sunlight before school, I identify the effect of start time relative to sunrise on academic performance conditional on student fixed effects and school characteristics.

The data include children moving across the time zone boundary at all ages between eight and fifteen, which allows estimation of the age-specific effect of school start times over a range of developmental stages. An additional hour of sunlight before school has almost no effect on math scores for pre-pubescent children, but a large and abrupt effect appears for girls at age 11 and boys at age 13. This pattern corresponds exactly to the gender-specific median age of an important pubertal transition (Campbell, Grimm, Bie, & Feinberg, 2012), which I take as evidence that the causal pathway is linked to the physiological changes that occur during puberty. Specifically, a one-hour delay in relative start times increases standardized math scores by 0.081 standard deviations for adolescents, but only 0.009 SDs for pre-pubertal children. In reading, an extra hour of sunlight before school increases scores by 0.057 SDs for adolescents and 0.061 SDs younger children. The difference between groups is not statistically significant in

[2] For any given school, clock start time is collinear with sunlight before school.

reading, though the adolescent estimate is more precise and can be tested as different from zero. As children move over the time zone boundary, the change in scores occurs within a year of the change in sunlight exposure and persists over time.

Later relative start times do not increase learning time for adolescents, as measured by absences. Absences are reduced by 0.869 percentage points for younger children. Differences in how absence is measured across school types (elementary, middle, and high schools) may be part of the reason behind the differences in outcomes I find here. I do not observe tardiness that does not result in an absence and therefore cannot rule it out as a causal channel, but the results are consistent with improved alertness and learning capacity as a result of later start times for adolescents.

The present paper builds on the current literature in two other ways. First, it provides evidence on whether improved achievement in high-morning-sunlight areas is a result of better learning throughout the year, or merely improvements in testing performance. Using variation in test timing over the sample years, I show that testing effects are unlikely to account for the math results. They may make up a portion of the gains from later start times in reading.

Second, it addresses a potentially important educational policy. Although moving start times later for all students would increase academic performance at a relatively low monetary cost (Jacob & Rockoff, 2011), interference with transportation and parental work schedules is a major concern for many districts. An alternative policy is to keep the same distribution of start times, but to adjust the opening order for schools in a way that is consistent with the physiological evidence: elementary schools, middle schools, and finally high schools. I show that most districts in the Florida panhandle do not follow this optimal pattern, but that the policy would increase math and reading scores by 0.06 and 0.04 SDs for high school students, with little negative effect for younger students. Although there may be other costs – in particular, young children might have to wait for the school bus in the dark – the present paper is the first to quantify the academic benefits of this policy.

Background

Previous research

There have been several recent studies investigating the effect of daily start times on academic achievement, though none have examined the role that pubertal changes play in the effects. Wahlstrom, Wrobel, & Kubow (1998) find that delaying school start times in Minneapolis public schools from 7:15 to 8:40 improved student sleep by 39 minutes and significantly decreased tardiness rates. Their measure of academic performance was teacher-assigned grades, where they found a positive but statistically significant effect.[3] A later paper by Hinrichs (2011) exploiting the same policy change finds no effect on ACT scores. Another approach is from Edwards (2012), who uses changes to busing schedules as a source of potentially exogenous variation in start times. He finds evidence that delayed start times increase achievement for middle school students. The effect seems to be smaller for elementary students, but he notes that this may be a result of start times being much later for younger children in his sample. The results are not available by gender, which makes inference on the importance of puberty difficult. Finally, Carrell, Maghakian, & West (2011) study freshmen cadets at the United States Air Force Academy who were randomly assigned different school schedules, and who belonged to cohorts with different first-period start times. Using this random variation, they

[3] Teacher-assigned grades may understate the effect of school-level interventions if teachers curve assigned grades within a given class and year.

find that having a start time of 7:00 a.m. (versus no class in first period) decreases achievement by about 0.15 SDs in that class, and by about 0.10 SDs in subsequent classes.

Sunlight, sleep, and puberty

The role of sunlight in determining sleep schedules is well known. Sleep patterns are partially controlled by the circadian rhythm, which synchronizes to a 24-hour cycle using the daily variation in light and darkness (Crowley et al., 2007). In the morning, light on the outside of the eyelids suppresses production of the hormone melatonin and stimulates brain processes to increase alertness; darkness at night increases melatonin levels and feelings of tiredness (Arendt, 2000).

One of the well-documented changes during adolescence is to the timing of sleep. As children move through puberty, nocturnal melatonin secretion is delayed several hours relative to adults and younger children (Carskadon, Acebo, & Jenni, 2004; Carskadon et al., 1997). The result is that adolescent sleep patterns become more owl-like, with later bedtimes and wake times, even holding the level of darkness fixed (Carskadon et al., 2004; Carskadon, Vieira, & Acebo, 1993; Crowley et al., 2007). Schools in the United States tend to begin early to accommodate after-school activities and parental work schedules, preventing adolescents from waking at their preferred later times and leading to an increasing disconnect between weekday and weekend sleep schedules during the school year (Jenni & Carskadon, 2012; Laberge et al., 2001). The result is low wakefulness and attention levels on school days (Lufi et al., 2011). More directly, sleep levels have large effects on cognitive performance (Sadeh et al., 2003; Walker & Stickgold, 2006).

Although boys and girls undergo similar sleep-related changes during adolescence, the age profile of puberty varies significantly by gender. Marshall and Tanner (1970) show that pubic hair development begins 1.5 years earlier for girls than for boys; there is a similar gap for attainment of other developmental thresholds. This variation in age at entry into successive pubertal stages generates an important testable prediction: if physiological changes are driving the increasing importance of school start times during high school, then the size of the start time effect will co-vary with the gender-specific entry into puberty. In contrast, other changes that might make start times more relevant to achievement – e.g., the transition to a block schedule, middle-school social pressures, or changes to after-school activities – likely affect both genders at the same age.

Identification strategy

The goal of this paper is to estimate the causal effect of school start times on academic achievement and behavioral outcomes. One approach would be to regress outcomes on start times, but because start times are chosen by the policy-maker, this approach would generate upwards-biased coefficients if better-managed schools tend to also start later in the day.[4]

Instead, the identification strategy exploits the relationship between sunlight and sleep, along with variation in sunrise time between locations. The intuition is that sleep patterns are linked partially to sunrise and sunset times, rather than clock time. This means that in terms of student sleep and alertness, the policy-relevant measure of school start time is start time relative to sunrise. For a given school, this is an unnecessary distinction: the choice of when to start classes according to the clock is equivalent to deciding when to start classes relative to sunrise.

[4] Better schools may also start earlier; for example, they may start earlier to accommodate after-school activities. This fundamental uncertainty about the direction of the bias from OLS underlines the importance of good instruments in this context.

Between schools in different locations, however, a given clock start time corresponds to different relative start times. This contrast is particularly stark at a time zone boundary. Suppose that there are two schools close together but on opposite sides of the boundary, where the sun rises at 6:00 a.m. in Central Time (CT) and 7:00 a.m. in Eastern Time (ET). If both schools begin classes at 8:00 a.m. local time, students attending the school in CT will have one more hour of sunlight before the morning bell.[5] To translate this insight into credible estimates, I track academic achievement as students move between schools on different sides of the time zone boundary. As students move from CT to ET, they are exposed to less sunlight before school, which I expect will decrease academic achievement. Conversely, a student moving from ET to CT gains sunlight before school and should see their test scores increase.

Formally, I use the time zone as an instrument for the amount of sunlight before school, which I refer to as the relative start time. I then regress academic and behavioral outcomes on instrumented relative start time to estimate the causal effect of relative start times.

The exclusion restriction in this setting is that time zone is uncorrelated with other school and student characteristics that might also affect achievement. This assumption might not be realistic in certain contexts. If, for example, I regressed achievement on instrumented time zone for the entire state of Florida, the identifying assumption would be that the only difference between schools in CT and ET relevant to student achievement is variation in relative sunrise times. Even conditional on a robust set of controls, this assumption is unlikely to hold. Instead, I include a set of student fixed effects and identify the coefficients of interest using only within-

[5] Children in CT will also have one less hour of sunlight after school. It is possible that this has an effect on academic outcomes, for example if less sunlight after school decreased sports participation and led to more homework time. As a policy matter, moving school start times later will always increase sunlight before school at the expense of sunlight after school; because I am interested in the effect of school start times as a policy, I consider this a feature of the approach.

student variation. This means that variation in the instrument comes only from students who move between time zones.

I relate outcomes to start times using the following functional form:

$$y_{it} = \delta_1 hours_{it} + \delta_2 hours_{it} \times \mathbb{1}[\text{puberty}] + X_{it}\beta + \gamma_i + \varepsilon_{it} \tag{1-1}$$

where y_{it} is the outcome of interest, $hours_{it}$ is the number of hours between sunrise and school start, X_{it} is a vector of controls, and γ_i is an individual fixed effect. The first stage instruments for relative start time with an indicator for time zone $timezone_{it}$:

$$hours_{it} = \alpha_{11} timezone_{it} + \alpha_{12} timezone_{it} \times \mathbb{1}[\text{puberty}] + X_{it}\theta_1 + \eta_{1i} + u_{1it} \tag{1-2}$$

$$hours_{it}\, \mathbb{1}[\text{puberty}] = \alpha_{21} timezone_{it} + \alpha_{22} timezone_{it} \times \mathbb{1}[\text{puberty}] + X_{it}\theta_2 + \eta_{1i} + u_{2it} \tag{1-3}$$

where η_i are individual fixed effects. The vector X_{it} typically includes longitude, which directly affects sunrise times, as well as school-level demographic controls to proxy for school quality.

Crucially, the estimate allows the effect of start time to vary by pubertal status. Based on the biological evidence, I expect that students' natural sleep patterns will become more out-of-sync with their school schedule as they enter puberty. I therefore expect that δ_1 in Equation 1-1 will be positive because later start times likely increase performance for children of all ages, and that δ_2 will be positive to reflect the greater benefits of later start times for adolescents.

One potential concern with this strategy is that the vast majority of cross-boundary moves are over a great distance. Long-distance moves may be inherently disruptive and therefore have an independent effect on academic outcomes. I address this concern by including in the sample students who move schools, but *not* across the time zone boundary. These students identify a set of dummies for 1, 2, and 3+ years after the move, disentangling the effect of moving from the effect of moving across a time zone boundary.

Data

Academic outcomes

The data come from Florida Department of Education (FDOE) administrative records for the fifteen school years from 1998-1999 through 2012-2013 (henceforth, 1999 through 2013). I exclude alternative schools, adult education centers, and virtual academies that may have non-standard start times. The primary outcome of interest is individual-level scores on the annual Florida Comprehensive Assessment Test (FCAT) in math and reading; this test is considered "high stakes" for students and schools. Students took the FCAT in math in grades 5 and 8 in years 1999 through 2000, grades 3 to 10 in 2001 through 2010, and grades 3 to 8 in 2011 through 2013. They took the FCAT in reading in grades 4 and 8 in 1999 through 2000 and grades 3 through 10 in 2001 through 2013. Scores are standardized by year and grade at the state level for each test, with a mean of zero and a standard deviation of one. In addition to the FCAT, the data include individual-level characteristics such as race, ethnicity, gender, free- or reduced-price lunch (FRL) eligibility, and absentee rates. I use student birthdays to calculate age at the start of the school year in September.[6]

The longitudinally-linked data allow me to follow students over time, as long as they remain within the Florida public school system. About 90% of students are matched year-to-year by social security number; the remaining students are matched by name and birthday. This matching process is conducted by the FDOE and appears to contain a small number of errors

[6] The FDOE uses September 1 as the kindergarten admission cutoff.

caused by multiple students with similar names or birthdays. To account for this, I exclude students who move backwards more than two grades, fail and then skip a grade, have a change in birthday, are older than 15, or change gender from year-to-year. In total, these deletions amount to about 7% of the original dataset. I lose few students in the longitudinal analysis; among students who took the third grade FCAT before 2009, I observe 93% taking an FCAT the following year and over 80% taking an FCAT five years later.

I restrict the sample in two main ways to address possible threats to identification. First, I focus on the area near the time zone boundary. This reduces the likelihood that there are different economic trends on either side of the boundary, which could mean that moves in one direction were disproportionately induced by job loss. Parental job loss is often a stressor for children and may itself have a negative impact on academic achievement; this could bias the results in either direction. The area near the time zone boundary is known as the Florida panhandle, and is generally seen as distinct from the rest of the state.[7]

Second, I limit the sample to students who make a substantial move, defined as consecutive appearances at schools farther than 25 miles apart. This restriction is largely targeted at the within-time zone movers; I want to ensure that these students are subjected to something comparable to the disruptive, long-distance cross-time zone moves. The exact choice of 25 miles as the cutoff is admittedly arbitrary; in the Appendix, I show that the main results are similar when using 15, 20, or 30 miles as the cutoff, or defining a move as a change in school district.

Table 1-1 displays summary statistics for third-graders in the panhandle. Note that this is a subset of the main estimation sample; I do not require that I observe a student in third grade to

---

[7] The panhandle includes the following 19 counties: Bay, Calhoun, Escambia, Franklin, Gadsden, Gulf, Holmes, Jackson, Jefferson, Lafayette, Leon, Liberty, Madison, Okaloosa, Santa Rosa, Taylor, Wakulla, Walton, and Washington. The time zone boundary approximately bisects the area.

include them in the main analysis. However, because I intend to show that test scores are directly affected by time zone through the start time channel, observed differences in test scores for older children are not informative about baseline characteristics. The third grade summary statistics in Table 1-1 are therefore as close to baseline summary statistics as is possible with the data, although there may already be some effect of differing relative start times.

Panel A presents school-level outcomes for all students in the panhandle (Column 1); for those who move more than 25 miles (Column 2); and for those who move more than 25 miles between time zones, disaggregated by direction of move (Columns 3 and 4). Column 5 tests the difference between Columns 3 and 4. Movers come from nearly identical schools as non-movers on all dimensions. Comparing within cross-boundary movers, CT-ET movers come from fairly similar schools as ET-CT movers across most measures; two differences stand out as large and statistically significant. First, the schools in ET have a much larger percentage of black students. This occurs because most black students in the sample are from Tallahassee and its surrounding suburbs in ET. Second, the district-level third grade reading score of the cross-time zone movers' schools is 0.08 SDs higher in CT than in ET. This would be problematic for identification if it implied that underlying peer quality improves when students move from ET to CT. However, this pattern may actually be a *result* of later relative start times in CT, because these students have already been treated with four years of later relative start times in grades K-3. In contrast, peer covariates like FRL, which are less affected by sunlight levels, are more similar between time zones. As a precautionary measure, I control for some characteristics of the peer populations with demographic share controls in the main specifications. In the Appendix, I show that the results are robust to the inclusion of controls for peer mean test scores.

Panel B presents individual-level characteristics. The movers are quite similar to the overall panhandle population, which bodes well for external validity. Movers are 11 percentage points more likely to be FRL relative to the non-movers, but equally likely to be black. Their test scores are slightly lower than the non-movers (0.09 and 0.08 SDs lower in math and reading, respectively), possibly reflecting stress from the upcoming move or slightly higher poverty rates among movers.

The characteristics of cross-time zone movers who begin in CT and those who begin in ET are balanced in terms of demographics, although the third grade math score is an insignificant 0.06 SDs lower for the CT-ET movers. The CT-ET movers also have one percentage point lower absentee rates than ET-CT movers.

Overall, Table 1-1 indicates that the two different types of cross-time zone movers are similar but not identical in terms of third grade characteristics and those of the schools they attend. Equality of baseline outcomes is not strictly required for the identification strategy; I make only the difference-in-differences assumption that the unobserved changes in average achievement had the students moved at a different time (or moved but not been exposed to a different relative start time) be the same for both types of mover. There are two main ways that this could be violated: if the ET-CT movers are on a different trend than the CT-ET movers, or if there are different changes in school quality over the move for different mover types.

The patterns of achievement in the years before the move provide evidence on the similarity of the underlying trend for each of the mover groups. Figure 1-1 displays pre-move trends for four types of movers – two within a time zone (CT-CT and ET-ET) and two across (CT-ET and ET-CT) – estimated from a regression of test scores on the number of years until move interacted with mover type. I include a vector of controls[8] and a fixed effect for the period preceding a move for each student. The year before the move is the excluded category. The Figure shows that the trend for each mover group is similar: in both math and reading, the test scores for each group are statistically indistinguishable from each other during the pre-move period. Time until move is also not a very strong predictor of academic achievement; for all but two of the group-time combinations, I cannot reject that there is no difference in achievement between that year and the year immediately preceding the move. This suggests that the groups are on similar underlying trajectories, and that variation in post-move outcomes can be attributed to changes in sunlight before school, rather than differential trends.

One slightly surprising finding is that math scores trend upwards for all groups in the years before the move. Long-distance moves are often a result of parental divorce or job loss, which may occur several years before the move actually takes place. Because both of these events can increase stress levels for children, it might be expected that in the absence of controls, test scores would decline leading up to a move. In the Appendix, I confirm this intuition; in a version of the same Figure without controls, I show that both math and reading scores unconditionally decline in the years before a move. Although I prefer the version with controls to maintain comparability with the main results, the substantive conclusion in both cases remains the same: there are no large differential trends that would threaten the identification strategy.

Another violation of the exclusion restriction would arise if school or neighborhood characteristics changed dramatically over the move. In the Appendix, I present evidence that

[8] I include all controls from the baseline regressions, which I discuss more in the Results section. They include age-gender dummies, longitude, and school-level demographic means (male, FRL, black, Asian, and Hispanic). The longitude and demographic coefficients are identified from small deviations in school location and school demographics in the years before the move, but have no substantive effect on the coefficients of interest. I include them for comparability with the main regressions.

changes in these characteristics are unlikely to drive the results. Taking the year before and after each move, I regress school characteristics on a set of student-move dummies and a dummy for each of the four types of move. Relative to the schools they started in, CT-ET movers move to schools with 4.5 percentage points fewer FRL students, 14.0 percentage points more black students, and a median zip code income $5,700 higher (ET-CT movers see approximately the opposite changes). In the absence of any other intervention, this might actually raise achievement for CT-ET movers given the strong relationship between average income and school quality, when in fact I see the opposite.

Imputing puberty

I do not directly observe the onset of puberty, and instead use data from the National Health and Nutrition Examination Survey (NHANES) to impute developmental stage by age and gender. NHANES is a nationally representative sample of US children ages 8 to 19, and includes information on Tanner Stage, a 1-5 scale of pubertal development based on pubic hair. I use the median age of entry into Tanner Stage 3 as the cutoff for adolescence, as changes in sleep patterns occur after the acceleration of pubertal development during Tanner Stage 3 (Campbell et al., 2012).[9]

The Appendix includes a figure displaying the cumulative share of children who have reached Tanner Stage 3 by gender and age; the median age of entry occurs at 11 for girls and 13 for boys. I use these ages as the start of puberty in the analysis.

---

[9] A second version of the Tanner Stage uses genital and breast development to demarcate stages. I use the pubic hair definition because the scale is more closely associated with pubertal changes in sleep patterns (Campbell, Grimm, Bie, & Feinberg, 2012), although using the alternate definition does not substantively change the main results. Using pubic hair Tanner Stage 2 or 4 changes the precision but not the direction of the results. Full results are available in the Appendix.

School start times

I define school start time as the start of the first class where learning takes place; this excludes homeroom and breakfast. Data were mostly available on school websites, and a research assistant followed up by phone with all remaining schools.

I did not collect information on historical school start times, which change with some regularity according to the school principals I spoke with while conducting the survey.[10] Given the identification strategy, the estimates will be consistent if there has been no change in the average start time for each time zone over the study period.[11] I believe that this condition is likely met: although there has been some recent discussion of school start time policy in the popular press, most of the data is from before this conversation reached the mainstream. Furthermore, the debate has never touched on whether early start times are more onerous for students with a later sunrise time.

School start times range from 7:00 a.m. to 9:30 a.m. local time. The average start time is 8:10 a.m., and the median is 8:00, which is similar to the national average (U.S. Department of Education, 2012). There is some heterogeneity with age: the median elementary school student starts school at 7:55, the median middle schooler at 8:25, and the median high schooler at 7:50. Nationwide, it is common to have high schools start earlier than the other schools in the district, so these broad patterns are not surprising.

I use NCES school location data to calculate sunrise times for each school. Combining these with the school start time data, I average the difference over the school year before the

[10] This means that any attempt to estimate Equation 1-1 by OLS would result in attenuated coefficients due to measurement error on the right hand side.
[11] Under a more restrictive linear relationship between achievement and start times, I require only that there has been no change in the difference in start times between the two time zones.

testing date to construct a measure of relative start time, measured as the number of hours between sunrise and school start times.

Results

First stage

The first stage is predicated on the idea that although school start times may differ across the time zone boundary, they do not do so enough to erase the one-hour difference in sunrise times. Figure 1-2 plots the hours of sunlight before school, or relative start time, in the years before and after a move for each of the four groups of movers. I estimate each point from a regression of relative start times on time relative to move for each group as well as an individual-move fixed effect and controls for longitude and school demographics. The year before the move is normalized to be zero; I adjust the level of the coefficients with the group mean of relative start times for one year before the move.[12] There are three important takeaways. First, students in Central Time have more sunlight before school than those in Eastern Time, as expected. Second, the cross-time zone movers neatly switch places as they move across the time zone boundary: the cross-time zone movers are now "treated" with the start time of the other time zone. This shift allows us to identify the effects of start time relative to sunrise using only within-student variation. Third, the lines generally overlap within time zones, indicating that those who switch time zones are likely not selecting into schools in a way that affects sunlight before school.

More formally, Panel A of Table 1-2 presents the first stage regression of relative start times on time zone.[13] The first row displays the main effect for all students, and the second row displays the interaction effect for pubescent students. The third row is the *p*-value from a test for

[12] A version of this graph with unconditional means for each group-time bin shows similar patterns.
[13] The Appendix includes robustness checks using additional controls including urbanicity, log income, school size, student/teacher ratio, and other levels of demographic aggregation. The results are similar to Table 1-2.

the combined significance of the effect for pubescents. Each specification includes individual and age-gender fixed effects. Column 1 has no additional controls. Column 2 adds longitude.[14] Columns 3 and 4 add demographic means at the district and school level, respectively. These demographic means include the percentage of students who are male, FRL, black, Hispanic, and Asian. Columns 5 through 7 are identical to Columns 2 through 4, but with the addition of indicator variables for 1, 2, and 3+ years after the move to account for potential disruption.[15]

All specifications yield similar estimates. I prefer Column 7 because it includes controls that address both disruption and potential changes in peer characteristics over the move. Across the columns, younger children in ET have about 25 fewer minutes of sunlight before school than children in CT, while those who have gone through puberty have about a 41-minute difference. The difference is less than 60 minutes for each age group, which is what I would expect if schools opened at the same clock time on either side of the time zone boundary. I take this as evidence that policymakers faced with later sunrise times may shift start times later to compensate, and that they may differentially shift elementary start times to prevent younger students from waiting for the bus in the dark.[16] The F-statistics for the first stage range from 825 to 2004, with an F-statistic of 1105 for the preferred model.

---

[14] I also consider adding latitude as a control. However, the study area has a relatively small north-south dimension – from the top to the bottom of the panhandle, the difference in average sunrise time over the school year is less than a minute. When I include latitude as a control, the main results are very similar but slightly smaller in magnitude. These robustness checks can be found in the Appendix.

[15] I consider specifications that control for the time until the move. This has almost no effect on the other coefficients in both the first and second stage, but I do not pursue this avenue to avoid controlling for information that the students may not have themselves.

[16] When I look at results by age, the difference in sunlight before school is 22-23 minutes for elementary school students (typically ages 8-10 in the data), 28-30 minutes for middle school students (ages 11-13), and 47-59 minutes for high school students (ages 14-15).

Effect of start times on academic achievement

Panels B and C of Table 1-2 contain estimates for the effect of relative start times on math and reading test scores. Each specification includes individual fixed effects and age-gender dummies,[17] and the columns add additional controls in the same order as Panel A.

In Panel B, the estimated effect of relative start times on math scores is similar after I add a control for longitude in Column 2. In all subsequent specifications, moving start times one hour later increases math scores for prepubescents by 0.009-0.020 SDs; none of the coefficients are close to statistically significant. For adolescents, later start times increase math scores by 0.077-0.084 SDs. Across specifications, both the adolescent level and the difference between adolescent and prepubescent scores is significantly different from zero at the 1% level.[18]

Panel C repeats the exercise for reading. The results are again consistent across the columns; in the preferred specification moving start times one hour later increases reading scores by 0.061 SDs for prepubescent students and by 0.057 SDs for adolescents. The overall effect for adolescents is statistically significant at the 1% or 5% level for all specifications, while for prepubescents it is either significant at the 5% or 10% level depending on the level of aggregation for the demographic controls. There is no statistical difference between pubertal and prepubertal effects. For adolescents, the effect size is larger in math than in reading across

[17] Test scores are normalized at the year-grade level, so if I included the entire state population the age-gender dummies would reflect only the age-varying gender gap. Because the sample is restricted to movers in the Florida panhandle, there may be additional age-varying differences relative to non-panhandle and non-mover students that the age-gender fixed effects pick up. They are particularly important to include because they function as a set of saturated dummy variables for puberty, which I interact with start time as an explanatory variable of interest.

[18] The difference in effect size by pubertal stage is striking, and corresponds with increasing sensitivity to start times during puberty. In the Appendix I estimate a version of Table 1-2 without the interaction. The average effect of start times on achievement is close to the average of the adolescent and pre-pubertal measures; the reading estimates are statistically significant but the math estimates are only sometimes statistically significant.

specifications, corroborating previous research on middle schoolers (Edwards, 2012; Ng, Ng, & Chan, 2009).

Mechanisms

There are (at least) two reasons why school start times might affect academic achievement. First, later start times relative to sunrise may make it easier to get to school on time, reducing absences and increasing time spent on instruction. Additionally, more sunlight before school may improve cognitive function by increasing sleep levels and alertness.

Panel D of Table 1-2 explores the relationship between start times and absences. Conditional on school or district level demographic controls, there is no statistically significant relationship between start times and absence rates for adolescents, although there is an estimated 0.9 percentage point decrease in absences for the younger students in the preferred specification. For all ages, later relative start times decrease absences, although the relationship is weaker for adolescents than for prepubescents, which is difficult to reconcile with the larger effects of start times on achievement I observe in math and reading. Comparing between age groups is somewhat fraught; because record-keeping is not standardized across schools, an elementary-aged child might be marked absent for the entire day when she is late in the morning, but a high schooler who is similarly late could be marked absent only for the first class but not as absent in the larger tracking system. However, that caveat addresses only differences between the age groups; in light of the moderate and imprecisely estimated effects on absences for all age groups, I think it is unlikely that reductions in absences are a major causal channel through which later relative start times translate into improved test scores.[19]

[19] I do not have data on tardiness, which could also be affected by start times.

The evidence is somewhat stronger in favor of sleep and alertness as the causal channel. The data do not contain information on sleep, so I use the Child Development Supplement (CDS) of the Panel Study of Income Dynamics (PSID) to estimate the effect of the time zone boundary on sleep. The CDS collected time use diaries for students in 1997, 2002, and 2007, along with geographic and demographic information. I regress hours of sleep on a dummy variable for residence in CT for children within 400 miles of the CT-ET boundary.[20]

Table A1-7 in the Appendix shows that prepubescent children in ET get 6 minutes less sleep per night during the week than children in CT.[21] The difference in sleep is reversed on the weekend as they attempt to correct the sleep deficit; students in ET sleep 4 minutes *more*. After the onset of puberty, both gaps widen: children in ET get 17 minutes less sleep per night during the week, and compensate with 13 minutes more sleep per night on the weekend.

These findings indicate that children in ET are more sleep-deprived than children in CT, and that this gap increases in adolescence. If school start times in the Florida sample are representative of start times elsewhere, this suggests a pass-through from relative school start times to sleep of 40-50%, which is comparable to the 46% found by Wahlstrom et al. (1998). Thus, moving from ET to CT increases both sleep and test scores (and increases them more for adolescents), suggesting that levels of sleep and alertness in the morning are important causal channels through which later school start times increase achievement. There may be other changes in time use – descriptive research indicates that later start times decrease time spent on extracurricular activities, leisure time for girls, and computer use for boys (Groen & Pabilonia,

[20] The publicly-available CDS does not geocode individuals at a sub-state level, so I exclude all observations from states with multiple time zones – including Florida. See the Appendix for more information on sample construction.
[21] All estimates reported here include demographic controls; see Column 2. I conservatively cluster by state. The difference in sleep between children in ET and CT is statistically different for adolescents but not for prepubescent children.

2015; Wahlstrom et al., 1998) – but it is difficult to reconcile the patterns of achievement by developmental status with an explanation *not* revolving around the transition to puberty. More importantly, from the perspective of a policymaker the distinction is moot: whether the causal channel is before-school time or after-school time, changing the school start time will affect both channels.

Heterogeneity by age and gender

Rather than allowing the effect of relative start times to vary by pubertal status as in Equation 1-1, it is possible to estimate each age-gender-start time interaction term separately. If the increasing importance of start times for math performance is a function of puberty, the effect sizes should grow in importance as a larger share of the gender enters puberty. This is precisely what occurs.

Figure 1-3 presents coefficients from a version of Equation 1-1 estimated separately by gender, with start time fully interacted with age. Because ages range from 8 to 15, this amounts to estimating:

$$y_{it} = \sum_{a=8}^{15} \delta_a \; \delta_1 hours_{it} \times \mathbb{1}[\text{age}=a] + X_{it}\beta + \gamma_i + \varepsilon_{it} \qquad (1\text{-}4)$$

where $hours_{it} \times \mathbb{1}[\text{age}=a]$ is instrumented by time zone interacted with age and X_{it} is the baseline vector of controls. Starting in the upper left corner of Figure 1-3, there is a sharp spike in the effect of school start times on math scores at age 11 for girls, precisely when the median girl enters Tanner Stage 3.

The effect of later school start times is statistically significantly different from zero for girls 11-13, but not for girls 10 or younger. For boys, in the upper right corner, the effect of start times on math scores is statistically indistinguishable from zero at the 10% level for ages 8 to 12, then jumps from 0.049 to 0.096 at 13 as the median boy enters Tanner Stage 3. The effect of start times is significantly different from zero at the 1% level for ages 14 and 15. This is evidence that the increasing importance of start times with age is driven by pubertal entrance, rather than other academic or behavioral changes.

The effect of start times on math scores is noticeably (though insignificantly) smaller for girls after age 13. One possible explanation is that certain stages of puberty are particularly important for sleep (Campbell et al., 2012), and girls have moved beyond this developmental stage by age 14. For example, Crowley et al. (2007) speculate that older adolescents may be less responsive to light than younger adolescents. However, there is no firm physiological evidence on sleep patterns or light sensitivity at a granular gender-age level, so resolution of this issue will have to wait for data which extends further into adolescence, especially for boys. There is persuasive evidence from Carrell at al. (2011) that start times have a large effect on achievement for college freshmen cohorts that include both boys and girls, so it is unlikely that the true effect is zero for 14 and 15 year old girls.

In reading, as one might expect from Table 1-2, there is no sharp change in the relationship between start time and achievement at the gender-specific puberty thresholds.

Heterogeneity by subgroup

Educational interventions often have a larger effect on disadvantaged students or students attending low-resource schools (see, e.g., Krueger & Whitmore, 2002). In this case, however, there are more similarities than differences in effect sizes across racial, economic, and gender groups. The standard errors are large, but the results suggest that changes to start times will benefit all students, rather than certain demographic groups.

In Table 1-3, I apply the baseline regression of test scores and absence rates on start times for each of six demographic subgroups: whites and minorities;[22] FRL and non-FRL; and male and female. In math, the effect sizes are similar between white and minority students in Columns 1 and 2. For pubescents, a one-hour delay in relative start times increases math scores by 0.093 SDs for white students and 0.081 SDs for minority students. In reading, the effect sizes are 0.040 and 0.132, respectively, though this difference is not significant. None of the estimated effects for absences are statistically significant.

Columns 3 and 4 contrast FRL and non-FRL students. The effect size for math scores is significantly larger for the non-FRL adolescents at 0.147 SDs per hour, compared to 0.048 SD for FRL adolescents. There are no statistically significant differences for reading scores or absence rates.

Finally, Columns 5 and 6 indicate that the effect of relative start times on achievement and absences is similar for boys and girls. The difference is never statistically significant, and the effect sizes for both groups are similar to the overall estimates of Table 1-2.

Persistence of start times

To this point, the analysis has not distinguished between a transitory and permanent effect of start times on academic achievement. This distinction could be important. If changing school start times from one year to another has an effect for (say) only one year while the student

[22] I count all non-white students as minorities. These results are not substantively affected by not counting Asians as minorities, or delineating the categories as black and non-black. In the latter case, however, the standard errors for the black sample are large.

adjusts her sleep schedule, the estimates (which are essentially the average of achievement before and after the move) would overstate the long-term effect by averaging a positive effect in the first year with a zero effect in all other years. This would mean that the estimates would not correctly predict the long-term change in achievement as a result of changes in start time policy. I explore this possibility in Table 1-4, where I estimate a version of the baseline regression with relative start time by pubertal status interacted with dummies for 1, 2 and 3+ years since move. Note that the pubescent effect is the total estimate for adolescents, rather than an interaction.

The results indicate that the short-term and long-term effects are quite similar; for prepubescent children the long-term math and reading coefficient is an insignificant 0.005-.011 SDs smaller. For adolescents, the math effect is 0.020 SDs lower in the long run; the difference is significant at the 1% level. The reading effect is 0.010 SDs higher in the long run; the difference between the short and long run is not statistically significant. In the long run, the effect is larger for adolescents than younger students in both subjects, although the difference is not statistically significant in reading. In both the short and long run, the adolescent effects differ from zero. I conclude that changes to start times improve math and reading achievement within a year of the change in sunlight exposure for adolescents, and the effects largely persist over time.

Learning versus testing

The positive effect of later relative start times on test scores has two potential causes: improved learning in the year leading up to the test, or better testing performance caused by increased alertness on the day of the test. The approach so far has been to estimate the combined effect of learning and testing. Fully disentangling the two effects would require separate instruments for start times during the year and on the day of the test, which are unavailable in the data.

The data allow us to answer a related but less definitive question: does the relationship between sunlight and achievement vary with the amount of baseline test-day sunlight, holding sunlight during the school year constant? If so, this implies that changes to test-day relative start times matter for achievement. Estimates of the marginal effect of later relative start times at different levels of test-day sunlight can be combined with a mild assumption of diminishing returns to sleep to generate a lower bound on the size of the test-day start time effect.

This strategy is possible in this context because the data contain variation in test-day relative start time that is separate from the cross-time zone variation in start times. During the study period, testing dates moved from late February to mid-April. This changed levels of sunlight on the day of the test, but had only a small effect on average sunlight levels during the school year when learning occurred. Using these policy changes, I find that the lower bound on the test-day effect is relatively high for reading, but low for math. I interpret this as evidence in favor of potential testing effects in reading, but not as a definitive rejection of testing effects in math.

During the study period, the FDOE pushed the testing period later in two discrete steps. The first change was particularly useful for this research, because it moved the testing period from before to after the start of Daylight Saving Time. DST begins with a time change on the second Sunday of March in most of the United States.[23] Clocks "spring forward," moving sunrise one hour later and reducing the amount of sunlight before school. Figure 1-4 charts sunlight

---

[23] There have been changes in DST dates in the recent past; before 2007 DST started on the first Sunday of April. This change is not relevant for this research, because testing occurred before DST began in all years before the switch in DST dates.

before school for 2000-2007, 2008-2009, and 2011-2013, corresponding to the three test-day policy eras.[24] In 2000-2007, testing took place just before the change to DST, meaning that there was a relatively large amount of sunlight before school; in ET, the average was 1 hour 20 minutes on the first day of testing. For 2008 and 2009, the test was moved two weeks later to directly after DST; the average amount of sunlight before school on the morning of the test in ET dropped to 28 minutes. In 2011, the test was moved one month later, increasing sunlight before school on the testing day to an average of 1 hour 9 minutes for 2011-2013.[25] Throughout the study period, the average sunlight before school in the school year leading up to the test barely changed, at 61, 56, and 59 minutes, respectively. Based on these differences, I group together 2000-2007 and 2011-2013 into a "late test time" treatment, and 2008-2009 into an "early test time" treatment.[26] As the testing date was moved back, preparation time increased for all students. However, because the early test time treatment occurred in the middle of the period (when the testing date was closest to the DST transition) the average difference in preparation time is only five days longer for the late test time treatment group. Furthermore, neither of the changes in testing date correspond to any major changes in testing procedure or curriculum I could find, suggesting that any differences in performance between the policy eras can be attributed to test-day sunlight.

It is tempting to estimate the effect of earlier relative start times on the day of the test by regressing test scores on a dummy variable for the testing era. However, test scores are

[24] Specifically, the Figure shows 2007, 2008, and 2011, but all are archetypes of their eras.
[25] I exclude 2010 from analysis in this section because DST occurred during the testing period in this year, meaning that I cannot assign the test to either pre- or post-DST. I also exclude 1999 because testing occurred one month earlier, in the first week of February, where the sunrise time is between the early and late period.
[26] The main difference between 2000-2007 and 2011-2013 is that the average relative start time in the year preceding the test was slightly earlier in 2011-2013 because the extra month of class time was almost entirely after the DST transition. Excluding 2011-2013 from the regressions does not change the conclusions.

standardized by the mean statewide score in each grade-year, so direct comparisons between years are not possible. I instead test whether the effect of full-year relative start times on achievement changes depending on test-day sunlight. I estimate a second stage of:

$$y_{it} = \varphi_1 hours_{it} \times \mathbb{1}[\text{child} \cap \text{late test time}]_{it} + \varphi_2 hours_{it} \times \mathbb{1}[\text{child} \cap \text{early test time}]_{it} + \lambda_1 hours_{it} \times \mathbb{1}[\text{puberty} \cap \text{late test time}]_{it} + \lambda_2 hours_{it} \times \mathbb{1}[\text{puberty} \cap \text{early test time}]_{it} + X_{it}\beta + \gamma_i + \varepsilon_{it} \quad (1\text{-}5)$$

where X_{it} includes, in addition to the usual controls, dummies for the policy eras and their interaction with puberty.

Because sunlight before school during the year leading up to the test is nearly identical between eras, the difference in coefficients for a given age group represents the change in the effect of one extra hour of test-day morning sunlight on test scores between two different margins: 1 hour 17 minutes from sunrise (the average in the late testing years) and 28 minutes from sunrise (the average in the early testing years). If the coefficients are the same, that implies either that the effect of test-day sunlight is identical at the two margins, or that the effect of test-day sunlight is zero.[27] If they are different, that implies there is some effect of test-day sunlight on at least one of the margins. A smaller coefficient in the late-testing years is consistent with diminishing marginal returns to test-day sunlight. Analogously to the main specification of Equation 1-1, I expect that $\lambda > \varphi > 0$, since later start times should improve performance more for adolescents than for younger students.

[27] The latter implication is technically a subset of the former, but the conceptual difference is important.

Table 1-5 presents the findings for math and reading. Unlike the main table, the coefficients estimate the full effect for adolescents, rather than the difference between adolescents and younger children. I begin by verifying in Columns 1 and 3 that excluding 1999 and 2010 does not substantively affect the baseline results.

Columns 2 and 4 estimate Equation 1-5, allowing for a differential effect of start times on achievement as a function of baseline test-day start times. In Column 2, the math results are unchanged from the main specification: moving relative start times one hour later increases achievement at a similar rate in the two eras for adolescents (0.096 SDs per hour in the early era versus 0.095 SDs in the late era), and the difference in estimates is statistically insignificant. Because I argue there should be diminishing marginal returns to more sunlight before school, I take the similarity in estimates between different test-day sunlight eras as evidence against test-day effects in math.[28] For younger children, the results are slightly more suggestive of testing effects, with larger effects for more sunlight on the test days with less sunlight before school (at 0.071 SDs per hour) than on the test days with more sunlight before school (at 0.022 SDs per hour). However, neither estimate statistically differs from zero, nor do they differ from each other.

In reading, the results are more strongly suggestive of testing effects. For younger children, one extra hour of morning sunlight increases test scores by 0.096 SDs in years with less sunlight before school (early testing years), while the effect is statistically insignificant and only 0.049 SDs in years with more sunlight before school (late testing years). For adolescents, the

[28] Technically, the similarity between the early- and late-test time coefficients cannot be read as a failure to reject testing as an important input into math achievement. It is instead a rejection of a nonlinear relationship between achievement and test-day sunlight --- it is consistent with an effect of test-day start times only if the relationship between achievement and start times is linear in the region between 28 minutes and 1 hour 17 minutes of sunlight before school.

effect during the relatively earlier testing era is 0.104 SDs per hour of sunlight, compared to 0.045 SDs in the late era. The difference in estimates is statistically significant for adolescents, suggesting that test-day sunlight may be important for reading achievement. Under the assumption that changes to test-day relative start times do not change the effect of start times during the school year, and that there are decreasing marginal returns to later test-day start times, this indicates that the test-day effect is bounded at a minimum of 0.059 SDs per hour for adolescents (calculated as 0.104-0.045) and 0.047 for prepubescents (0.096-0.049) in the early start time years. This bounded effect implies that testing is a more important causal channel than learning for reading achievement.

There is, however, one important reason why the result in reading should be taken with some caution. In both of the early-testing years, the testing period began almost immediately after the switch to DST; one day after in 2008 and two days after in 2009. Because clocks move forward during the spring DST transition, students can lose up to an hour of sleep, depending on how much they adjust their sleep times. There is strong evidence that the DST transition negatively affects sleep levels and alertness: Smith (2016) finds an increase in the number of fatal car accidents in the six days following DST. I therefore interpret the difference in coefficients between the early- and late-baseline years as the difference in the gains from an hour of sunlight on test-day with a baseline of 1 hour 17 minutes sunlight before school and the gains from an hour of sunlight on test-day with a baseline of 28 minutes of sunlight before school *and* up to an hour of sleep deprivation. I have no information on the testing date for each student, so I cannot further stratify the start time effect as a function of number of days since the DST transition. However, since the testing period was longer than one week in both 2008 and 2009, the test was likely taken a few days after the DST transition and perhaps as long as two weeks after, when transition-induced sleep loss has lessened. I therefore think that the safest interpretation is for moderate test-day effects in reading, of the same order as the full-year learning effects. At the very least, this result tells us that under an assumption of diminishing marginal returns to test-day sunlight, there are some situations (potentially including more sleep deprivation than is normal for this age group) where test-day sunlight has a large effect on academic achievement in reading. There is much more to be done to separately identify the effects of whole-year and test-day sunlight, but I leave this for future research.

Placebo time zone changes

The identification strategy in this paper leverages the discontinuity in sunrise times at the time zone boundary to estimate the effect of relative start times on academic performance. In a reduced form sense, I track students as they move east (west) over the time zone boundary in the Florida panhandle and find that scores decline (increase), as predicted by the earlier (later) relative start times.

Alternatively, perhaps moves to the east are score-decreasing for some reason unrelated to start times: schools are lower quality, or parents moving east get worse jobs and lower pay, which decrease investment in educational inputs. The baseline specification includes controls for longitude and school demographics, which together control for any variation in underlying school or family characteristics that is linearly correlated with the demographic controls or varies linearly from east to west over the panhandle. If there are nonlinearities in this relationship, however, the method could misattribute variation in unobserved non-start time inputs to variation in start times, biasing the estimates.

In this section, I estimate placebo regressions that attempt to rule out a non-start time explanation. I generate placebo boundaries in ten-mile increments from the true boundary; Figure 1-5 displays the estimated effect of moving over each placebo boundary, conditioning on true time zone, the regular vector of controls, and student fixed effects. I present estimates using cross-time zone movers as well as restricting to only within-time zone movers. In the placebo section of the Appendix, I demonstrate that schools very close to the time zone boundary adopt start times similar to their cross-boundary counterparts; this means that there is a treatment effect of moving to or from the region directly adjacent to the boundary, even when the move is within time zone. I therefore exclude a 25-mile area around the true boundary (a version of the placebo test without this exclusion is available in the Appendix).[29]

Figure 1-5 displays the estimated coefficients for moving over placebo boundaries, placed in ten-mile increments from the true time zone boundary. In math, the placebo coefficients for the adolescent interaction are always smaller than the true coefficient and usually significantly so. The true level coefficient is approximately zero, and the placebo coefficients bounce around that estimate, although I can sometimes reject they are zero. In reading, for both the within- and all-mover specifications, the placebo coefficients are almost always smaller than the true coefficients (and very imprecisely estimated when they are not). The true time zone-puberty interaction coefficient is approximately the same size as the placebos, although it is imprecise enough that I cannot differentiate it from zero in the main sample.

[29] Excluding this region is not necessary in the main specification, as the IV estimate accounts for treatment bleed across time zones. However, the results are substantively the same even excluding this donut; I estimate that moving start times one hour later would improve math scores by 0.065 SDs for adolescents, and would have little effect on prepubescent math scores or reading scores for either age group. The number of students also decreases, resulting in larger standard errors on these estimates.

In summary, I estimate regressions of outcomes on placebo time zones, and find little evidence of changes in outcomes over the placebo boundaries, suggesting that the gains in achievement from westward moves are a function of crossing over the true time zone boundary and being exposed to later relative start times, rather than improvements in some other input.

Other effects of cross-time zone moves

A final threat to the identification strategy is the possibility that moving between time zones has a direct effect on family income or other characteristics. If these changes have an independent effect on academic performance, the exclusion restriction would be violated. Gibson and Shrader (2015) show that a one-hour delay in sunrise time reduces wages by between 0.5 and 4.5%. Given Dahl and Lochner's (2012) estimate of a 0.06 SD decrease in test scores per $1,000 decrease in EITC income, this could explain much of the test score effect. I do not observe parental income, and so cannot directly control for this possibility. However, there are three reasons to expect that a measure of income is not an important missing variable in the analysis. First, jobs are a primary reason for moving long distances and are chosen by the parents; wages are an important factor in job choice. It is therefore unlikely that movers are immediately treated with the average difference in wages given the change in sunrise times over the move. In fact, Gibson and Shrader (2015) argue that housing prices adjust to eliminate the incentive to move, and document that housing is indeed more expensive in early-sunrise cities. Disposable income would then be flat over the move, eliminating any effect on academic achievement.

Second, in the sample zip-level income is higher in low-sunlight ET than in high-sunlight CT, which is the opposite of what is predicted by Gibson and Shrader (2015).[30] As demonstrated in the Appendix, the results are unchanged by controls for zip-level income.

Third, and most importantly, even if disposable income did increase as families moved over the time zone boundary, I would expect that children of all ages would benefit from the move. Instead, I observe larger increases in standardized test scores for pubertal children – and almost no increase for pre-pubertal children in math – suggesting that changes in sunlight before school over the pubertal transition are the most important causal factor.

Benefits of rearranging start times

Academic research and popular coverage of the potentially negative effects of early start times dates back at least as far as the late 1990's (Martin, 1999; Wahlstrom et al., 1998). The evidence from the medical and physiological literature has grown so compelling that the American Academy of Pediatrics (2014) recommends that middle and high schools delay start times to allow students to wake up no earlier than 8:00 a.m. Despite the growing consensus, schools continue to open early; the median high school *opens* at 8:00 a.m. (U.S. Department of Education, 2012).

School districts, particularly those in large urban areas, often open different types of schools at different times. This structure is convenient for parents dropping off children at different schools, because it guarantees that a child in middle school will not need to be dropped off at the same time as a child in high school. It also allows school districts to use the same buses more intensively, saving on transportation costs. However, of the 19 school districts in the

[30] This does not seem to be a function of education, since literacy is actually marginally lower in ET (Author calculations from the NCES 2003 National Assessment of Adult Literacy).

Florida panhandle, only 4 currently order their start times in the "efficient" way. Inflexible parental schedules often preclude moving start times later for all students since parents must be able to drop off their last child in time to get to work. In this section, I consider the academic effects of an alternative start time policy that better fits the physiological evidence but does not alter the overall distribution of start times: changing the opening order for different types of schools to elementary schools, middle schools, and finally high schools.

I operationalize this simple counterfactual by taking the average start time for each school type in each district, then assigning the earliest average start time to elementary schools, the next start time to middle schools, and the latest time to high schools. I adjust the mean start time for each district so that it is the same in the counterfactual as in the real world.[31] I take the difference in relative start times for the counterfactual and real worlds for each school type and apply the coefficients from Table 1-3, weighting by the number of children in each district-school type. On average, this moves elementary start times 22 minutes earlier, middle schools 13 minutes earlier, and high schools 44 minutes later.

Figure 1-6 displays the effect on test scores, separated by gender and race. The counterfactual policy has been constructed so that if start times have an identical effect on children of all ages, the average increase in test scores will be zero. However, because the gains from later start times are smaller for younger children than for older children, the procedure has the effect of raising average academic achievement. In both math and reading, the effect is slightly (and usually insignificantly) negative for all groups of students in elementary and middle

[31] A clarifying example: if a district has 800 students in grade 9-12 schools with a start time of 7:00, 800 students in grades 6-8 schools with a start time of 7:30, and 1200 students in K-5 schools with a start time of 8:00, the mean district start time is 7:34. I would then set counterfactual start times to 7:08 in elementary school, 7:38 in middle school, and 8:08 in high school, with an average start time of 7:34. The procedure keeps the counterfactual mean start time the same as the status quo, and maintains the half hour spread in start times between school types.

school. On average, elementary- and middle-school math and reading scores decline by 0.01 SDs. For high school students, the gains are large and statistically significant: in math, the proposed policy would increase minority student achievement in high school by 0.06 SDs in math and 0.08 SDs in reading. For white students, I expect that math scores would increase by 0.06 SDs and reading scores by 0.02 SDs. By gender, male high school students benefit slightly but insignificantly more compared to females. Using the coefficients from Table 1-2, the average effect is a 0.064 SD gain in math and a 0.044 SD gain in reading.

Furthermore, the high school results are good estimates for the overall change in achievement for each student by the end of high school. Table 1-4 shows that increases in academic achievement occur immediately after the move and persist for years. Taking the long-term estimates of Table 1-4 as given, the counterfactual would increase end-of-high school math scores by 0.05 SDs and reading scores by 0.037 SDs.[32]

One drawback of re-ordering start times would be that the youngest children may have to wait for the bus or walk to school in the dark. In December, the average sunrise would be only 53 minutes before school starts, with 12% of elementary school students having less than half an hour between sunrise and school start in the darkest month. This would likely mean that a substantial number of very young students might need to travel to school in the dark, which presents a significant drawback to this proposal. Moving all school start times later, rather than re-ordering schools, would not have this problem.

In summary, I demonstrate that adjusting school start times so that high school students have the latest start time would significantly increase achievement for older children at a very

[32] This math score is calculated by multiplying the long-term coefficient of 0.087 - 0.020 = 0.067 by the average change in high school relative start times, 44 minutes.

low academic cost for younger children. Even when start times are reordered such that the average start time across the district remains the same, there are non-trivial gains in average academic performance that would benefit students in all demographic groups. These gains must be weighed against the costs of having younger children traveling to school in the dark.

Conclusion

I investigate the effect of daily school start times on academic performance. Adolescents in particular struggle with early start times; the onset of puberty shifts the sleep schedule back several hours, making any given start time more onerous for high schoolers than for students in other age groups. The empirical strategy tracks academic performance in the same student before and after a cross-time zone move, which I use as an instrument for the amount of sunlight before school. Because the circadian rhythm is tied to variation in sunlight levels, this is a good approximation of a policy change in start times. Using a long individual panel from the state of Florida, I find that moving start times one hour later relative to sunrise would increase adolescent scores by 0.081 SDs in math and 0.057 SDs in reading. The increase in test scores can be observed immediately after the move and persists for as long as I can measure it. Taking advantage of the fact that girls enter puberty two years earlier than boys, I document that the effect of relative start times on math performance spikes precisely at the gender-specific age of median entrance into an important pubertal stage. Previous research, which has mostly focused on a smaller age range of the population, has been unable to fully explore changes in the effect of start times over the pubertal transition.

These effects are cost-effective compared to other proposals to improve educational achievement, such as smaller classrooms or higher-skilled teachers. Specifically, reducing class size in elementary schools from 22 to 15 increases scores by 0.15-0.20 standard deviations (Schanzenbach, 2006), and a 1 standard deviation improvement in teacher quality increases scores by approximately 0.10 standard deviations (Chetty, Friedman, & Rockoff, 2014). Changes to school schedules would likely be much cheaper. Jacob and Rockoff (2011) suggest that the cost of moving start times one hour later is less than $150 per student per year and potentially as low as free. In contrast, reducing class sizes by a third costs approximately $6,200 per student per year.[33] The cost of such a large improvement in teacher quality is more difficult to evaluate since the supply side of the teacher market is poorly understood. However, it is likely very large, if only because it would likely require hiring hundreds of thousands of new teachers.[34]

I simulate the effect of adjusting start times by school type to match students' developmental patterns while maintaining the same mean district start time. I estimate that this would increase math scores for high school students by 0.064 SDs and reading scores by 0.044 SDs, while having small and mostly statistically insignificant effects on scores for younger children. Alternatively, moving start times later across the board would increase achievement for all ages and demographics. In either case, adjustments on the start times margin seem to be significantly cheaper than adjustments to classroom size or teacher composition, suggesting that there may be large unrealized gains in this area.

There is one important caveat to the findings. Changes in school start times can increase achievement through either better learning in the year leading up to the test, or improved testing performance. I exploit a policy change in the testing date relative to Daylight Saving Time to

[33] These figures are from Schanzenbach (2006), inflated from 2002 to 2011 prices via the CPI.
[34] If teacher quality were distributed normally, then replacing the bottom half of teachers with average teachers would raise the average SD of teacher quality by only 0.4, and therefore test scores by 0.04 SD. According to the NCES, there were 3.7 million teachers in the United States in 2012. It is hard to imagine that finding 1.85 million new average-quality teachers could be done without significantly increasing wages.

learn whether test-day start times are important for achievement (but not by how much). I find suggestive evidence in favor of testing effects in reading but not math. The method is unable to precisely quantify the relative importance of testing and learning, but show that the magnitude is approximately the same for reading. I leave this as an important direction for future work.

Despite growing medical and physiological evidence that current school start times are too early for optimal adolescent cognitive functioning, there has been little policy response to move start times later. I add to this debate with direct evidence that more sunlight before school – or a later relative start time – increases academic achievement for children of all ages. The increase in scores is much larger for adolescents, implying that even when parental schedules preclude later start times for all children, districts can improve academic performance by adjusting the order in which school types open to correspond with students' changing sleep schedules. Specifically, high school students should begin school later in the day to compensate for pubertal changes that shift their circadian rhythm later, while elementary students should begin school the earliest. Despite the low costs of adopting this policy, the gains are quite large.

Study 2. A Hard Day's Night: The Spillover Effects of Teen Pregnancy on the Academic Performance of Younger Siblings

Siblings share (limited) parental resources, the same neighborhood environments, and similar genetics, and it seems probable that an unexpected change in one sibling could change the outcomes of children living under the same roof. However, given their shared context, it is difficult to analyze the effect of one sibling on another, and little is known about how a negative shock to one sibling affects the rest of the family, particularly in older children (Black et al., 2016; Breining, 2014; Breining, Daysal, Simonsen, & Trandafir, 2015; Nicoletti & Rabe, 2014; Yi, Heckman, Zhang, & Conti, 2015). One presumably large shock that directly affects one child and may have ripple effects in the family is the birth of a child to a teenage mother. I use novel longitudinal data to study how teenage pregnancy changes siblings' outcomes following the birth, relative to similar students who did not have a teen birth in the family.

While there is an expansive literature on the effects of teenage motherhood on the mother's own outcomes, there is little research on what happens to the rest of the mother's family.[35] Understanding family-wide effects matters for policymakers as well as those considering using siblings as a comparator in an analysis on teen mothers (e.g., Geronimus & Korenman, 1992). Adding a newborn to the home might have profound effects on the whole

---

[35] Prior research has reached conflicting conclusions about whether teenage pregnancy causes poor outcomes for the teen mother herself or whether it is a symptom of prior trends. Teenage parenthood is popularly understood to be a negative outcome for the mother, including reduced education attainment and worse long-term economic prospects (Kane, Morgan, Harris, & Guilkey, 2013; A. R. Miller, 2009). However, many such studies do not account for the negative selection into pregnancy and, among those who get pregnant, the positive selection into abortion (Ashcraft, Fernández-Val, & Lang, 2013; Fletcher, 2011; Geronimus & Korenman, 1992; Hotz, McElroy, & Sanders, 2005; Hotz, Mullin, & Sanders, 1997). In other words, the type of women who become teenage mothers may have limited economic opportunities, and even if they had delayed pregnancy they likely would have had poor outcomes relative to other women. Instead, many of these studies use miscarriage as an instrumental variable to estimate the effects of teen motherhood, generally finding null or small effects (Ashcraft et al., 2013; Hotz et al., 2005, 1997). Other work has argued that miscarriage is correlated with community-level factors, and after accounting for this correlation teenage pregnancy in the U.S. does indeed reduce high school graduation rates and annual income (Fletcher & Wolfe, 2009).

family, including family conflict, additional responsibilities, and loss of sleep (see, e.g., Meltzer & Montgomery-Downs, 2011). The grandparents of this child often take on child care responsibilities, which can take away their time to work outside the household, increase their stress levels, and take away the time available for their other children (Bailey, Haynes, & Letiecq, 2013; Chase-Lansdale, Gordon, Coley, Wakschlag, & Brooks-Gunn, 1999; East, 1998). After the baby's birth, new grandmothers monitor and communicate less with their non-parenting children (East, 1999), perhaps allowing the siblings to make choices that harm their human capital development. Understanding the full consequences of teenage motherhood matters for policymakers in the United States, which has the highest birth rate among teenagers of any industrialized country (Kearney & Levine, 2012).

However, to date there has been almost no research on the effects of teen motherhood on the outcomes of other children in the family.[36] The only prior work along these lines pertains to sibling fertility: the siblings of teenage mothers are more likely to become teenage parents themselves, particularly when siblings are close in age and in low-income households (Monstad, Propper, & Salvanes, 2011).[37] This paper represents the first research, to my knowledge, that studies the effects of teen motherhood on their siblings' human capital development. I make use of detailed longitudinal student data from a large anonymous Florida county-level school district linked to postsecondary outcome data from the National Student Clearinghouse to study the

---

[36] There is evidence that siblings can affect each other's educational outcomes in other contexts. Younger siblings with disabilities or health problems can negatively affect their older siblings' educational outcomes (Black et al., 2016; Breining, 2014; Breining, Daysal, Simonsen, & Trandafir, 2015), while higher-achieving older siblings can also positively affect their younger brothers and sisters (Joensen & Nielsen, 2015; Nicoletti & Rabe, 2014).

[37] Kearney and Levine (2015) show that the MTV show *16 and Pregnant* led to a decrease in teen pregnancy, thanks in part to the demonstration effect of the ordeals of teen pregnancy and parenthood. Other work finds that peer pregnancy increases own pregnancy, at least among schools without family-planning services (Fletcher & Yakusheva, 2016), while a friend's pregnancy, as opposed to a miscarriage, decreases pregnancy (Kapinos & Yakusheva, 2016). It is clear from this other research that teenagers are responding to a variety of signals.

effects of teen motherhood on siblings' test scores, dropout likelihood, juvenile justice participation, college attendance, and college completion.

The primary causal identification problem is that teen pregnancy is generally not an exogenous event for the family, and the pregnancy itself may be a symptom of family conflict and disruption (Ellis et al., 2003; B. C. Miller, Benson, & Galbraith, 2001). Thus, teenage mothers may be on a downward trajectory well before the birth. Indeed, I demonstrate that both teenage mothers and their siblings have falling test scores for several years before the birth of the child.[38] Unless researchers account for these underlying trends, the negative estimated consequences of a birth in the family may reflect these unobserved family factors rather than the spillover effects of teen motherhood per se. Many identification strategies that work for studying teen fertility (e.g., Buckles & Hungerman's (2016) study of the effects of condom distribution programs on teen fertility) cannot help to disentangle the spillover consequences of teen pregnancy, especially given that siblings tend to be relatively closely spaced. In this paper, I make use of the longitudinal nature of the school district data, in which children are observed annually throughout their schooling years. I conduct an event study analysis by matching children in families experiencing a teen pregnancy event to observationally equivalent children who were on the same trajectories, in terms of test scores, in the years leading up to their sisters' pregnancies.

I show that, in terms of test scores, siblings of teen mothers and their matched comparators are on a similar downward trajectory for several years prior to the commencement

[38] This pattern supports the argument in the teen mother literature that teen mothers are on a different path than other teenagers, and that simple cross-sectional analyses would overstate the negative consequences of teen motherhood (Ashcraft et al., 2013; Fletcher, 2011; Fletcher & Wolfe, 2009; Geronimus & Korenman, 1992; Hotz et al., 2005, 1997).

of their sisters' pregnancies. However, the siblings of teen mothers diverge after the birth. For the siblings of teen mothers, there is a marked decrease in test scores of 2.6 to 4.0 percentile points following the birth, relative to those on a similar trajectory before birth. Among the sisters of teen mothers, high school dropout is 9.5 percentage points more likely; the effect is null for brothers. Among brothers, the chance of encountering the juvenile justice system increases by 9.7 percentage points after the birth; the effect is null for sisters. In the longer term, obtaining any post-secondary degree or certification drops by 6.7 percentage points, on average, while the chance of obtaining a four-year degree drops by 7.0 percentage points. As with high school academic outcomes, the college estimates are larger among the sisters than the brothers of teen mothers. The college outcomes are also largest among the middle tertile of siblings. Very few of the bottom tertile students obtain a degree or certificate, while the top tertile is unchanged.

I use supplemental data from the American Time Use Survey to study whether time allocation may be a mechanism driving the results. Using a proxy for the siblings of teen mothers[39] and a differences-in-differences framework, I find that the sisters of teen mothers spend more time on childcare on weekdays and less time on homework on weekends, relative to the brothers of teen mothers and other teenage girls. On weekdays, brothers of teen mothers are more likely to be alone, less likely to be with parents, less likely to be supervised by adults, and less likely to be with other children under 13 than other males, while the sisters are more likely to be with children relative to both brothers of teen mothers and other females. Substitution of time from homework to childcare may drive the academic results for the sisters of teen mothers,

[39] The data do not identify if respondents live with a niece or nephew. Instead, I use teenagers who live with a child under five that is not their younger brother or sister, or their own child, but is still related to them in some as a proxy for siblings. These children could also be a young cousin, for example.

while the decreased supervision may allow brothers of teen mothers to encounter the juvenile justice system.

While not the primary contribution of this paper, I am also able to apply the same analytical strategy to add to the literature on the effect of teen pregnancy on mothers' own outcomes, though remaining selection issues are arguably more important in the own-outcomes case than they are in the sibling spillover case. Relative to female students who had been on a similar trajectory, following the birth the teen mothers display a marked decrease in test scores, an increase in grade repetition and high school dropout, and a decrease in college attendance and college graduation. Unlike their siblings, whose test scores begin to drop relative to matched comparators after the birth of the new child, teen mothers' relative test scores begin to drop in the year prior to the arrival of the baby.

The present analysis provides important evidence on the family-wide effects of teen pregnancy as well as evidence of the important role that siblings can have on each other. Current estimates of the costs of teen childbearing that do not account for siblings may underestimate the true value of successful intervention programs, and the evidence of sibling spillover may also offer a warning to researchers using "sibling fixed effects" to study various outcomes. If a shock affects one child directly it also has the potential to alter the paths of her siblings; if the effect moves in the same direction then sibling fixed effects would understate the true magnitude of the effects. I show that sibling fixed effects models produce lower estimates for the teen mothers than I find in my preferred analysis.

Data

Data come from an anonymous large Florida school district's administrative files for the 1989-1990 through 2004-2005 school years (henceforth, 1990 through 2005). Data are limited to one large county in Florida for students in families with at least two siblings, defined as co-resident children who share the same last name. The year of birth for the students ranges from 1974 to 1993.

I identify teen mothers in two ways. First, in the 2005 school year the district identified the school ID of parents if they also attended school in the district. The mother's school data is then connected to the child's date of birth, which is used to calculate the mother's age at birth (the "birthday method"). Mothers are identified with this first method even if they dropped out of the public school system, but if their children enrolled in school in the county in 2005. This method does not identify most births before 1989, because most of the children would have graduated by 2005. This is not a concern in the analysis, as I do not have the necessary pre-scores for those mothers anyway. The youngest children identified by this method were born in 2003, as they had entered a public pre-K program by 2005.

Second, until 2003 the district identified when students become mothers as long as they remained enrolled in public school in the county (the "district method"). This second method misses any teenage mothers who dropped out of school. Data were not reported for 2002, but limiting the analysis to mothers identified in 2001 and prior does not change the results. Combined, these methods identify those who became teen mothers until 2003, though the 2004-2005 data are retained to examine outcomes after the birth. I can combine the data in multiple ways: method 1 only, method 2 only, privileging the information in method 1 over method 2 (as in the main analysis), or privileging the information in method 2 over 1. Choice of method does not substantively change the results.

Last names and shared address at first observation identify siblings. The year of entry into teen motherhood is also the year that teen mothers' siblings became the aunts/uncles to teenage-parented children.[40] It is both the teen mothers and the younger siblings of teenage mothers who comprise my main groups of interest in the present study. Throughout the paper, I refer to the children of teenage parents as children and the siblings of teenage mothers (who are the aunts and uncles of the children) as siblings. Grandparents are the parents of the teenage mother. Students include teenage mothers, their siblings, and their classmates; the present analysis does not focus on the academic outcomes of the children of teenage mothers.

The identification of teenage mothers means that I under-identify teenage mothers (and their siblings) if a teen mother both dropped out of public school *and* her child did not attend the same school district – or if she both dropped out *and* her child attended the same school district but not in 2005. The population of interest may then be positively selected (if it misses teenage mothers who drop out or if the child attending an alternative school district implies negative selection) or negatively selected (if the child attending an alternative school district implies positive selection). Under-identification of the teen mothers could attenuate the results, leading to an under-estimate of the true effect size. Under-identification may also affect external validity.

[40] The test score data used in this analysis occurs annually, allowing a year-by-year comparison of outcomes as teenage mothers move through the pre-period, pregnancy, birth, and a post-period. Year t=0 is the first academic year that the baby appears in the home – but a portion of that year also occurs before the birth, when the teen mother-to-be is still pregnant. Similarly, t=-1 may contain almost all of the pregnancy (if the birth occurs at the beginning of the year) or none of it (if the birth occurs at the end of the year and year t=0 contains all of the pregnancy).

However, both methods of identifying mothers produce similar results, increasing confidence in the estimates.

There are several outcomes of interest. The first, most immediate outcome is the nationally norm-referenced individual-level scores on the annual California Test of Basic Skills (CTBS) and later the Stanford Achievement Test in math and reading. Tested grades differed by year and ranged from grades 1 through 10. Students took the tests in the spring of grades 3 through 8 in all years, and testing also occurred in grade 1 in 1990, grade 2 in 1990-1992, and grades 9-10 in 2000-2005. The data are reported on a 1-100 scale, representing the student's rank in the national distribution of test scores in each subject.[41] About 4.8% of student-test years are missing test data that ranks them on a national percentile scale, but do have data from the Florida Comprehensive Assessment Test (FCAT). For each grade and subject (math and reading), I impute national percentile ranking in the cases in which these national rankings are missing by regressing, for the cases in which I observe both tests, the national percentile rank on a cubic function of the FCAT. I then use these estimates to predict the estimated national percentile rank for those years missing data. That said, excluding these imputed years does not change the magnitude of the estimates. For brevity, I combine the math and reading scores to estimate the average percentile rank for each student in each year. Analyzing the data separately generally produces smaller results in reading and larger estimates in math, with both estimates going in the same direction as in the combined estimates.

The analysis also examines longer-term outcomes, including whether the student repeats a grade in at least one of the years following the birth, whether the student drops out of school

[41] There are also some students in Grade 11 in the data. Tests are given by student grade, not test grade, and some eleventh graders may have been required to retake the test. The analyses limit data to grades where policy indicated students should be tested, which in practice means dropping some scores in grades 1-2 during untested years.

after the birth, and whether the student first encounters the juvenile justice system after the birth.[42] Testing did not occur in every grade, so the number of observations is lower in the test score analysis than in the longer-term outcomes, particularly among mothers who are often older than tenth grade when they give birth. The National Student Clearinghouse (NSC) provides college-going data for the subsample of students expected to be in college 1997-2006 (about 60% of the data). Of these, about 38% did not appear in the college-going data, 59% are in the NSC data, and 2% blocked detailed reporting. I create three indicator variables using this data: ever attended any college, obtained any degree or certificate, and obtained a four-year degree.

Table 2-1 presents descriptive statistics. Relative to families without a teenage mother, both teenage mothers and their siblings are more likely to be eligible for the free- and reduced-price lunch (FRL) program, identify as black, have lower first-observed test scores, and attend schools with a higher proportion of these characteristics in their first observation in the data.[43] The average teen mother gave birth to the child at age 16.3 (range: 15-17), and her younger siblings were on average 13.1 at the time. All mothers are female, and only younger siblings are included in the sibling analysis.

Family trajectories

Kearney and Levine (2012) argue that teenage childbearing is a consequence of low economic prospects. Such prospects are not stationary over years, and this section investigates whether families that eventually have a teenage birth are on a stable or a downward trajectory

[42] Grade repetition is equal to one if the prior year grade equals the current grade. Dropout is equal to one if the student does not appear in a grade when they would be expected to appear. Juvenile justice exposure equals one if the district data indicated the student was sent through the county juvenile justice system. The age of majority in Florida is 18 (Interstate Compact on the Placement of Children, 2017).

[43] First observation is the first test score observation in the data, which may be different grades for different students. The modal grade is grade 3.

prior to the birth. This distinction matters in teenage pregnancy analysis: if the families are on a downward trajectory then a simple difference-and-difference approach would overstate the effect of teenage pregnancy on teenage mothers and their siblings. Indeed, this intuition of a differential trajectory is the reasoning behind the instrumental variables and other identification strategies pursued in the teenage mother literature (Ashcraft et al., 2013; Fletcher, 2011; Fletcher & Wolfe, 2009; Geronimus & Korenman, 1992; Hotz et al., 2005, 1997).

Figure 2-1 displays the trajectories of teen mothers and their siblings. Panel A includes individual fixed effects, and Panel B adds age fixed effects. Students from non-childbearing families are included to estimate the age fixed effects, which should capture any countywide patterns in test scores.[44] The figure shows that, after controlling for typical scores at a given age, there is a strong downward pattern in both teen mothers and their siblings. The scores begin dropping well before the year of the birth (t=0) or even pregnancy (around t=-1). Note that t=-5 should be taken with some caution for siblings, because many siblings were not old enough to take a test in that year (see Figure 2-2, which displays the percent of observations with test score data by year relative to birth for siblings and teen mothers). Conducting the analysis for years -4 to -1 does not change the overall results, and Table A2-1 in the Appendix shows that adding

[44] The analysis estimates the following model for both mothers and their siblings in Panel B:

$$Y_{it} = \beta_0 + \beta_{1,t} \sum_{-5}^{-1} Birth_{\text{year}_t=\text{t}} + \gamma_{\text{i}} + \mu_t + \varepsilon_{it}$$

where Y_{it} is the test score for individual i in year t, $Birth_{year\ t}$ is an indicator variable for each year relative to the birth of the first child in the family (t=-5 to -1, with t= 0 being the year of birth), γ_i is an individual fixed effect that controls for anything that is constant *within that individual* over the period, and μ_t is an age fixed effect that accounts for any constant differences in outcomes by age in the county. All students without a teen birth in the family form the control group in this analysis. Because there are student fixed effects, the purpose of these control students is to estimate the age fixed effects, as $Birth_{year\ t}$ =0 for all control observations. Standard errors are clustered by family ID.

school fixed effects does not change this pattern.[45] Overall, Figure 2-1 provides support for the theory that families who will eventually experience a teen pregnancy differ from families without teen pregnancy, where scores tend to be relatively flat over time, on average.

Analytic method

Research on teen pregnancy often compares teenage mothers to girls in families without a teenage mother. Selection into teen pregnancy means that such research may not provide reliable results: the mothers and siblings from the sorts of families likely to contain teen mothers are disadvantaged relative to students from the sorts of families unlikely to contain teen mothers, even without a baby in the home. If research does not account for pre-existing trends, it can falsely create the appearance of a causal effect of the birth. In other words, some other change may have led to both the teenage motherhood and lower outcomes in the mother and her siblings. For instance, consider a family without obvious problems and typical academic achievement in Year 1. Perhaps job loss caused intra-family conflict in Year 2, leading to a drop in academic performance. By Year 3 one of the sisters had also given birth as a teenage mother, as her siblings continued to struggle in school. Comparing the average outcomes in Year 1 and 2 (before the birth) to Year 3 (after the birth) would falsely attribute the drop in *average* scores to the birth, when it was actually the job loss and family conflict that led to worsening performance before the birth occurred.

The primary contribution of this paper tests whether teen birth changes the trajectory of the siblings of teen mothers. I create several matched control groups to estimate whether observably similar students diverge after the eventual teen mother gives birth. The first set of

[45] I do not include school fixed effects in the main analysis because changing schools itself might be a signal of diverging trajectories, and I do not want to control for contemporaneous changes.

control analyses use the student's age; family size (as measured by the number of siblings in the data); first-observed individual test score; first-observed school characteristics for percent black, percent FRL, and first test score; and first-observed individual indicators for identification as female, FRL, and black.[46] The siblings and their matched controls are not allowed to be the oldest siblings in the family, both because theoretically there is a stronger influence from older to younger siblings (Monstad et al., 2011) and because practically older siblings are unlikely to have the necessary test score data in the years leading up to the birth. Each sibling of a teen mother is only included once in the creation of the matched control; age is entered as his or her age at t=-1 (one year before birth). Each sibling is matched to five control students, who each receive a weight of 0.2 per match. Potential matched controls are included in each available year. The process allows the same control students to be used more than once for different siblings, possibly at different ages, but each control can only be matched to a given sibling once.[47] The control group produced by this procedure is referred to as the first-observed matched control. Some of the variables used in this procedure (e.g., whether a student identifies as black) are fairly stable over time, while some of them (e.g., test scores) vary.

In addition to the first-observed matched control, a trajectory-matched control adds prior-year trends to the algorithm. In both the sibling and teen mother analysis, the matched control

[46] Data on Hispanic students are not included as it could reveal the anonymous county though including data on Hispanicity at the individual or school level does not change the results. School average first-observed test scores provides an estimate of the mean first-observed test score in the first grade observed for the students (with a mode of grade 3) in the school, regardless of the level of the school. For instance, a higher mean first-observed test score in a middle school indicates that the students in that middle school had high scores in their first test in elementary school.

[47] Specifically, I use a logit model to predict the probability of being the sibling of a teenage mother based on these characteristics. The results of this analysis are available in Table A2-2 in the Appendix. I remove variables to minimize the Akaike information criterion (AIC); this is the model I use in subsequent analysis. Based on this probability, I select the five nearest control matches for each sibling of a teen mother, allowing the same controls to be used for different siblings at various ages, though each control can only be matched to a given sibling once. This requirement prevents a control very similar to a given sibling from being matched to the sibling at ages 13 and 14, for instance.

analysis uses the prior years that have over half of the available data. For the siblings of teen mothers, a majority of siblings have pre-birth test scores in years t=-4, -3, -2, and -1 (see Figure 2-2).[48] However, in theory the pregnancy itself in year t=-1 may affect outcomes, so the trajectory matching is limited to years t=-4, -3, and -2.

An additional condition in some analyses requires that the matched control be from the same micro-neighborhood as the siblings at the first observation in the data. Micro-neighborhoods are small areas, similar to block groups, identified by the county in the data. These are small areas, with an average of 103 students per neighborhood per year.

This analysis produces three trajectory matched control groups: scores only, scores plus observable controls, and scores plus controls plus a requirement that the matched control be from the same micro-neighborhood as the siblings. Table A2-2 in the Appendix displays the results of the logit models.[49]

Table 2-2 displays descriptive statistics for the siblings of teenage mothers and their matched control groups. The means for the siblings in Column 1 differ slightly from Table 2-1 because Table 2-2 adds the requirement that the siblings have at least two of the prior three test scores used in the matched matching procedures (the mean is 2.1 out of 3 observations; requiring all three prior observations increases the standard error but does not change the magnitude or direction of the results). Most characteristics of the five matched control types are similar to the

[48] Students must have at least two of the three years of prior data. For years missing data, the closest available prior year approximates the missing data. Missing test score data are replaced only in the creation of the matched controls. Indicators for missing data in each year are included in the logit model. First-observed matched controls must also have at least two of the prior three years of data to make the results comparable. Similar analysis by alternative timeframes or more or fewer years of data required yield qualitatively similar results.

[49] The preferred trajectory model predicts that siblings of teen mothers are more likely to be female, younger, come from larger families, identify as FRL, identify as black, and have first attended schools with more FRL students and higher first-observed test scores, holding other factors constant. The siblings of teen mothers are also less likely to be missing prior test score data, though both the siblings and the potential controls are required to have at least two of the three prior scores.

siblings of teen mothers in Column 1, but the prior test scores highlight the difference between the first-observed (Columns 2-3) and trajectory methods (Columns 4-6). Examining the columns with scores and controls (Columns 2 and 5), both methods produce control groups that are similar to the siblings at t=-4. At t=-2, however, the siblings and trajectory controls have lower scores than at t=-4, while the first-observed matches remain higher.

The final column requires the trajectory controls to be from the same micro-neighborhood as the teen mothers at first observation. Such a requirement may make unobservable local conditions equal between the siblings and controls, which should reduce unobserved bias (Cook, Shadish, & Wong, 2008). Indeed, Fletcher and Wolfe (2009) show that it is necessary to account for community factors in research designs that rely on miscarriages to create variation in teen parenthood, implying that neighborhoods can affect pregnancy outcomes in ways not picked up by other control variables. However, these local neighborhoods are small, and it could in theory be difficult to perfectly match the students on the observable differences within the pool of potential controls from the same neighborhood. In this final column, the matched controls are younger at t=0 and from slightly smaller families. In the second-to-last column, the matched controls without the neighborhood requirement are similar across all observable characteristics. The main analysis presents results from multiple matched control groups to show how different groups change the estimates.[50]

Using these matched control groups, the main analysis examines several outcomes of interest. Most of the analysis uses ordinary least squares (OLS) regression, as follows:

[50] Column 4, which only uses test scores in the trajectory match, is quite different on all observable characteristics except test scores, and estimates from this analysis should be interpreted with caution.

$$Y_i = \beta_0 + \beta_1 TeenMomSibling_i + X_{\mathrm{i}}\alpha + \varepsilon_i \tag{2-1}$$

where $TeenMomSibling_i$ is an indicator variable equal to one if the student is the sibling of a teen mom, X_i includes the characteristics used in the matching procedure from above, and ε_i is an error term with a mean of zero. The outcomes Y_i are observed once in the data (after the birth). For the siblings of teen mothers, the outcomes examined are the test score in the year of the birth (at t=0), whether the student repeats a grade in at least one of the years following the birth, whether the student drops out after the birth, whether the student encounters the juvenile justice system after the birth, college-going, and whether the student obtains any post-secondary certification/degree or a four-year college degree. An additional analysis includes all years of test score data except the year before the birth, which allows the inclusion of age and individual fixed effects, as follows:

$$Y_{it} = \beta_0 + \beta_1 TeenMomSibling_i \text{ X } PostBirth_t + \gamma_{\mathrm{i}} + \mu_t + \varepsilon_{it} \tag{2-2}$$

where $TeenMomSibling_i X PostBirth_t$ is equal to one after the birth for the siblings of teen mothers, γ_i is an individual fixed effect, and μ_t is an age fixed effect. The year before birth is excluded because pregnancy itself may be a treatment on the siblings. Thus, β_1 provides an estimate of whether the siblings of teen mothers' average scores diverge from what would be expected from a similar control student of the same age, after accounting for individual fixed effects. The age fixed effects capture year-over-year patterns in the population as a whole. This estimate provides an estimate of whether any test score effects persist over time after the birth.

A complementary analysis uses a similar strategy to examine whether the teen birth also changes the trajectories of mothers. The teen mother estimates are interesting by themselves, but they also provide a useful check on the causality of the sibling analysis. If it is family disruption caused by the appearance of a child that changes the trajectories of the siblings of teen mothers, we would expect that pregnancy itself might affect the mothers but not the siblings. Moreover, if a divergence occurs at different times, it is unlikely that some common external event led to pregnancy and drops in scores in the family overall.[51]

Results

The analysis begins by examining whether there is negative spillover from the teen birth to the siblings of the teen mother. That is, given that students are on a similar trajectory, do the paths of the siblings of teenage mothers diverge after the birth? Figure 2-3 displays the test score patterns for siblings, their first-observed matches, and their trajectory matches. The first panel does not require the matched controls to be from the same neighborhood at first observation; the second panel does. Each line displays the coefficient of a regression of national percentile rank on years relative to birth (*t*=-4 through *t*=+1) within the noted combined treatment and control population, holding individual and age fixed effects constant. The vertical line marks the end of the years used in the trajectory matches. To be included in the graph, the students had to have test scores observed both before the pregnancy and after the birth. The solid black (sibling) and gray (trajectory match) lines summarize the primary analytic strategy. While the lines move

[51] This analysis is limited to females, and the trajectory model adds t=-5 because teen mothers, who are older, have more pre-trend data. Using these matched control groups, the main analysis uses ordinary least squares OLS to estimate:

$$Y_i = \beta_0 + \beta_1 TeenMom_i + X_i\alpha + \varepsilon_i$$

where $TeenMom_i$ is an indicator variable equal to one if the student is a teen mom (at ages 15-17), X_i includes the characteristics used in the matching procedure from above, and ε_i is an error term with a mean of zero. The outcomes Y_i are observed once in the data (after the birth) and include the test score in the year of the birth (at t=0), whether the student repeats a grade in at least one of the years following the birth, whether the student drops out after the birth, whether the student encounters the juvenile justice system after the birth, college-going, and whether the student obtains any post-secondary certification/degree or a four-year college degree.

together in the four years before birth, there is a divergence after the birth, with the siblings continuing their decline in test scores and the controls leveling out.

Figure 2-3 also highlights the importance of using trajectory matches. After the birth at t=0, estimated difference between siblings and the first-observed matches are larger than the gap between siblings and their trajectory matches, but this divergence began well before the birth and cannot be caused by the new baby in the home. Overall, this figure provides evidence that there may be some negative spillover from teen motherhood, but it is less than what would be estimated from an analysis that did not account for prior trends.

The figure also displays how requiring the matched controls to be from the same neighborhood changes the estimates. This requirement limits the number of potential matches, which leads to matches that are less similar to the siblings on some observable characteristics. However, they may be more similar on unobservable characteristics, and notably the first-observed match without the neighborhood control (Panel A) has increasing scores over time. Given the low baseline scores for this population, this represents a regression to the mean, and it is a pattern that does not occur once the matches are required to be from the same neighborhood in Panel B. Overall, the post-pregnancy gap between the matched control and the siblings appears to be smaller when the matches are from the same neighborhood in Panel B, particularly for the first-observed match.

More formally, Table 2-3 shows the estimated differences in outcomes between siblings and various matched control options. Column 1 uses all children from non-childbearing families, while Column 2 uses the baseline match with neighborhood requirement. Column 3 is the trajectory match based only on the test scores, Column 4 uses the matches based on the additional observable controls, and the preferred matches in Column 5 include the neighborhood requirement. All estimates control for the same variables; it is the matched control group that changes. In the high school outcomes, the estimates in the first two columns are generally larger than those in the final three columns that use the trajectory-based matched control groups. For the test score in the year immediately following birth, the first column indicates that siblings score 6.2 percentile points lower, relative to all students (in all years) holding first-observed test score, demographics, and school characteristics constant (p-value<0.001). Limiting the comparison to first-observed matches, but still controlling for first-observed test score, demographics, and school characteristics, indicates a difference of 6.5 percentile points (p-value<0.001). However, these estimates are likely biased, as from the figures above it is clear that the families that eventually experience teen pregnancy are on a downward trajectory well before the birth. In the final two columns, the estimate is -4.0 percentile points for the trajectory plus controls match (p-value<0.01) and -4.0 percentile points in the trajectory plus neighborhood requirement match (p-value<0.05). The coefficient in Column 5 is 36% smaller than the coefficient in Column 1. Note that this drop-off is relative to a mean score at the 44th percentile for the preferred control group (see Table A2-3 in the Appendix for the outcome means for the preferred control group and subgroups).[52]

The fixed effect estimate in Row 2 can be interpreted as the average decrease in scores post-birth within siblings relative to their pre-birth average scores, holding individual and age

[52] Table A2-4 in the Appendix includes alternative outcomes. Converting the percentile scores into Z-scores with a mean of zero (at the 50th percentile) and a standard deviation of one results in estimated test score decreases of 0.127 SDs for the siblings of teen mothers in t=0. Dropping the observations that were imputed using FCAT scores and looking at math and reading separately results in largely similar estimates to the main results. If anything, math scores are larger than reading scores, though not statistically significantly so. Looking at tenth grade test scores rather than scores at t=0, for those whose sisters gave birth in grade 10 or earlier, results in similar coefficient sizes but larger standard errors, given that many tenth graders did not have test scores.

effects constant, with the various control groups contributing to the age coefficients. The estimated coefficients are again slightly larger for the models that do not account for trajectories, with an estimate of -3.5. percentile points (*p*-value<0.001) in the first column compared to -2.6 percentile points (*p*-value<0.05) in the final column, a 27% decrease.[53]

Rows 3-5 examine other high school outcomes. The magnitude of the coefficient for grade repetition declines across the columns, with no statistical effect by the final column, meaning that after accounting for test score trajectories, observable characteristics, and neighborhood characteristics, I find no effect of the appearance of the baby in the home. There is an increase in the probability of dropout (7.3 percentage points, *p*-value<0.05) and entering the juvenile justice system (5.8 percentage points, *p*-value<0.05) for the siblings of teen mothers in the final preferred column. For perspective, the control group has a mean dropout rate of 13.3% and a mean juvenile justice exposure rate of 7.8% after the match, so these estimates represent large percentage changes.

Finally, rows 6-8 examine post-secondary outcomes. The magnitudes of these results only slightly change when moving across the columns. The probability of ever attending any college decreases by 7.4 points though this is not statistically significant. The probability of obtaining any sort of degree or certification decreases by 6.7 percentage points (*p*-value<0.10) and obtaining a four-year degree decreases by 7.0 percentage points (*p*-value<0.05), relative to other teenagers from the same neighborhood who had been on a similar trajectory before the

[53] Row 3 of Table A2-4 in the Appendix includes alternative outcomes for the fixed effect model. Converting the percentile scores into Z-scores results in estimated test score decreases of 0.078 SDs for the siblings of teen mothers. Dropping the observations that were imputed using FCAT scores and looking at math and reading separately results in largely similar estimates to the main results.

birth. For reference, in the control population 55% attend any college, 26% obtain any degree or certificate, and 18% obtain a four-year degree.[54]

Dropout rate

The siblings of teenage mothers are more likely to drop out of high school than otherwise similar students from non-childbearing families. Because only those who remain in school have test scores, the comparison groups for test scores may be somewhat biased. Siblings are more likely to drop out than their controls, so this selection means that the previous models underestimate the true effects of teen pregnancy if dropouts would have had lower scores than non-dropouts.

Figure 2-4 displays the dropout rates by year relative to birth. The figure indicates that siblings have an increase in dropouts in years t=0 and 1, relative to both control groups. There is little opportunity for dropout in the earlier years, given that Florida law requires students to remain in school until age 16. This pattern implies that any selection problem should only apply to the oldest students in the population. Moreover, many of the dropouts occur after year t=0, which means that these eventual dropouts should have test scores in the year of the birth.

One method to address this selection is to assume that all those who dropped out of high school would have scored at the bottom of the test score distribution in the years they were missing scores. With this method, the estimated effect is a 3.8 percentile point drop for the siblings of teen mothers (p-value<0.05, see Table A2-4 in the Appendix). Alternatively, I can run a quantile regression to find the median estimated coefficient, which should be more robust to

---

[54] Due to the proximity of a local four-year-college, there is little difference when examining the probability of attending a four-year-college (51% of the control population), so I do not pursue this additional analysis. Note that attending a four-year college does not necessarily mean that a student is pursuing a bachelor's degree.

outliers than OLS methods. Under this method, the effect size is -3.0, has much larger standard errors, and is statistically indistinguishable from zero or the estimate in the preferred regression.

Overall, differences in dropout rates have minimal effects on the test score predictions in the analysis.

Patterns for siblings by sex, race, and FRL status

An additional concern with the results might be that they are driven by a particular subgroup; for instance, perhaps lower-income mothers have fewer supports. Alternatively, perhaps low-income and black communities, which have a higher prevalence of multi-generational families, offer support to handle teen childbearing (Burton, 1999; Fuller-Thomson, Minkler, & Driver, 1997). Sisters may be expected to help with children more than brothers (East, 1998), and a reduction in monitoring by the new grandparents (East, 1999) may affect the sexes differently. This section explores patterns by sex, race (black versus non-black), and FRL status (see Table 2-4). To be conservative, these estimates include the neighborhood requirement, as this requirement may account for unobserved neighborhood differences. The overall estimate from Table 2-3 is included in Column 1 for reference.

Sisters consistently have statistically significant test score effects, while the estimates for brothers are indistinguishable from zero, though the difference between the groups is never statistically significant. At t=0, the estimated test score effect is -5.4 for sisters (p-value<0.01) and -1.1 for bothers (p-value>0.10; p-value of difference between estimates=0.172). For dropout, girls experience larger effects than boys (9.5 versus 4.7 percentage points, p-value of difference=0.433). Conversely, boys whose sisters give birth are 9.7 percentage points more likely to enter the juvenile justice system after the birth, compared to boys on a similar trajectory pre-birth (*p*-value<0.05). There is no effect for girls, though again the difference between the estimates is not statistically significant (*p*-value of difference of estimates=0.217). Note that these changes occur on different margins, as 10.3% of boys in the matched control group are ever exposed to this system, compared to 5.8% of girls. In the longer term, girls experience larger effects on college attendance than boys (-8.8 versus -4.2 percentage points, *p*-value of difference=0.627), on obtaining any degree or certificate (-9.6 versus -2.7 percentage points, *p*-value of difference=0.358), and of obtaining a four-year degree (-8.6 versus -4.3 percentage points, *p*-value of difference=0.485). The standard errors are quite large on these estimates, but consistently point to larger academic effects for sisters than brothers of teen mothers.

The estimated effects on high school outcomes are similar in the black versus non-black comparison. For post-secondary outcomes, the college-going effects are larger for the black students, though the difference is never statistically significant. The estimated effects on test scores are also indistinguishable in the FRL versus non-FRL comparison, though the non-FRL students consistently have larger effect sizes.

Patterns by baseline test scores

Rather than breaking the analysis into binary demographic categories, I can also examine the results by baseline test score distribution. In Figure 2-5, the horizontal axes are the test scores at the first observation in the data, and the vertical axes are the various post-birth outcomes for the siblings and their trajectory-based matched controls. The gap in the dashed horizontal lines is the unadjusted mean difference between these groups. The gray vertical lines divide the baseline scores into tertiles. From this analysis, it appears that the difference in test scores between the treated and control teenagers are evenly spread across the baseline scores, while there is a difference in post-birth grade repetition only on the lower end of the distribution. Dropout is larger in the middle and top tertiles, while the largest differences in the college outcomes are in the middle group.

More formally, Table 2-5 displays the estimated results by tertile of the baseline scores. The regression follows the preferred analysis run separately for each tertile. As expected from Figure 2-5, there is no difference across tertiles for test scores. There is a difference in grade repetition across tertiles (a Hausman test of the difference on the coefficients has a *p*-value of 0.008), with siblings in the lowest tertile having a 15.2 percentage point increase relative to their matches in the lowest tertile. These lower-tertile students may be closer to the margin of grade repetition. Dropout and juvenile justice exposure do not differ across groups. Few (36.4%) of the low-tertile and many (69.9%) of the high-tertile students attend college, on average, but there is a large difference in college-going and completion by treatment status in the middle tertile. For these middle-tertile, marginal students, there is a 25.3 percentage point effect size for attending college (*p*-value of Hausman test across tertiles=0.005), a 16.7 percentage point effect size for obtaining a degree or certification (*p*-value of Hausman test=0.036), and a 12.6 percentage point effect size in obtaining a four-year degree (*p*-value of Hausman test=0.125).

Timing of the effect

The biggest concern in the present analysis is that some external event led to both the pregnancy and a downward drop in scores for the whole family. While the pregnancy itself is likely to affect the teen mother, it will not necessarily affect the siblings. Thus, one way to test whether some external event might have led to the drop in scores involves testing whether test scores drop in *t*=-1 (in the time leading up to and during pregnancy) or continue on the same path.

A formal test of parallel trajectories runs a fixed effects analysis by year relative to birth and interacts each pre-birth year with an indicator for being a sibling (with age fixed effects to match the previous analysis). There is no statistical difference in the interaction between sibling and the years leading up to year *t*=-2, confirming the efficacy of the match. In *t*=-1, which is not restricted to be the same between the siblings and their matched controls, the coefficient on the interaction term is small (-1.1 percentile points) and does not statistically differ from zero (*p*-value=0.541). This indicates that the divergence between siblings and their matched controls only occurs after the appearance of a child in year *t*=0.

Given that the divergence occurs in *t*=0, it is also possible to estimate the matching years by different timeframes. Figure 2-6 displays alternative trajectory specifications, with the requirement that to be included the data must have post-birth data in addition to the data required to be included in the matching in the first place. Adding the year of pregnancy (*t*=-4 to -1, Panel A), shifting the match by a year (*t*=-3 to -1, Panel B), or adding another pre-pregnancy year (*t*=-5 to -2, Panel C) in the trajectory procedure does not change the overall pattern. Panel C should be taken with some caution given the low number of participants with data in *t*=-5 (see Figure 2-2). Limiting the match to just two years (*t*=-4 to -3, Panel D) results in similar patterns in years *t*=-2 and *t*=-1, despite not being constrained to move together by the algorithm.

Teen mother analysis

Table A2-5 in the Appendix displays logit models predicting the probability of becoming a teen mother.[55] Table 2-6 displays descriptive statistics for the teen mothers and their matched control groups. Most of the five matched control types are quite similar to the teen mothers, though the teen mothers are somewhat more likely to be older and be the oldest sibling in the data. As in the sibling analysis, the prior test scores highlight the difference between the first-observed and trajectory methods. Both methods produce control groups that are similar to the teen mothers at t=-5. At t=-2, however, the teen mothers and trajectory controls generally have lower scores than at t=-5, while the first-observed matches are higher. Notably, the neighborhood-restricted match is also relatively high in this year, though the difference from the teen mothers is not statistically significant. The neighborhood requirement may reduce unobserved bias (Cook et al., 2008), but it can also make matching on the smaller pool of comparison students more difficult, as evidenced by the slightly smaller number of matched controls in the final column, relative to the second-to-last column.

The families of teenage mothers are on a downward trajectory relative to average families (see Figure 2-1). However, the downward pattern exhibited by the teen mothers could change following the pregnancy or the birth of the child. Figure 2-7 displays the test score patterns for teen mothers, their first-observed matches, and their trajectory matches. The first panel does not require the matched controls to be from the same neighborhood at first observation; the second adds that requirement. Each line displays the coefficients from regressions of national percentile rank on years relative to birth (t=-5 through t=0) within the noted combined treatment and

[55] The preferred trajectory model indicates that teen mothers are more likely to be older, be the oldest sibling in the family, identify as FRL and black, and attend schools with fewer black students and lower first-observed test scores. Teen mothers are also less likely to be missing prior test score data.

control population, holding individual and age fixed effects constant. The vertical line marks the end of the years used in the trajectory matches. To be included in the graph, the students had to have test scores observed both before the pregnancy and after the birth. The solid black (teen mothers) and gray (trajectory controls) lines again summarize the analytic strategy. While the lines move together in the years used in the matching, there is a divergence during the year closest to the pregnancy (t=-1), with the teen mothers increasing their decline in test scores and the controls leveling out. The test scores remain low for teen mothers in t=0, the year of the birth.

As in the sibling analysis, requiring the matched controls to be from the same neighborhood results in fewer matches for the mothers. However, the students from the same neighborhood may be more similar on unobservable characteristics. The figure with the neighborhood requirement has less divergence between the lines after the birth, though the lines are somewhat more erratic.

Table 2-7 examines the outcomes for the teen mothers. The test score estimates replicate what would be predicted from Figure 2-7. The naïve estimates in Column 1 indicate that teen mothers have test scores 7.4 percentile points lower than all female students from non-childbearing families in the year of birth, after controlling for all observable characteristics from Table 2-1 (p-value<0.001). The gap is 7.1 percentile points when the control population is limited to the matched control group based on first-observed characteristics (and still controlling for observable variables; p-value<0.001). However, these estimates may be biased, as the analysis shows that the families that eventually have teen childbearing are on a downward trajectory well before the birth. The estimate shrinks once the match is instead restricted to the trajectory-based matched control, which matches students who are on a similar downward trajectory in year t=-2 and prior. For the preferred estimate in Column 5, teen mothers achieve test scores 4.4 percentile points lower than students who do not give birth but who were previously on a similar trajectory (p-value<0.051). This is a 41% reduction in the estimated effect size compare to Column 1.

Rows 2 and 3 examine grade repetition and dropout. All methods indicate that teenage pregnancy is associated with large increases in the probability of repeating a grade and dropping out post-birth for teenage mothers. The probability of grade repetition is 17.2 percentage points higher relative to the preferred trajectory-based matched control group (p-value<0.001), while the probability of dropping out is 24.1 percentage points higher (p-value<0.001). There is a small decrease in the probability of exposure to the juvenile justice system in the naïve estimate (-1.7 percentage points, p-value<0.01), but this disappears in the matched estimates with controls.

Turning to the college-going data, teen mothers are much less likely to attend any college, relative to students on a similar trajectory without a birth in the family (-18.6 percentage points, p-value<0.001). Mothers are also less likely to obtain any degree or certificate (-12.5 percentage points, p-value<0.001) or obtain at least a four-year degree (-11.7 percentage points, p-value<0.001).

Next, I test whether the divergence in scores between the teen mothers and their matched controls starts in tests taken before the birth. Using the same parallel trends analysis from the sibling analysis, the scores for the years used in the matching process (t=-5 to -2) are statistically indistinguishable. The teen mothers score 4.8 percentile points lower than their matched controls at t=-1 (p-value<0.01). Recall that the siblings' scores do not drop until the child appears in the home (t=0).

Overall, this analysis indicates that teen motherhood is associated with poor academic outcomes. For completeness, A2-6 in the Appendix also conducts the analysis by subgroup for black, non-black, FRL, and non-FRL students. The effects are large and robust across subgroups in both the high school and college-going years. The non-FRL teen mothers have particularly large effect sizes, though the differences between the FRL and non-FRL effects are never statistically significant.

Instrumenting for method of identification of teen mothers

Teen mothers are identified in two ways: by the school district or using birth dates. One worry is that some teens are missing from the analysis, and what I use as a control group may actually contain some teenagers who are really siblings of teen mothers that I did not identify. This could attenuate my results. One way to identify this is to use one measure of teen births (using the birth date method) as an instrumental variable (IV) to estimate the other measure (using the school district method). I can use this method for both the teen mothers and their siblings. For the siblings, this analysis estimates the effect for the siblings whose sister's childbirth was identified by the district method, scaled up by the predicted increase in district method identification based on identification using the birthday method.[56] This should account for the group of unidentified siblings who had formerly been in the control pool, much like a treatment-on-the-treated analysis in other IV models.

---

[56] Specifically, the model estimates the following set of equations:

$$sib\_district_i = \alpha_0 + \alpha_1(sib\_bday_i) + X_i\theta + u_i$$

$$Y_i = \beta_0 + \beta_1(\widehat{sib\_district_i}) + X_i\varphi + \varepsilon_i$$

where X_i includes the vector of coefficients from the main regressions. The first stage is about 0.5, indicating that if a student is identified as a sibling through the birthday method, their chance of identification through the district method is about 50 percentage points higher.

For the teen mothers, under this method I find effect sizes that are at least twice as large as the preferred estimates and at least marginally statistically significant on test scores in t=0 (β=-9.566, p-value<0.05) and ever attending college (β=-43.497, p-value<0.001). For the siblings of teen mothers, under the IV method I find effect sizes that are at least twice as large as the preferred estimates and at least marginally statistically significant on test scores in t=0 (β=-13.641, p-value<0.001), ever attending college (β=-21.201, p-value<0.05), and obtaining any degree or certificate (β=-14.506, p-value<0.10). The other effects are about the same in both methods, though generally slightly larger.

The effects could be larger for two reasons. First, instrumenting may reduce the measurement error in the number of teenagers who become pregnant, thus increasing the accuracy of this estimate relative to the preferred estimate from the previous tables. Second, the IV measures local average treatment effects (LATE) for those who are predicted to have a birth under the district method using the birthdate method. If those teen mothers who have both measures (that is, do not drop out of school themselves *and* have children who stay in the district until 2005) have larger effect sizes, the IV could overstate the true average treatment effect for both the teen mothers and their siblings. For instance, those siblings whose sisters stay in school and remain in the district may be more advantaged than others, so this second interpretation is plausible if it is the more-advantaged siblings of teen mothers who are most hurt by their sister's childbirth. That interpretation is not supported by the analysis by tertile of baseline test performance, and I instead take the IV results as evidence that my main results may understate the true effects somewhat due to measurement error. I conservatively retain the main results without IV, but note that future estimated effects may be higher with less measurement error.

Mechanisms and checks

There are several reasons that teen motherhood could affect sibling outcomes, particularly given the differences observed by sex. Below, I examine several possibilities, including that the effects for the sisters of teen moms are driven by girls who become teen mothers themselves, differences in time allocation of the siblings of the teen mothers, differences in family composition, and whether the effects occur for older siblings. I also include two placebo tests.

Sisters and teen motherhood

One potential reason that sisters of teen mothers perform worse than either other teenage girls or the brothers of teen mothers is that these girls are more likely to become teen mothers themselves (Monstad et al., 2011), and the education effects are really capturing the effect of these girls becoming teen mothers. I examine this possibility by interacting the indicator variable for siblings with an indicator for whether that sister eventually becomes a teen mother herself. The analysis includes the subgroup of sisters of teen mothers and their female matched controls. The coefficient on the interaction should be interpreted as a descriptive exercise and should be interpreted with caution given the low number of sisters who become teen mothers and have the requisite pre-test data (N=16). However, if the effect for sisters who do not become teen mothers is statistically significant and comparable to the main estimate, it implies that teen motherhood for the sisters does not drive the results.

Panel A of Table 2-8 displays the results of this analysis and indicates that the effects for sisters who do not become teen mothers are comparable to the preferred estimates, with no differences between the effect sizes for sisters who do and do not eventually become teen mothers themselves. The exception is grade repetition, where the estimated effect is null for most sisters but 26.1 percentage points higher for the sisters who later become teen mothers themselves (*p*-value of difference<0.10). This could occur if becoming a teen mother itself increases grade repetition, as indicated by Table 2-7. Beyond grade repetition, sisters who become teen mothers do not appear to be driving the sibling results.

Teen time use

This section uses data from the American Time Use Survey data from 2003 to 2015, which samples individuals ages 15 and older from U.S. families.[57] The data includes details about the primary activity pursued throughout a 24-hour period (beginning at 4:00 a.m.), as well as who is with the respondent during the activities. If respondents report pursuing more than one activity at a time, they must select a "primary" activity. Interviewers also ask respondents about secondary childcare, defined as having a child under age 13 in the respondent's care while doing other activities. I restrict the analysis to 15-, 16-, and 17-year-olds to match the age range used in the main analysis.

The data does not directly identify all teen parents or the siblings of teen parents; instead I rely on the composition of the household of the respondent to approximate these categories. I divide the teenagers into six groups. There are teenage mothers and fathers, defined as females and males, respectively, living with their own children in their household. I exclude teenage fathers from the analysis given the very small number of individuals in this category. There are also the sisters and brothers of teenage mothers. There is no identification for whether you live with your niece or nephew (i.e., the child of your sister); instead, these groups are identified based on whether they live with a child under five years old who is not their sibling, or their own

[57] See American Time Use Survey Technical Note for additional details (Bureau of Labor Statistics, 2016).

child, but is related to them in some other way. Some of these children may be cousins or some other group, but it is the closest available proxy for siblings. If siblings are most likely to provide childcare for their nieces and nephews than some other child, the following analysis may underestimate the true amount of time sisters and brothers spend on childcare. The final two groups are other females and males, defined as 15-, 16-, and 17-year-old respondents who do not fall into one of the other categories.

I pursue two primary analyses: how respondents spend their time (activity analysis), and with whom they spend their time (social analysis). The main purpose of both analyses is to identify potential reasons for the differences in the outcomes between sisters and brothers, framed as a difference-in-difference approach. The difference between the other females and other males is taken as a baseline, and the difference between that gap and that of sisters and brothers of teen mothers is taken as the estimated gap attributed to gendered reactions to a young niece/nephew being in the home. This analysis faces similar challenges as the first-observed matching estimates in the main analysis, and may be biased if there are different gender-by-group trajectories before the observation (i.e., the parallel trends assumption is violated). For instance, perhaps the sisters of teen mothers would have spent time on childcare of some other family member even in the absence of their niece/nephew in the home. I have no way to test this, and the analysis should thus be taken as a descriptive exercise of potential pathways, rather than a causal estimate.

I group activities into five categories. Sleep is the most straightforward. Childcare includes any time spent with childcare as a primary activity such as physical care of children. A supplementary analysis also includes childcare as a *secondary* activity, meaning the teenager may be cooking, eating, watching TV, etc., as a primary activity, but the teenager also reports a child being in their care. School includes both being in school and homework activities, while work includes all work-related activities.[58] The remaining time is grouped into an "other" category that includes eating, socializing as a primary activity, and any other activities not included above. Sleep, childcare (as a primary activity), school, work, and other are mutually exclusive.

The social analysis examines the time spent in various types of social situations: time spent alone, time spent with parents, time spent with friends, time spent without supervision of an adult over 18 years old, and time spent with a child under 13 years old. These categories are not exclusive.

I conduct the analyses for all days, as well as separately by weekdays and weekends. Weekends also include holidays.[59]

Figure 2-8 displays the time spent on various primary activities by the teenagers for weekdays (Panel A) and weekends (Panel B). Due to space constraints, I do not display eight of the hours spent on sleep. Teen mothers spend the most time on childcare, though sisters appear to spend more time on childcare relative to brothers, possibly at the expense of schoolwork.

More formally, Table 2-9 displays the difference-in-difference analysis; it also controls for family size, number of children in the family, race/ethnicity, metropolitan status, and age. Beginning in Panel A, which includes all days in the analysis, none of the primary activities

---

[58] Sleep is defined as code t010101; childcare is defined as codes t0301*, t0302*, t0303*, t0401*, t0402*, and t0403*; any childcare adds time reported spent in secondary childcare for household and non-household children; school is defined as t0601* (class) and t0603* (homework) but not t0602* (extracurricular activities); and work is defined as t05*.

[59] Holidays include New Year's Day, Easter, Memorial Day, the Fourth of July, Labor Day, Thanksgiving Day, and Christmas Day. Data were not collected for New Year's Day in 2012 or Christmas Day in 2011 and 2014.

differ for the siblings of teen mothers. However, when childcare is measured as either a primary or a secondary activity, the gap between sisters and brothers of teen mothers is one and a half hours larger than the gap between the sisters and brothers of other families. Panel B indicates that this difference is larger on weekdays, at over two hours. On the weekends, the childcare difference is over one hour, but it is not statistically significant. The gender gap on homework is an hour in the opposite direction on the weekends, indicating that sisters of teen mothers spend less time on homework.

Table 2-10 examines with whom the students spend their time. Across all days (Panel A), sisters of teen mothers spend less time with their parents and more with children, relative to both brothers of teen mothers and other female teenagers. On weekdays (Panel B), brothers of teen mothers are more likely to be alone by over an hour, less likely to be with parents by half an hour, less likely to be supervised by adults by nearly an hour, and less likely to be with children by about half an hour, relative to other males. The sisters are more likely to be with children relative to both brothers of teen mothers and other females; the coefficient on the interaction is three and a half hours. On the weekends, sisters are less likely to spend time with parents by nearly two hours or with friends by over an hour and a half.

Though the analysis controls for several potential covariates, results should be interpreted with caution if gendered differences in time use in the sorts of families that experience teen childbearing differ from the sorts of families without teen childbearing, even in the absence of a teen birth. Still, the analysis does provide suggestive evidence of substitution from homework to childcare, which may drive the academic results for the sisters of teen mothers. Decreased supervision and increased alone time may allow brothers of teen mothers to encounter the juvenile justice system.

Family composition

If sisters are providing more childcare assistance, then having more younger sisters may benefit both the teenage mother and her sisters if that allows the work to be spread across more family members. Panel B of Table 2-8 examines the outcomes by family composition by adding an indicator variable equal to one if the sibling also has at least one other sibling younger than the teen mom and an indicator variable equal to one if at least one of the fellow younger siblings is a sister. For grade repetition and high school dropout, the effects seem to be largely driven by those siblings with a fellow younger sibling; this may be because the effects are more pronounced in larger families. However, having a younger sister alleviates this pattern. Indeed, for siblings with a fellow younger sister, the effects are null for grade repetition and high school dropout, though given the large standard errors this should be interpreted with some caution. When the analysis is limited to only females in Panel C, there is potential evidence of a protective effect for sisters for grade repetition, high school dropout, and receiving college degrees. Though the coefficients are large, however, they are rarely statistically significant, and I take these as suggestive. I find no evidence that teen mothers benefit academically from having younger sisters.[60]

Older sibling analysis

The main analysis focuses on younger siblings. If the primary driver of the effect is among those who share a home with the new baby, then siblings over eighteen, who are more

---

[60] The main difference by family composition is that teen mothers without any younger siblings in the data have a 3.4 percentage point decrease in juvenile justice exposure relative to other females who had been on a similar trajectory, while there is no effect for teen mothers with younger siblings.

likely to have moved out of the house by the time the baby arrives, should not be affected by the baby. This analysis should be interpreted with caution. The number of respondents is low, as they must both be eighteen or older at the time of the match and have enough years of test data to estimate a trajectory match from their tests. I use the six, seven, and eight years before the birth to estimate the trajectory, as these are the years with at least 50% of the data for the siblings who are 18 or older at the time of the birth.

The first column of Table 2-11 shows the results, estimated only for the college-going outcomes, as the high school outcomes were likely completed before the birth. The coefficients are large, but they also have large standard errors and are inconsistent in direction. None are statistically significant. The analysis provides some support for the hypothesis that being in the home is important for explaining the effect of teen motherhood on the siblings' outcomes, though future research with larger samples sizes should examine older sisters in more detail.

Falsification tests for siblings

I conduct two placebo tests to verify that the matching mechanism is not mechanically creating the estimated effects. The first test looks retroactively at test scores, grade repetition, and juvenile justice exposure before the match, predicting the difference four years before the match. The analysis uses the preferred trajectory match. These estimates should be null, as they are (see Column 2 of Table 2-11).

The second set of estimates randomly selects 200 families from non-childbearing homes who had siblings matched to the siblings of teen mothers in the main analysis. This randomly selected group is treated as a set of placebo siblings, with their placebo "age at birth" set as a random age among their match years if there were multiple years of matching.[61] Each of these placebo younger siblings is then matched to other potential matches from the sample of younger siblings from non-teenage-childbearing families who were not randomly selected to be placebos. The matching procedure is the same as in the main analysis. The results of these estimates are displayed in Column 3 of Table 2-11. Again, these placebo estimates should be null, as indeed they are.

The two placebo estimates indicate that there is no mechanical way that the matching strategy is creating differences among groups. Moreover, the retroactive estimates in Column 2 provide further evidence that the groups are matched well on baseline characteristics.

Discussion

The preceding analysis can be interpreted as causal if accounting for the several years of pre-trend data, plus other baseline characteristics, captures any differences between families with and without teen births. One concern could be that some unobserved event leads to teenagers giving birth. For instance, perhaps parental job loss increases the chance that older females have a child, and parental job loss is also associated with poor outcomes for all siblings. Then, the underlying cause of the poor outcomes would be the job loss, not the teen pregnancy itself. If the appearance of the child, and not general family trends, causes drops in performance, test scores should diverge from prior patterns in the year the child appears (and possibly continue into subsequent years), as seen in Figure 2-3.

The analysis provides support for interpreting the sibling analysis as causal. In the above example, the siblings and the matched controls continued on parallel paths in the year of the pregnancy; it is only after the birth that the paths diverge. Mothers, on the other hand, begin their

[61] Recall that a given potential match can be matched at multiple ages for different siblings in the main analysis.

drop on performance in year t=-1 (see Figure 2-7). If some external event led to both pregnancy and dropping scores in the family, then the drops should occur in the same year. Instead, the scores of teen mothers and their siblings change in different years, and the sibling scores do not change until the baby appears in the family.

In the teen mother analysis, any unobserved differences between mothers and non-mothers are likely to be negatively associated with the outcomes of interest, and thus this analysis provides an upper bound on the negative effects of teen pregnancy on various outcomes. The estimates in the teen mother analysis are larger than prior literature that uses miscarriage as an instrument (e.g., Fletcher & Wolfe, 2009). To some extent, comparing completed pregnancies to miscarriages may understate the true effect of motherhood relative to no pregnancy, given that miscarriage is associated with long-term psychological repercussions such as elevated anxiety and depression (Hunfeld, Wladimiroff, & Passchier, 1997; Lok & Neugebauer, 2007; Prettyman, Cordle, & Cook, 1993). The estimated effects for teen mothers in the present study are likely an upper bound on the true effect, but the 24.1 percentage point increase in high school dropout in my preferred estimate, for instance, is broadly consistent with the 8-9 percentage point decrease in diploma receipt found by Fletcher and Wolfe (2009) if we take that estimate as a minimum bound. My estimates are also in line with work using within-school matching estimators (Levine & Painter, 2003).

Notably, prior work using sibling fixed effects to examine sisters may understate the true effect for teen mothers. For instance, I find a 24.1 percentage point increase in high school dropout for teen mothers and a 9.5 percentage point increase for the sisters of teen mothers. When I conduct a within-family fixed effect estimate with my data, I find an estimated 17.6

percentage point drop for teenage mothers, relative to their non-childbearing sisters.[62] That is approximately the same as the 14.6 percentage point difference in my estimates,[63] demonstrating that sibling fixed effects estimates will understate the true effects in cases of sibling spillover.

Conclusion

The present analysis uses novel data to study several important questions. First, I present new evidence regarding spillover effects of teenage childbearing on the outcomes of the siblings of teen mothers. The mothers' sisters experience the largest academic effects in K-12, while the implications for brothers are larger in terms of juvenile justice exposure. Both have lower college-going and completion. These findings indicate that estimates of the cost of teen pregnancy that only focus on the teen mothers or their children will be understated. Research on long-acting reversible contraceptives found an approximate cost of $13,531 per prevented birth (Lindo & Packham, forthcoming). A one SD increase in student test scores is associated with a 12% increase in earnings at age 28 (Chetty et al., 2014). I take the 0.127 SD effect on siblings here (see Appendix), and assume that the increase to earnings is constant over time and each teen mother has 0.95 younger siblings (the median of the 1.89 average per family with children; U.S. Census Bureau, 2016). Taking $522,000 as the mean present value of earnings at age 12 for the U.S. (Chetty et al., 2014), preventing a teen birth would increase sibling earnings by approximately $7,500, though even an estimated effect of half that amount is still substantial.[64] The $7,500 figure represents over half the cost of a pregnancy prevented from Lindo and Packham, indicating that the benefit to teen pregnancy prevention among siblings alone could be

---

[62] This analysis uses all families with at least one sister who does and one sister who does not give birth as a teenager, comparing within the family using the Stata command *xtreg*.

[63] Calculated as a 24.1 percentage point effect for teen mothers minus the 9.5 percentage point effect for their sisters.

[64] Calculated as $522,000 X 12% per SD X 0.127 SDs per sibling X 0.95 siblings = $7,558.

quite substantial. A full cost-benefit analysis is beyond the scope of the present paper, but should consider the net social and individual benefits of teen childbearing reduction for the teen mother, her children, her sexual partner, and her family members.

Second, I show that analyses using sibling fixed effects will understate the effects of teenage pregnancy. The finding of unexpected family spillover is also a cautionary note to researchers in topics beyond teen pregnancy. Given the popularity of sibling fixed effects in economics and other disciplines, researchers should carefully consider pathways by which the direct policy or change of interest may also have spillover on the family.

Third, the analysis provides evidence that the families of eventual teen mothers are on a downward trajectory well before birth. If teen pregnancy is a symptom of family dynamics in struggling families, then public policy could focus its efforts on targeting and assisting students from these families. Programs already use individual warning signs for dropout prevention programs (e.g., Kennelly & Monrad, 2007), and teen motherhood is associated with behavior also associated with dropout (e.g., poor academic performance, grade repetition). Perhaps programs could incorporate family-wide patterns to identify who to target when attempting to reduce pregnancy and dropout.

Fourth, analysis that does not account for downward trajectories will overestimate the negative effects of teen pregnancy on the teen mothers and their siblings. Still, after accounting for these patterns, this paper finds a decrease in academic performance, an increase in grade repetition and dropout, and decreases in college attainment for the whole family.

Not every pregnancy leads to childbirth. In future research, other types of data sources could shed light on whether similar pre-pregnancy patterns occur for pregnancies with and without associated births.

The present analysis identifies some potential pathways through which siblings are affected by their sister's childbearing. Previous research has indicated that teen pregnancy might be "contagious" among sisters (Monstad et al., 2011). Here I find no evidence that the effects are driven by the sisters of teen mothers that eventually become teen mothers themselves in most estimates; the exception is in grade repetition, where the sisters who later become teen mothers are the only ones who have an increase in repetition. In a supplementary difference-in-difference analysis I study time use using a proxy for sibling childbirth. I show that sisters of teen mothers spend more time on childcare and less on homework than either other females or the brothers of teen mothers, whereas brothers have much more alone time. These differences provide suggestive evidence to explain the difference in academic and juvenile justice outcomes for the sisters and brothers of teen mothers.

Overall, the findings provide evidence that families that experience teen motherhood are on a downward trajectory well before birth, but that teen motherhood has short- and long-term negative effects on teen mothers and their siblings. Teen motherhood is more common among low-income and under-represented minority groups, and having a sister who gives birth may exacerbate problems faced by children in such families. The present research indicates that current estimates understate the true cost of teen pregnancy to families and society.

Study 3. A Day in the Life: The Effects of Acute Violent Crime on Sleep and Biological Stress Levels in Adolescents

Children exposed to recent incidents of local violence perform worse on assessments of academic skills, executive function, and effortful control, relative to children from the same or similar neighborhoods who were not recently exposed to violence (Gershenson & Tekin, 2015; Sharkey, 2010; Sharkey, Tirado-Strayer, Papachristos, & Raver, 2012). Despite evidence that violence has an acute, causal impact on academic performance, the mechanisms explaining how environmental stressors get "under the skin" or "into the minds" of children have received little attention. This study examines two potential causal pathways that might explain the effects of acute violent crime: sleep and the hypothalamic-pituitary-adrenal (HPA) axis.

Sleep is a stress-sensitive system (Hicken, Lee, Ailshire, Burgard, & Williams, 2013; Sadeh, 1996). Under the opponent processes model of sleep, individuals must suspend arousal to transition from wakefulness to sleep (Edgar, Dement, & Fuller, 1993). Crime may affect sleep through heightened vigilance arising from fear or stress, or through disruption from gunfire, sirens, shouting, and expanded police presence (R. E. Dahl, 1996; Hale et al., 2013; Lacoe & Sharkey, 2016). Insufficient sleep is associated with greater fatigue, more difficulty concentrating, and reduced executive functioning (Alapin et al., 2000; Sadeh et al., 2003; Talbot, McGlinchey, Kaplan, Dahl, & Harvey, 2010).

Stress exposure can also affect the HPA axis, a key biological stress response system, and its primary hormonal product, cortisol. Cortisol levels follow a predictable daily pattern. Cortisol builds overnight, and a sharp increase in cortisol about 30 minutes after waking is called the cortisol awakening response, or CAR (Pruessner et al., 1997). Normally, the CAR assists in the transition from sleep to waking and may provide a boost to help meet the expected demands of the upcoming day (Adam, Hawkley, Kudielka, & Cacioppo, 2006; Clow, Hucklebridge, Stalder, Evans, & Thorn, 2010; Fries, Dettenborn, & Kirschbaum, 2009; Vrshek-Schallhorn et al., 2012). Following the CAR, basal activity of the HPA axis typically decreases during waking hours, reaching the lowest levels shortly after sleep begins (Adam et al., 2006).

Stress exposure can change typical diurnal cortisol patterns (Adam, 2012; Sapolsky, Romero, & Munck, 2000). Acute stress is associated with an elevated CAR, with more chronic stress and burnout associated with a blunted CAR (Chida & Steptoe, 2009). High daily stress is associated with low waking cortisol, a low CAR, slower memory speed, and reduced attention continuity (Maldonado et al., 2008).

Although exposure to traumatic events such as hurricanes is associated with long-term changes to sleep and the HPA axis (Lavie, 2001; Vigil, Geary, Granger, & Flinn, 2010), it is unclear whether somewhat more common stressors such as violent crime have similar effects. To my knowledge this is the first study examining the short-term effects of proximity to nearby violent crime on sleep and cortisol.

The primary empirical strategy uses a within-person design. It measures the change in sleep or cortisol on the night/morning immediately following a local violent crime, relative to that same person's typical pattern. Because the comparison occurs within people, any changes to the outcomes cannot be caused by characteristics that are constant for a person (e.g., neighborhood quality). The important assumption in the design is that, within a given person, the timing of violent crime is random with respect to the timing of the diary study. Exogenous variation in the timing of local crime allows for causal inference.

Methods

Participants

The present study involves a sample of 82 adolescents (49% female) who participated in a take-home diary study in a large Midwestern city in fall 2012. Diary subjects were drawn from participants of a larger study of adolescent stress (N=379) submitted through the DePaul University and Northwestern University Institutional Review Boards. Adolescents were recruited through three public schools selected for having racially, ethnically, and socioeconomically diverse student bodies drawn from across the city. Information about the study was presented orally and in writing to potential participants during homeroom or lunch periods at their school. Parents were informed about the study at parent meetings, at report card pick-up days, and through parental consent forms. Participants with parental consent completed assent forms before data collection.

A sub-sample of 138 adolescents participated in the take-home diary study, including 94 adolescents who also received activity-tracking watches. One adolescent did not return a watch and 11 watches contained no usable data, leaving 82 adolescents with at least one full night of sleep data.[65] I obtained 87% of four expected nights of sleep data from the 82 adolescents. The mean age was 14.90 (SD=1.87, range=11.3 to 18.1), with 17% of participants identifying as black (n=14), 20% identifying as Hispanic white (n=16), 18% identifying as non-Hispanic white (n=15), 27% identifying as multi-ethnic (n=22), and the remainder identifying as another

[65] Six participants had only one night of usable sleep data and thus do not contribute to the estimated differences caused by crime. Those participants are retained because they contribute to the cortisol estimates. Their sleep and cortisol measures did not differ from the average of the other participants. Dropping them from the sleep analysis did not change the results, because they do not contribute to the sleep estimates.

category or missing race information. Table 3-1 provides descriptive statistics. Table 3-2 provides a correlation matrix.

Measures

Sleep. Adolescents participated in one of six collection periods in fall 2012. Participants wore an Actiwatch-64 (Mini-Mitter Respironics, Inc.) for four consecutive nights to objectively measure sleep. Each period began on a Saturday. Participants were instructed to register bedtime and wake time by pressing a button on the Actiwatch immediately before they went to bed and immediately after they woke up, respectively. These timestamps were cross-checked using adolescent self-report diaries. The validated Actiware Sleep software (version 3.4, Mini-Mitter/Philips Respironics) calculated various sleep parameters using 1-minute epochs, based on significant movement after at least 10 minutes of inactivity (Oakley, 1997).

The software calculated sleep measures including *bedtime* (time when the adolescent got into bed; mean=11:07 p.m.), *sleep latency* (lag from bedtime to sleep onset; mean=26 minutes), *wake time* (time at which the adolescent woke up for the last time in the morning; mean=7:00 a.m.), and *sleep duration* (time actually spent asleep from sleep onset to final waking; mean=6.5 hours). Adolescents completed a make-up day on Wednesday night if they missed a day of data collection. The adolescents had a mean of 3.4 nights of data out of an expected 4 nights.

Cortisol. Participants collected three salivary cortisol samples (immediately after waking, 30 minutes after waking, and at bedtime) each day for three consecutive days (typically Sunday, Monday, and Tuesday) during the same period as the sleep study. They also collected an evening sample on Saturday and wake and 30 minutes post-awakening samples on the final day (typically Wednesday). Participants used a passive drool technique by which they expressed unstimulated saliva through a small straw into a vial. Participants watched a saliva sample demonstration prior to the study and received reminder calls from the research team during the data collection to ensure they followed the protocol. Participants were instructed to avoid eating, drinking, and brushing their teeth 30 minutes prior to each sample collection. A kitchen timer pre-set to 30 minutes was provided to aid in the timing of the second daily sample. Participants were instructed to refrigerate their samples as soon as possible after collection. Most samples were taken at home and thus immediately refrigerated. Adolescents returned their samples to school at the end of the study week, at which point the research team paid the participants. Six of the 82 sleep participants returned no usable saliva samples, leaving N=76 cortisol participants. Cortisol participants completed 91% of expected samples. Samples were stored at -20°C before shipment to Trier, Germany, where they were assayed in duplicate using time-resolved fluorescent-detection immunoassay (Dressendörfer, Kirschbaum, Rohde, Stahl, & Strasburger, 1992).

Measures included *waking cortisol* (cortisol level at wake; mean=0.266 µg/dl; mean samples per person=3.7 out of an expected 4), *CAR* (increase from waking measurement to the measure occurring about 30 minutes later; mean=0.120 µg/dl; 3.2 of 4), and *bedtime cortisol* (cortisol level at bedtime; mean=0.066 µg/dl; 3.7 of 4). Analyses dropped any CAR samples taken more than 60 or less than 15 minutes after the waking sample. All 76 cortisol participants had at least one morning and bedtime sample, and 74 participants had data to calculate at least one CAR.[66]

Violent crime. Crime data were obtained from a geocoded file identifying the date and location of every violent crime reported to the city police department. I created an indicator

[66] One participant had only one morning cortisol value and three had only enough data to calculate one CAR. All participants had at least two bedtime cortisol samples. The participants with only one sample do not contribute to the fixed effect estimates.

variable equal to one if a violent crime occurred in a participant's home police beat in a given day, from midnight to midnight. Police beats are sub-divisions of police districts that represent areas that officers can patrol by foot. The city is divided into over 250 police beats. I had 80 instances of violent crime (43 assaults, 31 robberies, 4 criminal sexual assaults, and 2 homicides) in a participant's beat from Saturday through Wednesday of the participant's associated study week, with 51% of adolescents (N=42) having at least one crime in their beat during their participation period. For sensitivity analyses, exact distances between participants' addresses and crime locations were calculated.

Empirical strategy

The analysis uses individual fixed effects to measure the change in outcome from the person's typical pattern on the day following a local violent crime, as follows:

$$Y_{it} = \alpha + \beta Crime_{it} + \nu_t + \gamma_i + \varepsilon_{it} \tag{3-1}$$

where Y_{it} is the adolescent's objectively measured sleep or cortisol on a given day, $Crime_{it}$ is an indicator for whether crime occurred the preceding day, ν_t is a fixed effect for the day of the week, γ_i is an individual fixed effect, and ε_{it} is an idiosyncratic error term. The parameter of interest is β; it captures the causal effect of exposure to crime on the outcome. The inclusion of individual fixed effects enables us to compare sleep and cortisol outcomes for the same adolescent in a day with proximity to crime relative to a day without proximity to crime. Only those adolescents who experience a nearby crime during the study period contribute to this estimate (N=42). I control for day of the week to account for daily differences in sleep and cortisol patterns; participants with and without local violent crime contribute to this estimate. All standard errors are robust and clustered within individuals. I examine outcomes described in the measures section in separate analyses.

Additional analyses add indicators for two and three days after the violent crime to measure the persistence of changes in the response to the acute crime event. I add an indicator for the day before a violent crime as a placebo test: outcomes should not change in anticipation of a crime occurring the following day.

I also separate the analysis by type of violent crime (robbery, assault, criminal sexual assault, and homicide) to test whether responses differ by severity of crime. This analysis should be interpreted cautiously given the low N for certain types of crime. I also examine whether using distance cutoffs rather than an indicator for crime within a police beat provides similar estimates.

Results

Sleep

Table 3-3 displays the sleep results. Because some measures are mathematically related to each other, these should be viewed as related outcomes rather than independent tests. From Panel A, adolescents had significantly later bedtime (by 26 minutes, p=0.043) on evenings immediately following a violent crime in their police beat, relative to their typical bedtime, controlling for day of the week. A robustness check tested whether weekends (defined as Saturday night sleep) drove the results. When running the same estimate but excluding weekend sleep (Panel B), the estimated bedtime effect was 38 minutes later (p=0.027). The standard error was also larger, at least partly because the number of observations dropped from 281 to 211. This estimate did not statistically differ from the estimate including weekends for any of the sleep estimates, indicating that the effects are not driven by weekend sleep.

The first four panels of Figure 3-1 display the effects of violent crime on sleep for days subsequent and prior to acute crime. Each chart was derived from a regression that included interactions between lagged indicators for whether a crime occurred within an individual's police beat during a given day, as well as individual and day-of-the-week fixed effects. Participants' bedtime increased on the night immediately following a crime (p-value=0.041), but the effect dissipated by the second night. Day -1 tested a placebo: sleep should not have changed in *anticipation* of a crime occurring the following day. I found no evidence of an effect in the placebo days or for any of the other outcomes on any day.

Panel C in Table 3-3 displays the analysis by type of crime. Results should be interpreted with caution, given the small number of severe crimes (especially homicide and criminal sexual assault) during the study. Homicide generated the strongest effect on sleep: bedtime increased by 1.8 hours (p=0.000) and total sleep decreased by 1.1 hours (p=0.000). Criminal sexual assault had a large effect on sleep latency (increased by 29 minutes, p=0.026) and total sleep (decreased by 1.0 hours, p=0.002). Coefficients for assault were generally large relative to Panel A, but the standard errors were large. Specifically, bedtime increased by 39 minutes (p=0.042). Effects for robbery were never statistically significant. As a check, I removed robbery from the Panel A estimate (Panel D). This check resulted in larger estimated effects, but also larger standard errors for bedtime (estimated at 39 minutes, p=0.030). The other estimates remained null.

The sample lacked the power to analyze differences in subgroups or other measures. Some exploratory estimates are included in the Appendix; analyses tested whether the effects differed by neighborhood crime level, pubertal status, or time of day of the crime. These suggest that effects on bedtime and sleep duration may be higher for those in low-crime areas and those in the earlier stages of puberty. Perhaps those who rarely experience crime and younger adolescents are more sensitive to acute violent crime, though this needs further exploration. Crime that occurs later in the day may also have larger effects, perhaps due to increased awareness of the crime or through increased arousal that disrupts the transition to sleep. These estimates should be interpreted with caution given the small sample size and represent important areas for future research.

Cortisol

Next, I reviewed the effects of violent crime on same-night bedtime, next-day waking, and next-day CAR cortisol measures (Table 3-4). I found a 0.133 μg/dl increase in next-day CAR in Panel A, representing a 111% increase over the mean CAR (p=0.025). I found no effects for other cortisol outcomes. Robustness checks that dropped two CAR outliers or excluded weekends (defined as Sunday morning; see Panel B) did not substantially change the results.

The final two panels of Figure 3-1 display cortisol effects over time. Each chart was derived from a regression that included interactions between lagged indicators for whether a crime occurred within an individual's police beat during a given time period, as well as individual and day-of-the-week fixed effects. The effect was limited to the CAR on the day following violent crime (p=0.032), and effects dissipated by the next day. As in the sleep panels in Figure 3-1, prior-day placebo tests did not show anticipatory effects.

Panel C of Table 3-4 separates crime by type. Homicide was associated with the largest effects, with a 0.453 μg/dl increase in CAR the day following a homicide, relative to an individual's typical CAR (*p*=0.000). Same-night cortisol also increased by a marginally significant 0.033 µg/dl (*p*=0.80). Assault was associated with a 0.221 µg/dl increase in CAR (*p*=0.011) and a marginally significant 0.085 µg/dl decrease in waking cortisol (*p*=0.059). Criminal sexual assault was associated with a 0.177 µg/dl increase in cortisol on the night of the crime (*p*=0.012). Effects for robbery were never statistically significant. Removing robbery from the Panel A estimate resulted in a larger coefficient for the effect on CAR (estimated at 0.218 µg/dl, *p*=0.003; see Panel D). The other measures remained statistically null in Panel D.

Sleep can change cortisol patterns, and one concern could be that changes to sleep caused by violent crime could affect cortisol measures, rather than the violent crime affecting cortisol directly. Panel E includes results that control for prior-night sleep length and wake time. Longer sleep length is associated with higher waking cortisol, a lower CAR, and lower bedtime cortisol. The other coefficients are largely unchanged, with an estimated 0.140 µg/dl increase in CAR following a violent crime, holding sleep length and wake time constant (*p*=0.053).[67]

The Appendix includes exploratory analyses that tested whether the effects on cortisol differed by neighborhood crime level, pubertal status, or time of day of the crime; these were largely null given the low power. If anything, these suggest that effects on CAR may be higher for those in low-crime areas and those in the later stages of puberty. The pubertal effect is somewhat surprising, given that HPA axis reactivity increases with the pubertal transition (Romeo, 2010; Stroud et al., 2009). The effect of time of day is also not clear in this analysis. These estimates should be interpreted with caution given the small sample size.

[67] Note that the N is slightly lower in Panel E due to the increased requirement that the adolescent have both valid sleep and cortisol data. This particularly reduces the number of available bedtime cortisol measures, as the first bedtime cortisol was measured on Saturday night and there was no Friday night sleep measurement.

Sensitivity to distance of crime from home

The primary analyses used police beats as measures of local neighborhoods, as they account for local landmarks and roads that separate different localities. Here, I consider a violent crime as local if it occurred within various distance cutoffs from a participant's home (see Figure 3-2; range=0.33 to 1.0 miles). Each line represents the estimated coefficient from a separate regression. The figure includes five outcomes: bedtime, sleep latency, sleep duration, next-day waking cortisol, and next-day CAR.

The final panel includes a measure of the number of crimes that occurred at each distance. More proximate crimes were necessarily rarer and generally had larger standard errors. Forty percent of adolescents experienced a violent crime within 0.33 miles of their home during their study participation, while the rate was 67% by 0.5 miles.

Coefficients were generally farther from zero at closer cutoffs. A violent crime within a half mile of an adolescent's home (equivalent to four city blocks) corresponded to a 30 minute increase in bedtime (p=0.036), null effects for latency and wake time, and a 39 minute decrease in sleep duration (p=0.043). There were no consistent cortisol effects using the distance measure, meaning that the CAR estimate is less robust to alternative specifications than the sleep estimates.

Discussion

I explored the relation between an acute external stressor (prior-day local violent crime) and two stress-responsive systems: sleep and the HPA axis. There were nearly 1.2 million violent crimes in the United States in 2015 (U.S. Department of Justice, 2016). Crimes may affect victims, but the present study indicates that it may also affect nearby residents. Local prior-day violence disrupted sleep and increased next-day CAR. I argue that the findings are causal because, within a person, whether local violent crime happens to occur on a given day in the study is random from the perspective of each participant.

The effects on sleep and the CAR were generally large for homicide, moderate for assault, and null for robbery. These patterns could occur either because adolescents reacted more strongly to more violent crime, or because news of such crimes spread more prominently through social networks, media, or police presence. The analysis by type of crime should be treated cautiously, given the low number of participants who experience local criminal sexual assault or homicide in this study.

Estimated effect sizes for sleep were generally larger for more proximate crimes; such crimes may be more likely to be communicated or even directly overheard. The effects for CAR were not statistically significant across distance measures, making these results less robust than the sleep analyses. This difference may be partially related to differences in reliability across days; the Cronbach's alpha for CAR was 0.53, compared to 0.72 for bedtime.

Many of the estimates were not statistically significant. Latency did not change following a crime, perhaps because the students went to bed later rather than tossing and turning after having gone to bed. Wake time increased by a statistically insignificant 11 minutes. Inflexible school start times may prevent adolescents from adjusting their wake time. I predicted an increase in same-day bedtime cortisol and a decrease in next-day waking cortisol, though neither was statistically significant.

A key limitation of the study is the low N. The 18-minute decrease in sleep duration and 18% decrease in waking cortisol are meaningfully sized, but the large standard errors prevented differentiating them from zero (p=0.260 and 0.202, respectively). Future research with larger sample sizes should replicate this analysis, as well as examine subgroup differences, including by gender and race/ethnicity. Preliminary analyses suggest important interactions in the effects by neighborhood crime, pubertal status, and time of day of the crime, which could be important areas for future research with larger samples sizes.

Because the data do not allow identification of adolescents who were direct witnesses of violent events, the estimates are averaged across exposed and non-exposed adolescents. If the effect of violence is heterogeneous across these two types of adolescents (with direct witnesses arguably experiencing more stress), the estimates should be considered a lower bound for the effect that directly exposed adolescents would have experienced. Similarly, if robberies and assaults are underreported and measured with error, the true estimates would be larger in absolute value.

Future research should add objective measures of cognitive functioning to examine relations between semi-random stressful events, sleep, cortisol, and daily cognitive functioning.[68] Such research may offer a mechanism explaining some of the prior findings on acute local violence and performance on standardized tests (Gershenson & Tekin, 2015; Sharkey, 2010; Sharkey et al., 2012).

[68] The diary study included measures of cognitive performance, but missing data across sleep, cortisol, and cognitive outcomes precluded using all three measures in the present study.

Concluding Remarks

The three papers presented in this dissertation broadly address racial/ethnic and socioeconomic disparities in the United States education system. The papers represent the beginning of a research agenda based around the Stress Disparity Model laid out by Heissel, Levy et al. (2017). I use the model to argue that differences in stress exposure across demographic groups create achievement gaps through effects on sleep and the HPA axis, which in turn affect learning and test-day academic performance. The papers presented in this dissertation examine components of that that model; my future work will expand on the model further.

Paper 3 focuses on the left-hand side of the Stress Disparity Model, examining the effect of unexpected nearby violent crime on students' sleep and cortisol patterns. On the night following a violent crime, adolescents have later bedtimes. Adolescents also have disrupted cortisol patterns the following morning. Supplementary analyses using varying distances of the crime to the child's home address confirm more proximate crimes correspond to later bedtimes and possibly less sleep overall. Given the changes to sleep following acute violent crime, Paper 1 provides an important link from sleep to academic achievement. Sleep is at least partially regulated by the sun, and I find that more hours of sunlight before school improves academic achievement, particularly for adolescents in math.

While Paper 3 examines stressors at the community level, Paper 2 focuses on stressors at the family level as it examines how the birth of a baby to a teenage mother affects her siblings. I show that methods that compare students based on first-observed characteristics, rather than trajectories over several years, systematically overestimate the negative effects of teen pregnancy because of these pre-pregnancy trends. However, when compared to students on a similar trajectory in families from the same neighborhood but without teenage childbearing, the siblings of teen mothers have worse test scores, higher high school dropout rates, lower college attendance, and lower college graduation. A similar analysis provides a maximum bound for the effects of teen pregnancy on mothers' own outcomes, finding lower test scores, more grade repetition, higher high school dropout, lower college-going rates, and lower college graduation, relative to females from the same neighborhood who had been on a similar downward trajectory but who did not experience a teenage birth.

All three studies are related to the achievement gap. Given that lower-income individuals are more likely to live in higher-crime areas, and that sleep and cortisol affect academic achievement, Study 3 could plausibly provide a pathway by which crime contributes to gaps in academic achievement (Buckhalt, 2011; Gershenson & Tekin, 2015; Joëls et al., 2006; Lupien et al., 2002; Maldonado et al., 2008; Schwabe & Wolf, 2010; Sharkey, 2010; Sharkey et al., 2012; Vedhara et al., 2000; Walker & Stickgold, 2006). My paper not only provides a novel identification strategy and results, but it is also serves as a call for future research. Studies with larger sample sizes could examine differences by subgroups. New studies could also add measures of cognitive functioning to directly study the relationship between acute violent crime, sleep, the HPA axis, and cognitive functioning as a potential mechanism behind the achievement gap. Results from such studies could be used to estimate the academic costs of neighborhood violence – and how much could be gained by reducing it.

Teen pregnancy rates differ by race/ethnicity and socioeconomic status (Child Trends Data Bank, 2016; Curtin, Abma, Ventura, & Henshaw, 2013), but results from Study 2 indicate that teen pregnancy likely has little effect on the nationwide academic achievement gap, mainly due to the relative rarity of teen pregnancy. A back-of-the-envelope calculation finds that entirely eliminating teen pregnancy among 15- to 17-year-olds would reduce the black-white achievement gap by 0.6% in math (0.4% from teen mothers and 0.2% for their siblings) and 1.3% in reading (1.0% from teen mothers and 0.3% for their siblings).[69]

Even with more generous assumptions, changes to teen pregnancy are unlikely to meaningfully reduce the overall academic achievement gap in the United States, but for individual families and communities reductions in teen pregnancy may offer important ways to improve life outcomes. The public health community's focus on teen pregnancy indicates that the public and policymakers consider teen pregnancy an important area of emphasis, and programs that effectively reduce teen pregnancy may have higher benefits than previously

---

[69] I take the black-white achievement gap as 0.90 SD in math and 0.66 SD in reading (Stanford CEPA, 2016) and the teen birth rate as 67.8 per 1,000 black girls and 20.6 per 1,000 white girls (Child Trends Data Bank, 2016; Curtin, Abma, Ventura, & Henshaw, 2013). Achievement gap reductions calculated by multiplying the estimated effect in SD by subject and racial category by 0.95 (the median number of younger siblings per teen mother, from the conclusion of Study 2), the 15-17-year-old birth rate, and dividing by two to translate from girls to all children. The estimated effect of having a sister give birth for black siblings is 0.091 SD in math and 0.103 in reading; the estimated effect for black teen mothers is 0.177 SD in math and 0.234 SD in reading. The estimated effect for white siblings is 0.081 SD in math and 0.155 in reading; the estimated effect for white teen mothers is 0.194 SD in math and 0.114 SD in reading (results available upon request). Thus, to estimate the effect of eliminating births for black siblings in math the calculation is:

$$\Delta_{math,\ black,\ siblings} = 0.091 \text{ SD/younger sibling} * 0.95 \text{ younger siblings/teen birth} * 67.8 \text{ teen births/1,000 girls} * 1 \text{ girl/2 students} = 0.0029 \text{ SD}$$

Similarly, to estimate the effect of eliminating births for white siblings in math the calculation is:

$$\Delta_{math,\ white,\ siblings} = 0.081 \text{ SD/younger sibling} * 0.95 \text{ younger siblings/teen birth} * 20.6 \text{ teen births/1,000 girls} * 1 \text{ girl/2 students} = 0.0008 \text{ SD}$$

This would decrease the achievement gap by 0.0021 (0.0029-0.0008), or 0.2% (0.0021/0.9000). The direct math effect for black teen mothers is calculated as:

$$\Delta_{math,\ black,\ mothers} = 0.177 \text{ SD/younger sibling} * 67.8 \text{ teen births/1000 girls} * 1 \text{ girl/2 students} = 0.0060 \text{ SD}$$

These estimates may somewhat overstate the true effects, as they counts each birth as the first baby for the family, though they may understate the effect if families with teen births are larger than typical families. In either case, the differences are unlikely to substantially change the calculation.

calculated. For instance, the improvement in test scores from reducing teen pregnancy is 86% higher when including the effects on the siblings of teen mothers; the benefits of teen pregnancy reduction for high school dropout and obtaining any post-secondary degree are 27% and 51% higher, respectively.[70]

Study 1 offers perhaps the most direct option for intervention available to school districts among the three papers presented here. I show that schools can improve academic performance by reordering start times, such that elementary schools start first, followed by middle schools and finally high schools, even without changing the average school start time in the district. Alternatively, schools with older students could shift their start times later without changing the times for elementary schools, which could limit concerns with young students waiting for the bus in the dark. Such changes are more cost-effective than other education policy proposals, such as smaller class sizes or higher-skilled teachers. Even without changing official school start times, districts and schools can make decisions that change the effective start times for students. For instance, some schools require students in need of remediation to show up early during "zero hour" before the first official period.

Black and lower-SES students have worse sleep quality and quantity than white and higher-SES students, respectively (Adam et al., 2007; Hale et al., 2013; Heissel, Levy, et al., 2017), so districts could proactively target changes in school start times accordingly. If, for instance, busing constraints require that some schools start early, schools could be chosen to minimize negative academic consequences for already low-achieving schools. Schools could also conduct remediation sessions after school rather than before.

---

[70] Calculated by multiplying the sibling effect by 0.95 (the median number of younger siblings per teen mother). This number is then divided by the teen mother effect to estimate the percentage difference.

The three papers are connected in additional ways related to my future research agenda. Specifically, the papers examine how adolescents spend their time, whether that's focused on sleep (Studies 1 and 3) or daily activity (how the sisters of teen mother spend their time relative to other teen females and the brothers of teen mothers, Study 2). This broadly fits my other work on teacher time use in turnaround schools (Heissel & Ladd, 2016) and teen time use regarding technology, face-to-face interactions, and sleep (Tavernier, Heissel, Sladek, Grant, & Adam, forthcoming). Future work will examine the time use changes induced in students by stressors, with a particular focus on family-wide time use substitutions in teen parenting families.

Some of my present work focuses on sleep and cortisol as biological drivers of changes in academic performance, and future work with Emma Adam, Jennifer Doleac, David Figlio, and Jonathan Meer will examine changes to cortisol measures induced by high-stakes testing in New Orleans charter schools, as well as how differences in stress responsivity are related to performance on the high-stakes test. Other work with Claudia Persico and David Simon will examine the effects of air and water pollutants on academic outcomes. Again, given differences in stress exposure and pollution by racial/ethnic groups and socioeconomic status, these studies provide another link between demographic groups and academic achievement gaps.

Overall, the present research examines several unexpected sources of stress and disruption that have received little attention in the academic literature – sunlight exposure, a sister's teen childbearing, and nearby violent crime. These factors may explain academic performance at the individual level, as well as some portion of the academic achievement gap broadly.

Tables

Tables – Study 1

Table 1 - 1: Sample characteristics, Florida Panhandle movers in third grade

Panel A: School characteristics					
FRL (fraction)	0.54	0.55	0.56	0.56	0.000
	[0.27]	[0.24]	[0.21]	[0.30]	(0.038)
Male (fraction)	0.51	0.51	0.51	0.51	0.003
	[0.02]	[0.02]	[0.03]	[0.03]	(0.004)
Black (fraction)	0.25	0.26	0.20	0.37	-0.168***
	[0.27]	[0.28]	[0.22]	[0.47]	(0.057)
Hispanic (fraction)	0.04	0.03	0.03	0.03	-0.008
	[0.04]	[0.04]	[0.02]	[0.07]	(0.008)
Asian (fraction)	0.02	0.02	0.01	0.01	0.004
	[0.02]	[0.02]	[0.02]	[0.02]	(0.003)
District Grade 3 math scores (SD)	0.11	0.11	0.12	0.08	0.039
	[0.22]	[0.25]	[0.21]	[0.34]	(0.043)
District Grade 3 reading scores (SD)	0.15	0.15	0.17	0.09	0.084**
	[0.22]	[0.23]	[0.17]	[0.34]	(0.041)
District Grade 3 absentee rates	4.54	4.48	4.39	4.74	-0.358
	[0.77]	[1.13]	[1.70]	[1.40]	(0.227)
1999 median income by zip, logged	10.67	10.64	10.59	10.62	-0.036
	[0.27]	[0.26]	[0.26]	[0.36]	(0.051)
Student/teacher ratio	15.43	15.72	15.40	15.80	-0.400
	[1.20]	[1.40]	[2.15]	[1.71]	(0.273)
Charter school (fraction)	0.02	0.01	0.01	0.02	-0.015
	[0.12]	[0.07]	[0.05]	[0.14]	(0.017)
Urban (fraction)	0.27	0.24	0.18	0.27	-0.086
	[0.48]	[0.47]	[0.49]	[0.63]	(0.084)
Panel B: Individual characteristics					
FRL (=1)	0.55	0.66	0.67	0.69	-0.017
	[0.50]	[0.47]	[0.47]	[0.46]	(0.025)
Male (=1)	0.52	0.51	0.51	0.52	-0.003
	[0.50]	[0.50]	[0.50]	[0.50]	(0.026)
Black (=1)	0.26	0.26	0.25	0.26	-0.008
	[0.44]	[0.44]	[0.44]	[0.44]	(0.023)
Hispanic (=1)	0.04	0.04	0.02	0.03	-0.008
	[0.19]	[0.20]	[0.15]	[0.18]	(0.009)
Asian (=1)	0.02	0.01	0.01	0.01	0.002
	[0.13]	[0.11]	[0.10]	[0.09]	(0.005)
Math score (SD)	0.11	0.02	-0.06	0.00	-0.064
	[0.96]	[0.92]	[0.88]	[0.88]	(0.047)
Reading score (SD)	0.15	0.07	0.00	0.00	0.003
	[0.97]	[0.93]	[0.90]	[0.93]	(0.048)
Absentee rate	4.52	5.60	5.44	6.46	-1.026***
	[4.44]	[5.18]	[5.16]	[5.60]	(0.325)
Observations	186,278	13,788	713	726	

Sample is all third graders in the panhandle. Categorical variables reported as 0-1. Absentee rate is reported as the percentage (0-100) of days missed in the school year to ease interpretation. SD in square brackets. SE in parentheses and clustered at the school level in Panel A, unclustered in Panel B. $* \; p < 0.10$, $** \; p < 0.05$, $*** \; p < 0.01$.

Table 1 - 2: Academic and behavioral outcomes on start times with student fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: First stage, relative start time (hours)							
CT(=1)	0.471**	0.345**	0.424**	0.415**	0.346**	0.424**	0.415**
	(0.016)	(0.021)	(0.020)	(0.020)	(0.021)	(0.020)	(0.020)
CTXPuberty	0.264**	0.265**	0.306**	0.265**	0.265**	0.306**	0.265**
	(0.012)	(0.011)	(0.012)	(0.011)	(0.012)	(0.011)	(0.012)
P(CT+CT X Puberty=0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Panel B: IV estimates, math test scores (SDs) on relative start time							
Start time - sunrise (h)	-0.063**	0.014	0.020	0.010	0.012	0.020	0.009
	(0.026)	(0.041)	(0.036)	(0.035)	(0.041)	(0.036)	(0.035)
Start time X puberty	0.099***	0.074***	0.058***	0.074**	0.073***	0.057***	0.073***
	(0.018)	(0.020)	(0.021)	(0.019)	(0.020)	(0.021)	(0.019)
P(Start+Start X puberty=0)	0.042	0.002	0.001	0.001	0.002	0.001	0.001
Cragg-Donald F-stat	1101.18	404.14	588.90	405.14	405.14	588.76	542.01
Panel C: IV estimates, reading test scores (SDs) on relative start times							
Start time - sunrise (h)	0.064**	0.088**	0.081**	0.061*	0.087**	0.081**	0.061*
	(0.027)	(0.041)	(0.037)	(0.036)	(0.041)	(0.037)	(0.036)
Start time X puberty	-0.005	-0.014	-0.023	-0.005	-0.013	-0.023	-0.004
	(0.018)	(0.021)	(0.022)	(0.020)	(0.021)	(0.022)	(0.020)
P(Start+Start X puberty=0)	0.000	0.004	0.008	0.014	0.004	0.008	0.014
Cragg-Donald F-stat	1230.00	485.69	637.13	618.88	486.65	637.22	619.26
Panel D: IV estimates, absence rate (%) on relative start times							
Start time - sunrise (h)	-0.937***	-1.885***	-0.696	-0.856*	-1.860***	-0.718	-0.869*
	(0.361)	(0.594)	(0.476)	(0.487)	(0.590)	(0.474)	(0.485)
Start time X puberty	0.481**	0.846***	0.365	0.443*	0.857***	0.395	0.469*
	(0.245)	(0.295)	(0.286)	(0.268)	(0.294)	(0.285)	(0.268)
P(Start+Start X puberty=0)	0.062	0.008	0.264	0.206	0.010	0.274	0.219
Cragg-Donald F-stat	689.75	273.69	425.19	383.57	274.18	425.38	383.62
Longitude	No	Yes	Yes	Yes	Yes	Yes	Yes
District quality	No	No	Yes	No	No	Yes	No
School quality	No	No	No	Yes	No	No	Yes
Time since move	No	No	No	No	Yes	Yes	Yes

Dependent variable as noted in panel heading. Test scores measured in SDs normalized at the grade-year level for the entire state. Absentee rate is the fraction of days the child missed school. Start time and its interaction with puberty are instrumented by time zone. Sample is all children who moved. All specifications include age-gender dummies and individual fixed effects. Sample size is fixed within panels: 34,018 students and 115,778 student-years in Panel A, 24,768 students and 99,835 student-years in Panel B, 25,191 students and 104,791 student-years in

Panel C, and 15,906 students and 66,263 student-years in Panel D. Standard errors in parentheses and clustered at the individual level. ∗ $p < 0.10$, ∗∗ $p < 0.05$, ∗∗∗ $p < 0.01$.

Table 1 - 3: Academic and behavioral outcomes on start time, by group with student fixed effects

	White (1)	Non-white (2)	Non-FRL (3)	FRL (4)	Male (5)	Female (6)
Panel A: Math Test Scores (SDs)						
Start time - sunrise (h)	0.021	-0.017	0.045	-0.015	0.027	-0.008
	(0.039)	(0.095)	(0.055)	(0.046)	(0.050)	(0.050)
Start time X puberty	0.072***	0.098**	0.102***	0.063**	0.076***	0.072***
	(0.022)	(0.046)	(0.032)	(0.025)	(0.027)	(0.028)
P(Start+Start X puberty=0)	0.000	0.182	0.000	0.137	0.003	0.069
Cragg-Donald F-stat	459.66	84.63	177.22	373.97	263.79	277.79
Number of students	17013	7755	10052	14716	12380	12388
Observations	70535	29300	40140	59695	49436	50399
Panel B: Reading Test Scores (SDs)						
Start time - sunrise (h)	0.034	0.135	0.072	0.056	0.055	0.072
	(0.040)	(0.092)	(0.056)	(0.047)	(0.051)	(0.050)
Start time X puberty	0.006	-0.003	-0.028	0.006	0.006	-0.018
	(0.024)	(0.046)	(0.035)	(0.025)	(0.028)	(0.029)
P(Start+Start X puberty=0)	0.113	0.018	0.215	0.037	0.060	0.101
Cragg-Donald F-stat	516.36	100.07	221.60	407.29	289.00	333.87
Number of students	17264	7927	10284	14907	12560	12631
Observations	73872	30919	42458	62333	51752	53039
Panel C: Absence Rate (%)						
Start time - sunrise (h)	-0.357	-2.012	-1.094	-0.619	-0.564	-1.277*
	(0.531)	(1.312)	(0.737)	(0.625)	(0.622)	(0.752)
Start time X puberty	-0.193	1.723***	0.298	0.533	0.201	0.794**
	(0.324)	(0.622)	(0.411)	(0.343)	(0.379)	(0.377)
P(Start+Start X puberty=0)	0.123	0.720	0.089	0.840	0.379	0.346
Cragg-Donald F-stat	320.62	58.76	116.36	270.00	193.18	190.14
Number of students	10613	5293	6383	9523	8019	7887
Observations	45654	20609	26483	39780	32994	33269

Dependent variable as noted in panel heading. Test scores measured in SDs normalized at the grade-year level for the entire state. Absentee rate is the percent of days the child missed school. Start time and its interaction with puberty are instrumented by time zone. Sample is all children who moved more than 25 miles. All specifications include age-gender dummies, longitude controls, school demographic means (FRL, male, black, Asian, and Hispanic) and individual fixed effects. Standard errors in parentheses and clustered at the individual level. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 1 - 4: Persistence in effects of relative start time on student outcomes, with student fixed effects

	Math Scores		Reading Scores	
	(1)	(2)	(3)	(4)
Start time - sunrise (h) (prepubescent)	0.009	0.007	0.061*	0.052
	(0.035)	(0.036)	(0.036)	(0.036)
Start X moved two years ago (pre)		0.002		0.011
		(0.009)		(0.009)
Start X moved 3+ years ago (pre)		-0.011		-0.005
		(0.012)		(0.012)
Start time - sunrise (h) (pubescent)	0.082***	0.087***	0.057**	0.048**
	(0.025)	(0.026)	(0.023)	(0.024)
Start X moved two years ago (pub)		-0.016***		-0.004
		(0.006)		(0.006)
Start X moved 3+ years ago (pub)		-0.020***		0.010
		(0.007)		(0.007)
P[Start (pre) = Start (pub)]	0.000	0.000	0.826	0.861
P[Start (pre) = Start (pub), long run]		0.000		0.577
Cragg-Donald F-stat	542.01	107.47	619.26	124.19
Number of students	24,768	24,768	25,191	25,191
Observations	99,835	99,835	104,791	104,791

Dependent variable as noted in panel heading. Test scores measured in SDs normalized at the grade- year level for the entire state. Start time and its interaction with puberty are instrumented by time zone and the interaction of time zone and puberty. Sample is all children who moved more than 25 miles. All specifications include age-gender dummies, longitude controls, school demographic means (FRL, male, black, Asian, and Hispanic) and individual fixed effects. Standard errors in parentheses and clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1 - 5: Academic outcomes, for testing before and after Daylight Saving Time

	Math Scores		Reading Scores	
	(1)	(2)	(3)	(4)
Start time - sunrise (h) (prepubescent)	0.030		0.056	
	(0.038)		(0.038)	
Start time - sunrise (h) (pubescent)	0.096***		0.060**	
	(0.027)		(0.024)	
Start time X prepubescent, late test time		0.022		0.049
		(0.039)		(0.039)
Start time X prepubescent, early test time		0.071		0.096**
		(0.046)		(0.047)
Start time X adolescent, late test time		0.095***		0.045*
		(0.030)		(0.026)
Start time X adolescent, early test time		0.096***		0.104***
		(0.025)		(0.026)
Era X puberty controls	No	Yes	No	Yes
P[Early = late test (Prepub)]		0.165		0.192
P[Early = late test (Adol)]		0.967		0.001
Cragg-Donald F-stat	468.563	229.684	542.050	269.539
Number of students	23,618	23,618	24,152	24,152
Observations	89,707	89,707	94,515	94,515

Dependent variable as noted in panel heading. Test scores measured in SDs normalized at the grade-year level for the entire state. Start time and its interactions are instrumented by time zone and the interaction of time zone and interactions. Sample is all children who moved more than 25 miles. All specifications include age-gender dummies, longitude controls, school demographic means (FRL, male, black, Asian, and Hispanic), time since move dummies, and individual fixed effects. Sample includes years 2000-2013 excluding 2010, when testing took place over the DST time change. Standard errors in parentheses and clustered at the individual level. * $p < 0.10$,** $p < 0.05$,*** $p < 0.01$.

Tables – Study 1 Appendix

Table A1 - 1: Academic outcomes on school start time for varying mover definitions, with student fixed effects

	Math (SDs)					Reading (SDs)				
	(1) district	(2) 15 mi	(3) 20 mi	(4) 25 mi	(5) 30 mi	(6) district	(7) 15 mi	(8) 20 mi	(9) 25 mi	(10) 30 mi
Start time - sunrise (h)	0.037	0.029	0.014	0.009	0.009	0.037	0.034	0.026	0.061*	0.053
	(0.034)	(0.033)	(0.031)	(0.035)	(0.037)	(0.036)	(0.034)	(0.032)	(0.036)	(0.038)
Start time X puberty	0.036**	0.038**	0.070***	0.073***	0.060***	0.007	0.011	0.018	-0.004	-0.008
	(0.018)	(0.017)	(0.018)	(0.019)	(0.022)	(0.019)	(0.018)	(0.018)	(0.020)	(0.023)
P(Start+Start X puberty=0)	0.001	0.002	0.000	0.001	0.004	0.029	0.025	0.033	0.014	0.049
Cragg-Donald F-stat	610.14	611.40	677.49	542.01	542.98	684.27	701.42	766.47	619.26	612.31
Number of students	33712	35744	28969	24768	21557	34144	36197	29393	25191	21957
Observations	143921	153462	120233	99835	84165	150800	160997	126110	104791	88408

Dependent variable as noted in panel heading. Test scores measured in SDs normalized at the grade-year level for the entire state. Start time and its interaction with puberty are instrumented by time zone and the interaction of time zone and puberty. All specifications include age-gender dummies, longitude controls, school demographic means (FRL, male, black, Asian, and Hispanic), and individual fixed effects. Standard errors in parentheses and clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A1 - 2: Academic and behavioral outcomes on start time, with student fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
								Panel A: Math Test Scores (SDs)							
Start time - sunrise (h)	0.012	0.011	0.009	0.020	0.028	0.028	0.031	0.036	0.037	0.009	0.014	0.014	0.003	0.012	0.012
	(0.041)	(0.035)	(0.035)	(0.036)	(0.037)	(0.036)	(0.035)	(0.036)	(0.036)	(0.035)	(0.034)	(0.034)	(0.036)	(0.034)	(0.034)
Start time X puberty	0.073***	0.051***	0.054***	0.057***	0.037*	0.039*	0.065***	0.041**	0.043**	0.073***	0.050***	0.053***	0.076***	0.050***	0.053***
	(0.020)	(0.019)	(0.019)	(0.021)	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
P(Start+Start X puberty=0)	0.002	0.005	0.005	0.001	0.003	0.002	0.000	0.001	0.000	0.001	0.003	0.002	0.001	0.004	0.003
Cragg-Donald F-stat	405.14	593.76	604.37	588.76	593.74	606.81	580.48	599.33	612.93	542.01	640.62	655.01	534.48	638.53	655.08
Number of students	24768	23516	23516	24768	23516	23516	24545	23294	23294	24768	23516	23516	24765	23514	23514
Observations	99835	91853	91853	99835	91853	91853	98751	90852	90852	99835	91853	91853	99823	91846	91846
								Panel B: Reading Test Scores (SDs)							
Start time - sunrise (h)	0.087**	0.061*	0.061*	0.081**	0.075**	0.074**	0.071**	0.065*	0.065*	0.061*	0.049	0.048	0.051	0.046	0.046
	(0.041)	(0.035)	(0.035)	(0.037)	(0.037)	(0.036)	(0.035)	(0.036)	(0.035)	(0.036)	(0.034)	(0.034)	(0.036)	(0.034)	(0.034)
Start time X puberty	-0.013	-0.009	-0.008	-0.023	-0.022	-0.022	-0.011	-0.013	-0.013	-0.004	-0.003	-0.003	0.000	-0.002	-0.002
	(0.021)	(0.020)	(0.020)	(0.022)	(0.021)	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)	(0.019)	(0.019)	(0.020)	(0.019)	(0.019)
P(Start+Start X puberty=0)	0.004	0.015	0.015	0.008	0.014	0.014	0.008	0.015	0.014	0.014	0.027	0.027	0.025	0.030	0.030
Cragg-Donald F-stat	486.65	679.86	687.26	637.22	648.26	671.04	656.76	675.89	697.05	619.26	729.44	746.01	616.60	725.65	742.75
Number of students	25191	24048	24048	25191	24048	24048	24963	23823	23823	25191	24048	24048	25189	24045	24045
Observations	104791	96788	96788	104791	96788	96788	103547	95641	95641	104791	96788	96788	104776	96776	96776
								Panel C: Absence Rates							
Start time - sunrise (h)	-1.860***	-1.463***	-1.431***	-0.718	-0.709	-0.695	-0.848*	-0.789*	-0.772*	-0.869*	-0.874*	-0.859*	-0.965**	-0.904*	-0.880*
	(0.590)	(0.505)	(0.502)	(0.474)	(0.483)	(0.479)	(0.460)	(0.471)	(0.467)	(0.485)	(0.467)	(0.464)	(0.492)	(0.470)	(0.466)
Start time X puberty	0.857***	0.677**	0.637**	0.395	0.330	0.304	0.439	0.353	0.320	0.469*	0.384	0.365	0.491*	0.396	0.367
	(0.294)	(0.278)	(0.275)	(0.285)	(0.286)	(0.283)	(0.274)	(0.278)	(0.275)	(0.268)	(0.268)	(0.265)	(0.269)	(0.270)	(0.266)
P(Start+Start X puberty=0)	0.010	0.012	0.011	0.274	0.182	0.166	0.156	0.117	0.103	0.219	0.091	0.087	0.151	0.081	0.077
Cragg-Donald F-stat	274.18	413.70	416.25	425.38	431.70	439.86	453.02	458.47	467.24	383.62	451.74	458.44	373.38	447.12	454.86
Number of students	15906	15130	15130	15906	15130	15130	15906	15130	15130	15906	15130	15130	15903	15128	15128
Observations	66263	61128	61128	66263	61128	61128	66263	61128	61128	66263	61128	61128	66252	61122	61122
Urban and log income	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Size and S/T ratio	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
District controls	No	No	No	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No
District grade 3 controls	No	No	No	No	No	No	Yes	Yes	Yes	No	No	No	No	No	No

School controls	No	No	No	No	No	No	No	No	No	Yes	Yes	Yes	No	No	No
School-grade controls	No	No	No	No	No	No	No	No	No	No	No	No	Yes	Yes	Yes

Dependent variable as noted in panel heading. Test scores measured in SDs normalized at the grade-year level for the entire state. Absentee rate is the fraction of days the child missed school. Start time and its interaction with puberty are instrumented by time zone. Sample is all children who moved more than 25 miles. All specifications include age-gender dummies, longitude, and individual fixed effects. Standard errors in parentheses and clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A1 - 3: Outcomes on school start time, with latitude and school test grade scores

	Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)
Start time - sunrise (h)	0.009	-0.035	0.015	0.061*	0.035	0.051
	(0.035)	(0.033)	(0.037)	(0.036)	(0.034)	(0.037)
Start time X puberty	0.073***	0.085***	0.073***	-0.004	0.004	-0.001
	(0.019)	(0.019)	(0.020)	(0.020)	(0.020)	(0.020)
Latitude controls	No	Yes	No	No	Yes	No
Third grade district scores	No	No	Yes	No	No	Yes
P(Start+Start X puberty=0)	0.001	0.029	0.001	0.014	0.069	0.035
Cragg-Donald F-stat	542.01	631.95	508.46	619.26	715.55	589.27
Number of students	24768	24768	24288	25191	25191	24730
Observations	99835	99835	97483	104791	104791	102276

Dependent variable as noted in panel heading. Test scores measured in SDs normalized at the grade- year level for the entire state. Start time and its interaction with puberty are instrumented by time zone and the interaction of time zone and puberty. Sample is all children who moved more than 25 miles. All specifications include age-gender dummies, longitude controls, school demographic means (FRL, male, black, Asian, and Hispanic), and individual fixed effects. Standard errors in parentheses and clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A1 - 4: Florida school and peer characteristics on move

	% FRL (1)	% male (2)	% black (3)	% Hispanic (4)	% Asian (5)	S/T (6)	Med income (7)
Move, ET-ET	-4.494***	-0.452***	0.186	-0.100	0.263***	0.258***	1010.277*
	(0.726)	(0.118)	(0.801)	(0.224)	(0.059)	(0.081)	(601.359)
Move, CT-CT	-1.681***	-0.316***	-0.582**	0.110***	-0.011	0.190***	-429.606***
	(0.280)	(0.054)	(0.227)	(0.037)	(0.025)	(0.038)	(162.849)
Move, ET-CT	0.115	-0.009	-15.350***	0.025	0.426***	0.124	-4778.338***
	(0.923)	(0.162)	(1.015)	(0.183)	(0.084)	(0.103)	(731.901)
Move, CT-ET	-4.513***	-0.557***	13.965***	0.495***	0.023	0.113	5729.001***
	(0.939)	(0.163)	(1.010)	(0.166)	(0.088)	(0.101)	(752.117)
P(ET-CT=CT-ET)	0.002	0.029	0.000	0.105	0.003	0.944	0.000
Observations	31763	31763	31763	31763	31763	31763	27747

Dependent variable as noted in panel heading. Regression is of school/zip summary stat on move, with student X moving event FE. Standard errors in parentheses and clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A1 - 5: Alternative definitions of puberty

	Math (SDs)				Reading (SDs)			
	(1) Preferred	(2) Stage 2	(3) Stage 4	(4) BG	(5) Preferred	(6) Stage 2	(7) Stage 4	(8) BG
Start time - sunrise (h)	0.009	0.011	0.032	0.025	0.061*	0.057	0.056	0.058
	(0.035)	(0.036)	(0.035)	(0.035)	(0.036)	(0.036)	(0.036)	(0.036)
Start time X puberty	0.073***	0.064***	0.029	0.040**	-0.004	0.003	0.006	0.002
	(0.019)	(0.019)	(0.020)	(0.019)	(0.020)	(0.020)	(0.021)	(0.020)
P(Start+Start X puberty=0)	0.001	0.003	0.005	0.008	0.014	0.012	0.002	0.010
Cragg-Donald F-stat	542.01	566.32	444.15	542.35	619.26	655.35	487.58	615.52
Number of students	24768	24768	24768	24768	25191	25191	25191	25191
Observations	99835	99835	99835	99835	104791	104791	104791	104791

Dependent variable as noted in panel heading. Test scores measured in SDs normalized at the grade-year level for the entire state. Absentee rate is the fraction of days the child missed school. Start time and its interaction with puberty are instrumented by time zone and the interaction of time zone and puberty. Sample is all children who moved more than 25 miles. All specifications include age-gender dummies, longitude controls, school demographic means (FRL, male, black, Asian, and Hispanic) and individual fixed effects. Standard errors in parentheses and clustered at the individual level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A1 - 6: Academic and behavioral outcomes on start time, with student fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: First stage, relative start time (hours)							
CT (=1)	0.598***	0.475***	0.585***	0.547***	0.475***	0.584***	0.547***
	(0.015)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Observations	115778	115778	115778	115778	115778	115778	115778
Panel B: IV estimates, math test scores (SDs) on relative start time							
Start time - sunrise (h)	-0.005	0.047	0.048*	0.044	0.045	0.048*	0.043
	(0.019)	(0.034)	(0.028)	(0.029)	(0.034)	(0.028)	(0.029)
Cragg-Donald F-stat	2254.173	744.796	1120.532	1002.330	746.364	1120.434	1003.020
Panel C: IV estimates, reading test scores (SDs) on relative start times							
Start time - sunrise (h)	0.061***	0.081**	0.069**	0.059**	0.080**	0.069**	0.059**
	(0.019)	(0.032)	(0.028)	(0.028)	(0.032)	(0.028)	(0.028)
Cragg-Donald F-stat	2587.05	911.72	1209.23	1151.57	913.31	1209.80	1152.03
Panel D: IV estimates, absence rate (%) on relative start times							
Start time - sunrise (h)	-0.664**	-0.539***	-0.549	-0.670*	-0.510***	-0.559	-0.672*
	(0.275)	(0.501)	(0.391)	(0.407)	(0.499)	(0.389)	(0.405)
Longitude	No	Yes	Yes	Yes	Yes	Yes	Yes
District quality	No	No	Yes	No	No	Yes	No
School quality	No	No	No	Yes	No	No	Yes
Time since move	No	No	No	No	Yes	Yes	Yes
Cragg-Donald F-stat	1394.52	475.67	721.91	669.77	476.44	722.82	669.98

Dependent variable as noted in panel heading. Test scores measured in SDs normalized at the grade-year level for the entire state. Absentee rate is the fraction of days the child missed school. Relative start time instrumented by time zone. Sample is all children who moved more than 25 miles. All specifications include age-gender dummies and individual fixed effects. Sample size is fixed within panels: 34018 students and 115778 student-years in Panel A, 24768 students and 99835 student-years in Panel b, 25191 students and 104791 student-years in Panel C, and 15906 students and 66263 student-years in Panel D. Standard errors in parentheses and clustered at the individual level. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table A1 - 7: Hours of sleep by time zone

	(1)	(2)	(3)
Central	0.081	0.103	
	(0.088)	(0.131)	
Puberty	-0.451***	-0.804***	-0.676***
	(0.055)	(0.122)	(0.134)
Weekend	1.421***	1.192***	1.229***
	(0.102)	(0.158)	(0.158)
Eastern X weekend	-0.107	-0.166	-0.102
	(0.156)	(0.194)	(0.188)
Eastern X puberty	0.218	0.183	0.257
	(0.139)	(0.185)	(0.195)
Weekend X puberty	0.384***	0.616***	0.586***
	(0.087)	(0.161)	(0.150)
Eastern X weekend X puberty	-0.215	-0.149	-0.229
	(0.168)	(0.239)	(0.224)
P(Central + Central X weekend = 0)	0.830	0.566	
P(Central + Central X puberty = 0)	0.074	0.085	
Demographic controls	No	Yes	No
Student fixed effects	No	No	Yes
Observations	6,084	3,737	6,084

Dependent variable is hours of sleep per night. Sample is all children 6-19 in the Child Development Supplement of the Panel Study of Income Dynamics within 400 miles of the ET-CT time zone boundary in a state with a single time zone. Demo- graphic controls in Column 2 include gender, race, and FRL status. Standard errors in parentheses and clustered at the state level. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Tables – Study 2

Table 2 - 1: Descriptive statistics

	(1) Families w/o teen birth, all	(2) Teen mothers	(3) Siblings of teen mothers
Female	48.475	100.000	50.312
	(49.977)	(0.000)	(50.051)
Age at birth	N/A	16.308	12.694
		(0.722)	(2.374)
# of siblings in data	2.385	2.581	3.029
	(0.701)	(0.879)	(1.056)
Oldest sibling	45.655	57.196	0.000
	(49.811)	(49.526)	(0.000)
FRL	35.827	74.393	78.586
	(47.949)	(43.687)	(41.065)
Black	12.162	57.757	56.757
	(32.684)	(49.441)	(49.593)
First-observed test score %	59.319	39.824	42.209
	(26.623)	(24.192)	(26.084)
School avg. FRL	38.050	47.438	47.463
	(17.650)	(15.522)	(15.253)
School avg. Black	15.306	18.942	21.090
	(10.448)	(8.503)	(12.234)
Mean school first-observed test %	58.605	54.303	53.535
	(9.185)	(7.996)	(8.879)
N	102700	535	481

Mean coefficients; SD in parentheses. Families without teen births include all children from non-teenage-childbearing families. Teen mothers include all teen mothers from families of two or more where the mother gives birth at age 15-17. Siblings include all younger siblings from families where an older sister gave birth at age 15-17.

Table 2 - 2: Descriptive statistics, matched controls for siblings

		First-observed match		Trajectory match		
	(1) Siblings of teen mothers	(2) Scores & controls	(3) Scores, controls & nbhd	(4) Score only	(5) Scores & controls	(6) Scores, controls & nbhd
Female	57.692	57.115	53.675	47.788	58.846	54.932
	(3.527)	(1.662)	(1.934)	(1.606)	(1.678)	(1.890)
p	.	0.882	0.317	0.011	0.767	0.490
Age at birth	13.913	13.941	13.521	14.499	13.967	13.690
	(0.086)	(0.061)	(0.062)	(0.059)	(0.055)	(0.064)
p	.	0.792	0.000	0.000	0.599	0.037
# of siblings in data	2.947	2.920	2.786	2.637	2.907	2.766
	(0.090)	(0.045)	(0.043)	(0.030)	(0.052)	(0.043)
p	.	0.789	0.107	0.001	0.697	0.070
FRL	79.327	79.615	77.273	47.308	79.712	78.723
	(3.051)	(1.320)	(1.602)	(1.616)	(1.307)	(1.472)
p	.	0.931	0.551	0.000	0.908	0.858
Black	56.731	56.827	53.578	20.673	57.115	55.513
	(3.944)	(1.730)	(2.084)	(1.352)	(1.742)	(1.975)
p	.	0.982	0.479	0.000	0.929	0.782
Test score % (four years before match)	47.224	47.828	42.169	47.077	47.540	43.794
	(2.092)	(1.048)	(1.260)	(0.930)	(0.963)	(1.255)
p	.	0.796	0.039	0.949	0.891	0.160
Test score % (two years before match)	41.197	46.954	47.309	41.062	41.366	42.881
	(1.812)	(0.916)	(1.021)	(0.798)	(0.818)	(0.943)
p	.	0.005	0.003	0.945	0.932	0.409
School avg. FRL	47.473	47.457	47.106	43.510	47.821	47.257
	(1.245)	(0.562)	(0.694)	(0.559)	(0.542)	(0.640)
p	.	0.990	0.796	0.004	0.797	0.877
School avg. Black	18.875	18.722	19.161	16.953	19.102	19.153
	(0.592)	(0.276)	(0.323)	(0.289)	(0.275)	(0.305)
p	.	0.815	0.671	0.004	0.727	0.675
Mean school first-observed test %	55.129	55.149	55.265	56.569	54.957	55.195
	(0.618)	(0.286)	(0.334)	(0.266)	(0.268)	(0.309)
p	.	0.977	0.846	0.032	0.798	0.924
N	208	1014	926	1028	1022	939

Mean coefficients; SE in parentheses (clustered by family ID). Siblings include all younger siblings from families where a sister gave birth at age 15-17 who had at least two of three years of pre-data. First-observed matches include matches from non-teenage-childbearing families to siblings based on first-observed characteristics. Trajectory matches include matches from non-teenage-childbearing families to siblings based on three-year test score trends and other observable characteristics. Includes *p*-value of *t*-test between matches and siblings.

Table 2 - 3: Estimated effects of teen birth on various outcomes for siblings

	(1) All younger siblings	(2) Matched on first-observed scores, controls, & nbhd	(3) Matched on trajectory scores only	(4) Matched on trajectory scores & controls	(5) Matched on trajectory scores, controls, & nbhd
Test scores at t=0	-6.173***	-6.460***	-2.643	-4.020**	-3.959*
	(1.006)	(1.593)	(1.777)	(1.513)	(1.573)
N	38871	844	825	905	835
Test scores, with age and individual FE	-3.508***	-3.139**	-2.271*	-2.307*	-2.554*
	(0.903)	(1.040)	(1.029)	(1.019)	(1.040)
Observations	406147	6585	7781	7594	6761
N	87008	1059	1212	1156	1074
Repeats grade in t=0 or later	5.107*	5.371	6.207	4.908	2.946
	(2.313)	(3.599)	(4.045)	(3.497)	(3.576)
N	86265	1134	1236	1230	1147
Drops out in t=0 or later	0.698	7.176*	8.699*	9.962***	7.325*
	(1.660)	(2.972)	(3.386)	(2.932)	(2.964)
N	85402	1134	1236	1230	1147
Juvenile justice in t=0 or later	6.141***	7.667**	7.497*	6.676**	5.766*
	(1.716)	(2.606)	(2.931)	(2.488)	(2.661)
N	85402	1134	1236	1230	1147
Ever attends any college	-13.640***	-12.724*	-10.645*	-7.445	-7.406
	(3.041)	(4.931)	(5.006)	(4.647)	(4.885)
N	44974	755	870	863	782
Obtains any degree or certificate	-5.895*	-5.967	-5.610	-6.333	-6.680^{+}
	(2.518)	(3.928)	(4.209)	(3.889)	(3.937)
N	44974	755	870	863	782
Obtains a 4-year degree	-6.114**	-6.566*	-3.064	-6.515*	-7.022*
	(1.945)	(3.205)	(3.773)	(3.161)	(3.211)
N	44974	755	870	863	782

Note: Robust standard errors clustered by family ID. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Analyses include all controls from Table 2-1. Column 1 includes all children from non-teenage-childbearing families as controls. Column 2 includes matches to younger siblings from non-teenage-childbearing families to the siblings based on the first-observed characteristics; matches must be from the same neighborhood. Column 3 includes matches to younger siblings from non-teenage-childbearing families to the siblings based on three-year score trends. Column 4 adds other observable control variables from Table 1. Column 5 adds the requirement that the match must be from the same neighborhood at first observation. Scores at t=0, repeats grade, drops out, juvenile justice, and college-going outcomes include one weighted observation from the siblings and their controls. Column 1 test scores at t=0 includes multiple observations per individual control. Fixed effects models include all observations from the siblings and their matched controls, excluding t=-1.

Table 2 - 4: Estimated effects of teen birth on various outcomes for siblings by subgroups

	(1) Baseline	(2) Female	(3) Male	(4) Black	(5) Non-black	(6) FRL	(7) Non-FRL
Test scores at t=0	-3.959*	-5.354**	-1.128	-2.164	-5.521*	-3.151^{+}	-6.193^{+}
	(1.573)	(1.962)	(2.526)	(1.915)	(2.666)	(1.742)	(3.427)
N	835	467	368	454	381	648	187
Test scores, with age	-2.554*	-3.841**	-0.917	-2.567^{+}	-2.449	-1.960^{+}	-5.004*
and individual FE	(1.040)	(1.315)	(1.718)	(1.390)	(1.536)	(1.171)	(2.204)
Observations	6761	3712	3049	3715	3046	5206	1555
N	1074	584	490	581	493	826	248
Repeats grade in t=0	2.946	5.687	-1.898	4.759	-0.013	1.802	5.370
or later	(3.576)	(4.547)	(5.537)	(4.837)	(5.469)	(4.040)	(6.869)
N	1147	631	516	651	496	899	248
Drops out in t=0 or	7.325*	9.453*	4.708	5.447	9.703^{+}	7.021*	7.600
later	(2.964)	(4.115)	(4.489)	(3.637)	(5.053)	(3.443)	(5.894)
N	1147	631	516	651	496	899	248
Juvenile justice in	5.766*	2.940	9.701*	6.249^{+}	4.353	4.453	10.604^{+}
t=0 or later	(2.661)	(2.916)	(4.794)	(3.666)	(3.846)	(3.017)	(6.159)
N	1147	631	516	651	496	899	248
Ever attends any	-7.406	-8.838	-4.166	-10.415	-5.410	-3.412	-17.742^{+}
college	(4.885)	(6.610)	(7.349)	(6.881)	(7.286)	(5.791)	(9.058)
N	782	431	351	424	358	593	189
Obtains any degree	-6.680^{+}	-9.609^{+}	-2.712	-7.218	-6.511	-4.972	-9.193
or certificate	(3.937)	(5.633)	(5.150)	(5.101)	(6.001)	(4.310)	(8.683)
N	782	431	351	424	358	593	189
Obtains a 4-year	-7.022*	-8.602^{+}	-4.357	-10.530**	-4.156	-6.611*	-6.883
degree	(3.211)	(4.396)	(4.362)	(3.384)	(5.249)	(3.115)	(8.710)
N	782	431	351	424	358	593	189

Note: Robust standard errors clustered by family ID. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All analyses based on the trajectory matches with other controls and same neighborhood requirement. Column 1 is the preferred estimate from the Table 3; later columns repeat the analysis by subgroups.

Table 2 - 5: Estimated effects of teen birth on various outcomes for siblings by tertile

	(1) Lowest tertile	(2) Middle tertile	(3) Highest tertile	(4) *p*-value of Hausman test across tertiles
Test scores at *t*=0	-3.593*	-3.253	-2.610	0. 949
	(1.729)	(3.211)	(2.560)	
N	274	278	283	
Repeats grade in *t*=0 or later	14.382*	-9.123	1.500	0.020
	(6.672)	(5.824)	(5.492)	
N	416	379	352	
Drops out in *t*=0 or later	0.056	9.929+	9.265+	0. 254
	(4.949)	(5.116)	(4.724)	
N	416	379	352	
Juvenile justice in *t*=0 or later	6.200	0.560	8.577*	0. 412
	(5.240)	(4.366)	(4.170)	
N	416	379	352	
Ever attends any college	-13.125	-23.960**	9.639	0.009
	(8.716)	(9.070)	(7.288)	
N	264	249	269	
Obtains any degree or certificate	-5.810	-18.016**	2.940	0.075
	(4.590)	(5.601)	(8.482)	
N	264	249	269	
Obtains a 4-year degree	-5.265**	-13.357**	-2.215	0.149
	(1.763)	(4.062)	(7.946)	
N	264	249	269	

Note: Robust standard errors clustered by family ID. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All analyses based on the trajectory matches with other controls and same neighborhood requirement. Column 1 is the preferred estimate from Table 3; later columns repeat the analysis by tertile of first-observed test scores, with the highest tertile being the highest scorers.

Table 2 - 6: Descriptive statistics, matched controls for teen mothers

		First-observed match		Trajectory match		
	(1) Teen mothers	(2) Scores & controls	(3) Scores, controls & nbhd	(4) Score only	(5) Scores & controls	(6) Scores, controls & nbhd
Age at birth	16.306	16.155	15.541	15.480	16.148	15.587
	(0.031)	(0.048)	(0.060)	(0.047)	(0.046)	(0.055)
p	.	0.009	0.000	0.000	0.004	0.000
# of siblings in data	2.589	2.572	2.547	2.499	2.538	2.537
	(0.039)	(0.027)	(0.030)	(0.021)	(0.026)	(0.030)
p	.	0.724	0.402	0.046	0.283	0.297
Oldest sibling	57.744	51.667	50.027	44.792	50.833	50.294
	(2.162)	(1.351)	(1.380)	(1.236)	(1.314)	(1.383)
p	.	0.017	0.003	0.000	0.006	0.004
FRL	74.570	76.615	76.644	50.781	74.687	76.750
	(1.906)	(1.073)	(1.212)	(1.289)	(1.112)	(1.205)
p	.	0.353	0.362	0.000	0.958	0.337
Black	57.361	60.104	57.082	24.740	61.146	57.616
	(2.165)	(1.322)	(1.534)	(1.144)	(1.317)	(1.497)
p	.	0.286	0.917	0.000	0.141	0.924
Test score % (five years before match)	40.501	43.258	41.871	38.709	39.850	39.999
	(1.386)	(0.725)	(0.869)	(0.687)	(0.687)	(0.790)
p	.	0.082	0.407	0.252	0.677	0.756
Test score % (two years before match)	37.651	40.009	41.535	37.601	36.936	39.665
	(1.354)	(0.731)	(0.816)	(0.679)	(0.693)	(0.762)
p	.	0.127	0.014	0.973	0.639	0.197
School avg. FRL	47.553	48.031	47.675	44.190	48.184	47.822
	(0.676)	(0.408)	(0.491)	(0.433)	(0.441)	(0.480)
p	.	0.550	0.885	0.000	0.440	0.748
School avg. Black	19.030	19.431	19.381	17.268	19.112	19.369
	(0.381)	(0.215)	(0.259)	(0.224)	(0.226)	(0.254)
p	.	0.365	0.451	0.000	0.855	0.465
Mean school first-observed test %	54.205	54.372	54.664	56.227	54.240	54.596
	(0.349)	(0.201)	(0.249)	(0.217)	(0.220)	(0.246)
p	.	0.682	0.290	0.000	0.934	0.364
N	523	1824	1705	1877	1836	1709

Note: Mean coefficients; SE in parentheses (clustered by family ID). $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Teen mothers include all females who gave birth at age 15-17 who had at least two of four years of pre-data. First-observed matches include matches from non-teenage-childbearing families to teen mothers based on first-observed characteristics. Trajectory matches include matches from non-teenage-childbearing families to teen mothers based on four-year test score trends and other observable characteristics. Includes *p*-value of *t*-test between matches and teen mothers.

Table 2 - 7: Estimated effects of teen birth on various outcomes for teen mothers

	(1) All females	(2) Matched on first-observed scores, controls, & nbhd	(3) Matched on trajectory scores only	(4) Matched on trajectory scores & controls	(5) Matched on trajectory scores, controls, & nbhd
Test scores at t=0	-7.445***	-7.054***	-1.830	-3.934**	-4.388**
	(1.470)	(1.566)	(1.570)	(1.473)	(1.481)
N	19179	939	1070	889	967
Repeats grade in t=0 or later	25.426***	18.858***	17.557***	16.660***	17.244***
	(2.294)	(2.540)	(2.629)	(2.493)	(2.528)
N	42255	2189	2361	2320	2193
Drops out in t=0 or later	18.587***	26.414***	21.887***	22.692***	24.051***
	(2.186)	(2.253)	(2.417)	(2.230)	(2.270)
N	41832	2189	2361	2320	2193
Juvenile justice in t=0 or later	-1.675**	-1.112	-1.487*	-0.892	-0.903
	(0.578)	(0.702)	(0.683)	(0.697)	(0.688)
N	41832	2189	2361	2320	2193
Ever attends any college	-20.765***	-20.879***	-18.267***	-20.923***	-18.577***
	(3.314)	(3.705)	(3.681)	(3.562)	(3.674)
N	22169	1291	1541	1451	1324
Obtains any degree or certificate	-13.117***	-15.279***	-12.638***	-12.767***	-12.540***
	(2.375)	(2.807)	(2.619)	(2.457)	(2.656)
N	22169	1291	1541	1451	1324
Obtains at least a 4-year degree	-11.617***	-13.720***	-11.641***	-11.953***	-11.693***
	(1.811)	(2.012)	(1.748)	(1.739)	(1.888)
N	22169	1291	1541	1451	1324

Note: Robust standard errors clustered by family ID. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Analyses include all controls from Table 1. Column 1 includes all females from non-teenage-childbearing families as controls. Column 2 includes matches to females from non-teenage-childbearing families to the teen mothers based on the first-observed characteristics; matches must be from the same neighborhood. Column 3 includes matches to females from non-teenage-childbearing families to the teen mothers based on four-year score trends. Column 4 adds other observable control variables from Table 1. Column 5 adds the requirement that the match must be from the same neighborhood at first observation. All outcomes include one weighted observation from the siblings and their controls. Column 1 test scores at t=0 includes multiple observations per individual control.

Table 2 - 8: Mechanisms of the effects of teen birth for siblings

	(1) Test scores at t=0	(2) Repeats grade in t=0 or later	(3) Drops out in t=0 or later	(4) Juvenile justice in t=0 or later	(5) Ever attends any college	(6) Obtains any degree or certificate	(7) Obtains a 4-year degree
Panel A: Sibling interacted with whether the sibling becomes a teen mom (female only)							
Sibling of teen mother	-5.035*	2.352	10.752*	1.082	-9.204	-9.823	-10.890*
	(2.085)	(4.680)	(4.384)	(2.809)	(7.061)	(6.051)	(4.312)
Sibling X Becomes a teen mom	-2.151	23.752^{+}	-9.258	13.237	2.470	1.447	15.447
	(4.835)	(14.223)	(10.473)	(10.310)	(17.470)	(14.548)	(14.151)
N	467	631	631	631	431	431	431
Panel B: Sibling interacted with family composition							
Sibling of teen mother	-2.611	0.672	3.240	8.236*	-9.962	-4.671	-4.558
	(2.153)	(4.714)	(3.060)	(3.618)	(6.200)	(5.191)	(4.230)
Sibling X Has other younger sibling	-0.092	9.580*	7.782^{+}	-4.944^{+}	8.687	0.359	1.136
	(1.807)	(4.103)	(4.393)	(2.861)	(7.014)	(5.950)	(5.148)
Sibling X Has other younger sister	-3.023	-14.619**	-6.311	4.322	-10.014	-5.828	-8.473
	(2.782)	(5.524)	(6.158)	(4.212)	(10.170)	(7.597)	(6.598)
N	835	1147	1147	1147	782	782	782
p(sum of coef.)=0	0.011	0.350	0.293	0.053	0.124	0.065	0.006
Panel C: Sibling interacted with family composition (female only)							
Sibling of teen mother	-4.384	2.756	5.710	6.989^{+}	-11.405	-4.659	-2.925
	(2.714)	(6.141)	(4.407)	(4.069)	(8.485)	(7.700)	(5.766)
Sibling X Has other younger sibling	1.588	11.635^{+}	5.762	-4.839	4.368	-14.245^{+}	-9.931
	(2.412)	(5.924)	(7.547)	(3.300)	(11.339)	(8.262)	(6.510)
Sibling X Has other younger sister	-5.047	-15.979*	-2.620	0.107	-1.315	12.643	3.376
	(3.072)	(7.428)	(9.600)	(3.073)	(14.604)	(10.991)	(8.299)
N	467	631	631	631	431	431	431
p(sum of coef.)=0	0.002	0.777	0.149	0.489	0.372	0.460	0.128

Note: Robust standard errors clustered by family ID. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All analyses based on the trajectory matches with other controls and same neighborhood requirement. Panel A interacts sibling with an indicator for whether the sibling becomes a teen mom later. Panel B interacts sibling with (1) an indicator for having a fellow younger sibling and (2) an indicator for having a fellow younger sister. Includes p-value of a test that the effect for younger siblings with at least one fellow younger sister (the sum of all coefficients) differs from zero. Panel C repeats Panel B for females only.

Table 2 - 9: Time use for siblings and other teenagers

	(1) Sleep	(2) Childcare as primary activity	(3) Any childcare time	(4) School	(5) Work	(6) Other
			Panel A: All Days			
Sibling	16.267	5.499	-18.805	-26.190	3.352	1.072
	(21.543)	(7.317)	(15.518)	(26.764)	(17.133)	(29.596)
Female	-4.509	4.664***	17.061***	12.342*	0.724	-13.221**
	(3.561)	(0.850)	(2.942)	(4.965)	(2.791)	(5.042)
Sister (Sibling X Female)	-13.854	12.588	94.448**	8.259	7.476	-14.469
	(32.043)	(13.270)	(34.621)	(38.033)	(22.911)	(40.448)
Observations	7440	7440	7440	7440	7440	7440
			Panel B: Weekdays			
Sibling	7.038	-0.111	-22.416*	-5.363	11.826	-13.390
	(33.128)	(3.721)	(9.928)	(52.192)	(34.215)	(45.718)
Female	-1.618	5.609***	20.103***	-1.346	-2.313	-0.331
	(4.706)	(1.070)	(3.784)	(7.648)	(3.751)	(6.958)
Sister (Sibling X Female)	-18.773	5.930	123.312**	-22.648	7.719	27.773
	(46.915)	(8.847)	(45.159)	(66.458)	(39.135)	(59.782)
Observations	3554	3554	3554	3554	3554	3554
			Panel C: Weekends			
Sibling	9.791	9.349	-20.600	-0.930	-2.496	-15.714
	(26.201)	(11.756)	(25.150)	(8.891)	(17.821)	(35.385)
Female	-5.449	3.865**	14.211**	20.661***	3.668	-22.745***
	(4.856)	(1.293)	(4.424)	(3.049)	(4.109)	(6.442)
Sister (Sibling X Female)	16.478	22.866	70.076	-28.100*	4.550	-15.795
	(39.419)	(24.446)	(49.561)	(11.763)	(30.290)	(51.276)
Observations	3886	3886	3886	3886	3886	3886

Note: Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Outcomes are time use on various activities in minutes. Any childcare time (Column 3) includes any time spent on childcare as a primary or secondary activity. Other (Column 6) includes any activity not included as sleep, childcare as a primary activity, school, or work. Categories are exclusive except any childcare time (Column 3). Population includes 15-, 16-, and 17-year-old American Time Use Survey respondents who do not live with their own child (i.e., non-parents). Displays the estimated difference in time use (in minutes) for the siblings of teen mothers, female teens, and sisters of teen mothers (the interaction of sibling and female). Excluded category is male teens. Sibling identified as respondents with related, non-own, non-sibling child under the age of five in their household. Analysis controls for family size, children in family, race/ethnicity, metropolitan status, and age. Panels B and C split the analysis by weekdays and weekends (Saturday/Sunday/holiday) time use.

Table 2 - 10: Time spent with various individuals for siblings and other teenagers

	(1) Alone	(2) With parents	(3) With friends	(4) No adult supervision	(5) With a child
		Panel A: All days			
Sibling	6.346	3.073	11.090	28.187	-2.102
	(32.750)	(29.330)	(31.559)	(31.714)	(31.869)
Female	-14.253*	23.332***	-5.341	-13.812**	18.884***
	(5.619)	(4.566)	(5.659)	(4.920)	(4.332)
Sister (Sibling X Female)	32.194	-69.972^{+}	-50.540	-47.441	124.562*
	(45.005)	(36.338)	(40.349)	(44.992)	(60.856)
Observations	7440	7440	7440	7440	7440
		Panel B: Weekdays			
Sibling	79.098*	-31.105^{+}	-0.847	58.831*	-34.359**
	(40.163)	(17.746)	(43.030)	(25.079)	(13.271)
Female	-17.643*	22.517***	-2.874	-15.632**	19.038***
	(7.785)	(4.821)	(7.610)	(5.993)	(5.296)
Sister (Sibling X Female)	-82.665	16.725	-0.232	-71.479	215.036**
	(55.410)	(31.510)	(56.921)	(46.467)	(80.391)
Observations	3554	3554	3554	3554	3554
		Panel C: Weekends			
Sibling	-35.981	9.043	24.993	17.767	13.276
	(41.077)	(44.674)	(44.558)	(47.654)	(51.745)
Female	-13.059^{+}	26.079***	-7.315	-13.771^{+}	18.459**
	(7.701)	(7.188)	(8.275)	(7.417)	(6.734)
Sister (Sibling X Female)	103.501	-117.304*	-101.756^{+}	-56.969	47.229
	(63.934)	(56.395)	(52.823)	(68.701)	(82.245)
Observations	3886	3886	3886	3886	3886

Note: Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Outcomes are time spent with specified groups in minutes. Categories are not exclusive. No adult supervision (Column 4) includes time spent without parents or others over 18 years old. With a child (Column 5) includes time spent with a child under age 14. Population includes 15-, 16-, and 17-year-old American Time Use Survey respondents who do not live with their own child (i.e., non-parents). Displays the estimated difference in time use (in minutes) for the siblings of teen mothers, female teens, and sisters of teen mothers (the interaction of sibling and female). Excluded category is male teens. Sibling identified as respondents with related, non-own, non-sibling child under the age of five in their household. Analysis controls for family size, children in family, race/ethnicity, metropolitan status, and age. Panels B and C split the analysis by weekdays and weekends (Saturday/Sunday/holiday) time use.

Table 2 - 11: Estimated effects of teen birth on outcomes for alternative treatment groups

	(1) Siblings over 18 years old	(2) Pre-scores for siblings (*t*=-4)	(3) Placebo from non-childbearing families
Test scores	---	-1.550	-1.445
		(1.191)	(1.411)
N		664	1034
Repeats grade	---	-0.289	0.376
		(0.543)	(3.192)
N		954	1675
Drops out	---	---	1.825
			(2.518)
N			1675
Juvenile justice	---	-0.095	0.629
		(0.097)	(1.567)
N		954	1675
Ever attends any college	-13.886	---.	-1.324
	(10.199)		(4.001)
N	103		1040
Obtains any degree or certificate	6.233	.---	-0.037
	(11.746)		(3.390)
N	103		1040
Obtains a 4-year degree	2.473	.---	-0.585
	(10.555)		(3.030)
N	103		1040

Note: Robust standard errors clustered by family ID. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All analyses based on the trajectory matches with other controls and same neighborhood requirement. Column (1) conducts the preferred analysis for siblings over 18 years old. Column (2) examines the outcomes for the preferred control group four years before the birth. Column (3) conducts the preferred analysis using 200 randomly selected families without teen childbearing in their family who had been used in the preferred analysis as controls. These former controls are treated as a placebo group and matched to a new set of matches, and the analysis displays the comparison between these placebos and their matched controls using the preferred specification.

Tables – Study 2 Appendix

Table A2 - 1: National percentile pre-trends, by group

	Teen mothers			Siblings		
	Individual FE	Individual & age FE	Individual, age & school FE	Individual FE	Individual & age FE	Individual, age & school FE
4 years before birth	-0.689 (0.804)	-0.668 (0.803)	-0.476 (0.807)	-0.714 (1.447)	0.063 (1.435)	0.015 (1.434)
3 years before birth	-2.015^{*} (0.863)	-2.762^{**} (0.866)	-2.432^{**} (0.864)	-5.709^{***} (1.477)	-4.282^{**} (1.444)	-4.134^{**} (1.447)
2 years before birth	-2.066^{*} (1.048)	-2.788^{**} (1.047)	-2.835^{**} (1.043)	-6.807^{***} (1.538)	-5.161^{***} (1.502)	-4.923^{**} (1.511)
1 years before birth	-5.898^{***} (1.233)	-5.489^{***} (1.249)	-5.879^{***} (1.242)	-8.447^{***} (1.574)	-6.778^{***} (1.548)	-6.704^{***} (1.545)
Observations	405606	405606	405586	405155	405155	405135
N	87030	87030	87026	86918	86918	86914

Note: Robust standard errors clustered by family ID. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Models include mothers and their siblings 1 to 5 years before birth. All children from non-childbearing families are included to estimate age and school fixed effects (included as noted in headings). Year t=-5 is the excluded category.

Table A2 - 2: Logit models predicting becoming a sibling of a teen mother

	First-observed controls		Trajectory controls		
	All	AIC-restricted	Scores only	Scores and controls	AIC-restricted
Female	0.272^{+}	0.266^{+}		0.293*	0.296*
	(0.141)	(0.141)		(0.142)	(0.142)
Age	-0.170***	-0.169***		-0.205***	-0.211***
	(0.035)	(0.035)		(0.042)	(0.041)
# of siblings in data	0.228***	0.228***		0.226***	0.225***
	(0.062)	(0.062)		(0.062)	(0.062)
FRL	1.239***	1.268***		1.163***	1.163***
	(0.198)	(0.195)		(0.199)	(0.199)
Black	1.604***	1.621***		1.495***	1.481***
	(0.160)	(0.152)		(0.162)	(0.158)
First-observed test score	-0.002				
	(0.003)				
First-observed school avg. FRL	2.410*	2.468*		2.241^{+}	2.329*
	(1.150)	(1.111)		(1.145)	(1.109)
First-observed school avg. Black	-0.556			-0.500	
	(1.175)			(1.175)	
First-observed school avg. first-observed test score	0.034	0.038^{+}		0.035	0.039^{+}
	(0.026)	(0.022)		(0.025)	(0.023)
Scores, year t=-2			-0.021***	-0.011*	-0.011*
			(0.005)	(0.005)	(0.005)
Scores, year t=-3			-0.017**	-0.013*	-0.013*
			(0.006)	(0.006)	(0.006)
Scores, year t=-4			0.009^{+}	0.014*	0.014*
			(0.005)	(0.005)	(0.005)
Missing score data, year t=-2			-0.530*	-0.151	
			(0.259)	(0.267)	
Missing score data, year t=-3			-1.570**	-1.160*	-1.132^{+}
			(0.585)	(0.589)	(0.587)
Missing score data, year t=-4			-0.170	-0.474**	-0.467**
			(0.165)	(0.180)	(0.180)
Constant	-8.834***	-9.327***	-4.855***	-7.748***	-8.086***
	(2.045)	(1.805)	(0.161)	(2.079)	(1.843)
Observations	128576	128576	128576	128576	128576

Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Model predicts probability of becoming a sibling of a teenage mother, among the younger siblings of teenage mothers and younger siblings from non-teenage-childbearing families. Requires at least 2 of 3 prior observations. Final column limits variables by minimizing the Akaike Information Criterion (AIC).

Table A2 - 3: Descriptive statistics for outcome variables for sibling matched controls

	(1) Baseline	(2) Female	(3) Male	(4) Non-black	(5) Black	(6) Non-FRL	(7) FRL
Test scores at *t*=0	44.248	45.255	42.977	55.431	34.167	61.173	39.387
	(25.286)	(24.290)	(26.474)	(24.155)	(21.830)	(25.214)	(23.141)
Repeats grade in *t*=0 or later	30.174	26.232	34.979	27.391	32.404	18.182	33.415
	(45.926)	(44.033)	(47.746)	(44.652)	(46.846)	(38.664)	(47.201)
Drops out in *t*=0 or later	13.250	12.852	13.734	12.174	14.111	7.727	14.742
	(33.921)	(33.500)	(34.461)	(32.739)	(34.847)	(26.768)	(35.477)
Juvenile justice in *t*=0 or later	7.834	5.810	10.300	5.435	9.756	2.727	9.214
	(26.884)	(23.416)	(30.432)	(22.698)	(29.700)	(16.328)	(28.942)
Ever attends any college	54.980	58.621	50.459	53.353	56.410	69.767	50.446
	(49.789)	(49.319)	(50.082)	(49.971)	(49.655)	(46.072)	(50.048)
Obtains any degree or certificate	26.330	30.542	21.101	28.863	24.103	44.767	20.677
	(44.076)	(46.122)	(40.871)	(45.389)	(42.829)	(49.883)	(40.539)
Obtains a 4-year degree	17.599	19.212	15.596	20.117	15.385	31.977	13.191
	(38.110)	(39.451)	(36.343)	(40.155)	(36.130)	(46.786)	(33.873)

Note: Mean coefficients; SE in parentheses (clustered by family ID) for noted population from the trajectory matches based on trajectory test scores, other controls, and same neighborhood requirement.

Table A2 - 4: Estimated effects of teen birth on alternative test score measures for siblings

	(1) Z-scores	(2) Mean scores, no imputation	(3) Math only	(4) Reading only	(5) Using minimum replacement for dropouts	(6) Using minimum replacement for dropouts, with quantile regression
Test scores at *t*=0	-0.127*	-4.526**	-5.081**	-3.040	-3.803*	-2.956
	(0.050)	(1.567)	(1.818)	(1.896)	(1.582)	(4.397)
N	835	799	825	824	843	843
Test scores in grade 10	-0.130*	-2.040	-2.904	-3.324	-3.644+	-3.812
	(0.058)	(2.456)	(2.313)	(2.257)	(1.905)	(4.785)
N	454	314	447	443	454	485
Test scores, with age and individual FE	-0.079*	-2.865**	-3.595**	-1.623	-2.554*	---
	(0.033)	(1.067)	(1.324)	(1.169)	(1.040)	
Observations	6761	6379	6630	6653	6761	
N	1074	1074	1074	1074	1074	

Note: Robust standard errors clustered by family ID except for quantile regression. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All analyses based on the trajectory matches with other controls and same neighborhood requirement. Row 1 outcomes are mean national percentile rank for math and reading test scores in the year of birth. Row 2 outcomes are mean national percentile rank for math and reading test scores in tenth grade, for those who were in tenth grade or earlier in the year of the birth. Row 3 conducts a fixed effects analysis including all observations from the siblings and their matched controls, excluding *t*=-1. Column 1 translates the mean national percentile rank into a Z-score with a mean of 0 (at 50 in the distribution) and a SD of 1. Column 2 does not use imputation from the current-year FCAT scores for those missing their national percentile rank in each year. Columns 3 and 4 separate the analysis by math and reading, respectively. Column 5 replaces all students with missing test score due to dropping out of high school with the minimum score. Column 6 conducts the analysis as a quantile regression at the median.

Table A2 - 5: Logit models predicting teen motherhood

	First-observed controls		Trajectory controls		
	All	AIC-restricted	Scores only	Scores and controls	AIC-restricted
Age	0.251***	0.252***		0.283***	0.276***
	(0.024)	(0.024)		(0.031)	(0.029)
# of siblings in data	0.072			0.075	
	(0.058)			(0.058)	
Oldest	0.303**	0.275**		0.328**	0.296**
	(0.106)	(0.103)		(0.107)	(0.104)
FRL	0.824***	0.836***		0.719***	0.732***
	(0.140)	(0.138)		(0.142)	(0.139)
Black	1.573***	1.576***		1.429***	1.442***
	(0.122)	(0.122)		(0.124)	(0.124)
First-observed test score	-0.017***	-0.017***			
	(0.002)	(0.002)			
First-observed school avg. FRL	-0.173			-0.385	
	(0.718)			(0.716)	
First-observed school avg. Black	-1.540^{+}	-1.515^{+}		-1.478^{+}	-1.418^{+}
	(0.851)	(0.838)		(0.861)	(0.853)
First-observed school avg. first-observed test score	-0.035*	-0.032***		-0.035*	-0.028**
	(0.016)	(0.009)		(0.016)	(0.009)
Scores, year t=-2			-0.012*	-0.005	
			(0.005)	(0.005)	
Scores, year t=-3			-0.011*	-0.006	-0.009*
			(0.005)	(0.005)	(0.004)
Scores, year t=-4			-0.009^{+}	-0.007	-0.007
			(0.005)	(0.005)	(0.005)
Scores, year t=-5			-0.013**	-0.008^{+}	-0.009^{+}
			(0.005)	(0.005)	(0.005)
Missing score data, year t=-2			0.360**	-0.077	
			(0.120)	(0.129)	
Missing score data, year t=-3			-0.910***	-1.192***	-1.189***
			(0.199)	(0.202)	(0.202)
Missing score data, year t=-4			-0.415**	-0.459**	-0.459**
			(0.152)	(0.149)	(0.148)
Missing score data, year t=-5			-0.700***	-0.208	-0.201

			(0.132)	(0.136)	(0.136)
Constant	-7.707***	-7.778***	-3.296***	-7.367***	-7.679***
	(1.332)	(0.732)	(0.120)	(1.379)	(0.786)
Observations	130811	130811	130811	130811	130811

Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Model predicts the probability of becoming a teenage mother, among the eventual teenage mothers and females from non-teenage-childbearing families. Requires at least 2 of 4 prior observations. Final column limits variables by minimizing the Akaike Information Criterion (AIC).

Tables – Study 3

Table 3 - 1: Descriptive statistics

	Mean	SD	Minimum	Maximum	Weekday Cronbach's Alpha	N
Black	0.17	0.38	0.00	1.00		82
Hispanic white	0.20	0.40	0.00	1.00		82
Non-Hispanic white	0.18	0.39	0.00	1.00		82
Multiethnic	0.27	0.45	0.00	1.00		82
Other	0.18	0.39	0.00	1.00		82
Age	14.90	1.86	11.27	18.11		82
Male	0.51	0.50	0.00	1.00		82
Violent crime ever=1	0.51	0.50	0.00	1.00		82
Mean bedtime	23.12	1.07	20.17	26.12	0.72	82
Mean latency	0.43	0.47	0.04	3.03	0.44	82
Mean sleep duration	6.54	0.87	3.98	9.13	0.61	82
Mean wake time	7.00	0.98	4.89	10.33	0.76	82
Mean waking cortisol	0.27	0.19	0.01	1.48	0.88	76
Mean CAR	0.12	0.17	-0.23	0.65	0.53	74
Mean bedtime cortisol	0.07	0.07	0.00	0.37	0.66	76

Sleep outcomes measured in hours. Cortisol outcomes measured in μg/dl. Includes Cronbach's alpha of outcome measures during the week.

Table 3 - 2: Correlation matrix for outcome measures

	Bedtime	Latency	Sleep duration	Wake time	Bedtime cortisol	Waking cortisol	CAR
Bedtime	1.000						
Latency	-0.261	1.000					
Sleep duration	-0.672	-0.106	1.000				
Wake time	0.303	0.073	0.328	1.000			
Bedtime cortisol	-0.073	0.066	0.002	0.074	1.000		
Waking cortisol (next day)	-0.112	0.056	0.206	0.227	0.222	1.000	
CAR (next day)	0.145	-0.230	-0.204	-0.235	-0.154	-0.457	1.000

Correlations between outcomes measured in the main analysis. Note that some measures are mathematically related to each other.

Table 3 - 3: Effect of acute violence on sleep measures

	Bedtime	Latency	Sleep Duration	Wake Time
	Panel A: Overall violent crime			
Crime	0.438*	0.017	-0.293	0.176
	(0.213)	(0.067)	(0.259)	(0.195)
	Panel B: Excluding weekends			
Crime	0.629*	-0.017	-0.402	0.105
	(0.264)	(0.082)	(0.298)	(0.217)
	Panel C: Type of violent crime			
Assault	0.645*	-0.062	-0.429	0.129
	(0.312)	(0.080)	(0.377)	(0.248)
Criminal Sexual Assault	0.686	0.499*	-1.042**	0.050
	(1.075)	(0.220)	(0.327)	(0.223)
Homicide	1.811***	-0.053	-1.138***	0.242
	(0.198)	(0.090)	(0.257)	(0.206)
Robbery	0.164	0.073	-0.073	0.235
	(0.272)	(0.107)	(0.260)	(0.254)
	Panel D: Excluding robbery from violent crime			
Crime, Excluding Robbery	0.649*	-0.030	-0.484	0.081
	(0.294)	(0.080)	(0.332)	(0.217)
N Adolescents	82	82	82	82
Observations	281	281	281	281

Robust standard errors clustered at the individual level and included in parentheses. Models include individual and day of the week fixed effects. Outcomes measured in hours. Listed N and Observations for all panels except Panel B; when excluding weekends N=77 and Observations=211. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3 - 4: Effect of acute violence on cortisol measures

	Same-day Bedtime Cortisol	Next-day Waking Cortisol	Next-day CAR
Panel A: Overall violent crime			
Crime	0.005	-0.048	0.133*
	(0.019)	(0.038)	(0.058)
Panel B: Excluding weekends			
Crime	0.005	-0.076	0.174*
	(0.022)	(0.049)	(0.042)
Panel C: Type of violent crime			
Assault	-0.019	-0.085^{+}	0.221*
	(0.028)	(0.044)	(0.085)
Criminal Sexual Assault	0.177*	0.056	0.190
	(0.069)	(0.119)	(0.119)
Homicide	0.033^{+}	-0.038	0.453***
	(0.018)	(0.057)	(0.076)
Robbery	0.019	-0.020	0.019
	(0.023)	(0.041)	(0.051)
Panel D: Excluding robbery from violent crime			
Crime, Excluding Robbery	-0.001	-0.064	0.218**
	(0.024)	(0.042)	(0.071)
Panel E: Controlling for sleep measures			
Crime	0.003	-0.043	0.140^{+}
	(0.016)	(0.043)	(0.071)
Sleep duration	-0.015*	0.025**	-0.030*
	(0.007)	(0.009)	(0.012)
Wake time	0.021*	-0.015	-0.012
	(0.010)	(0.017)	(0.018)
N Adolescents	76	76	74
Observations	283	281	244

Robust standard errors clustered at the individual level and included in parentheses. Models include individual and day of the week fixed effects. Outcomes measured in µg/dl. Listed N and Observations for all panels except Panels B and E. When excluding weekends N=76 and Observations=232 for bedtime cortisol, N=75 and Observations=215 for waking cortisol, and N=73 and Observations=188 for CAR. When including sleep measures N=74 and Observations=229 for bedtime cortisol, N=76 and Observations=241 for waking cortisol, and N=72 and Observations=212 for CAR. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Tables – Study 3 Appendix

Table A3 - 1: Effect of acute violence on sleep measures

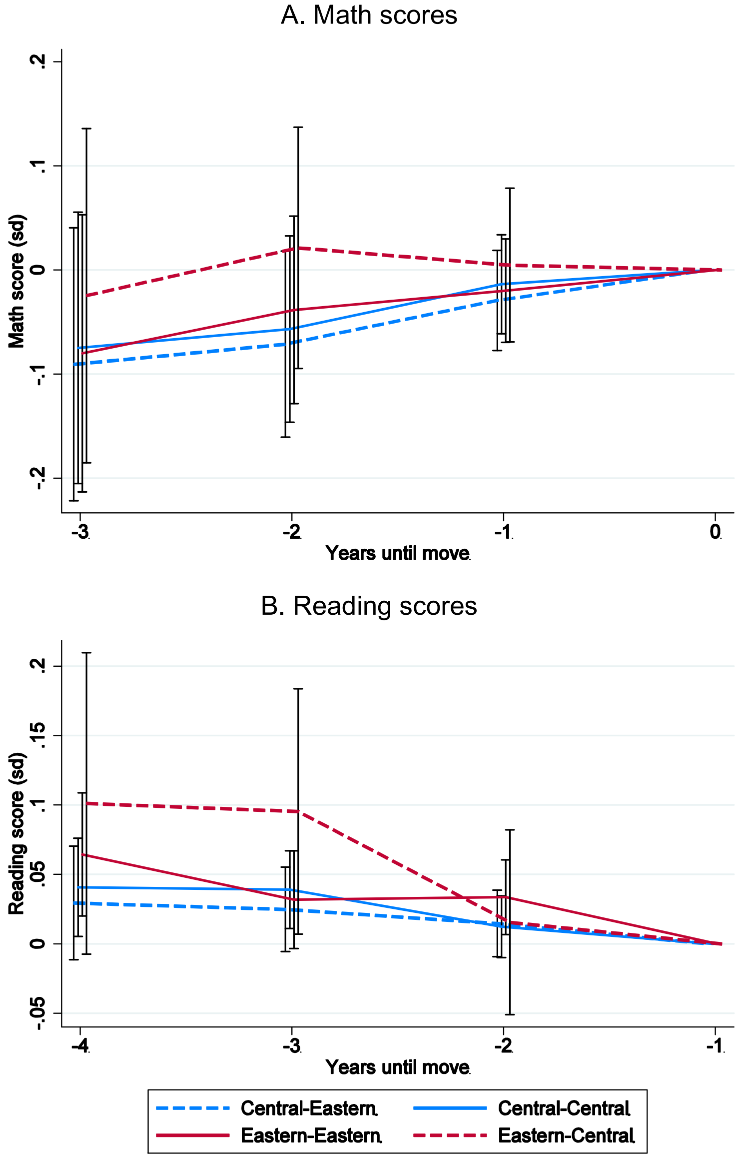
	Bedtime	Latency	Sleep Duration	Same-day Bedtime Cortisol	Next-day Waking Cortisol	Next-day CAR
	Panel A: Difference by neighborhood violence					
Crime Effect	0.629	0.010	-0.789	0.009	-0.142^{+}	0.227
	(0.401)	(0.073)	(0.569)	(0.020)	(0.082)	(0.143)
Crime X Above Median	-0.295	0.011	0.765	-0.007	0.141	-0.143
Neighborhood	(0.439)	(0.099)	(0.603)	(0.032)	(0.090)	(0.151)
p(sum=0)	0.135	0.807	0.916	0.909	0.966	0.081
	Panel B: Differences by pubertal status					
Crime	0.874	-0.163	-0.678	-0.033	-0.111*	0.046
	(0.576)	(0.137)	(0.474)	(0.036)	(0.052)	(0.094)
Crime X Puberty	-0.293	0.072	0.247	0.022	0.039	0.032
	(0.283)	(0.075)	(0.215)	(0.022)	(0.023)	(0.054)
*p(*highest stage=0)	0.986	0.686	0.838	0.395	0.924	0.132
p(lowest stage=highest stage)	0.303	0.344	0.254	0.327	0.105	0.551
	Panel C: Differences by time of day					
Crime	0.589*	0.039	-0.508^{+}	0.010	-0.046	0.112*
	(0.284)	(0.098)	(0.302)	(0.025)	(0.048)	(0.051)
Early X Crime	-0.303	-0.043	0.430	-0.009	-0.004	0.038
	(0.372)	(0.114)	(0.379)	(0.026)	(0.055)	(0.081)
p(sum=0)	0.303	0.957	0.816	0.954	0.271	0.077
N Adolescents	82	82	82	76	76	74
Observations	281	281	281	283	281	244

Robust standard errors clustered at the individual level and included in parentheses. Models include individual and day of the week fixed effects. Outcomes measured in hours for sleep and in µg/dl for cortisol. High-crime areas defined as those that had an above-median number of violent crimes in their police beat in the months of the study. "Puberty" is the mean on the four-point Peterson pubertal scale for each person, with lower numbers indicating earlier stages of puberty. Early crime defined as before 4:00 p.m. (the median crime time). $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figures

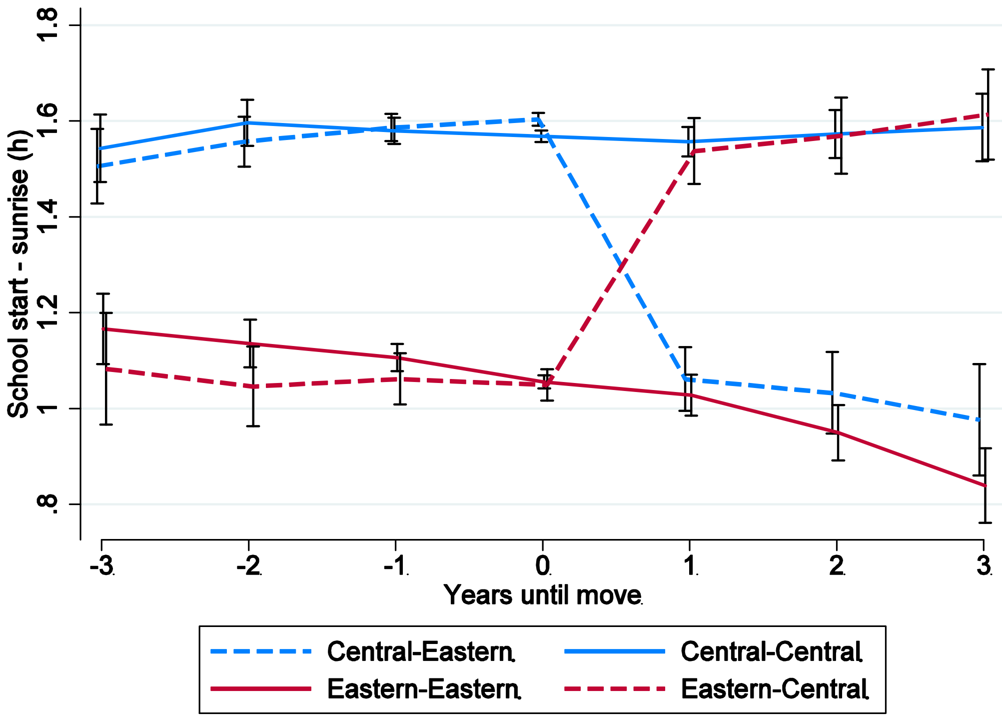
Figures – Study 1

Figure 1 - 1: Pre-move trends in academic outcomes, by mover type



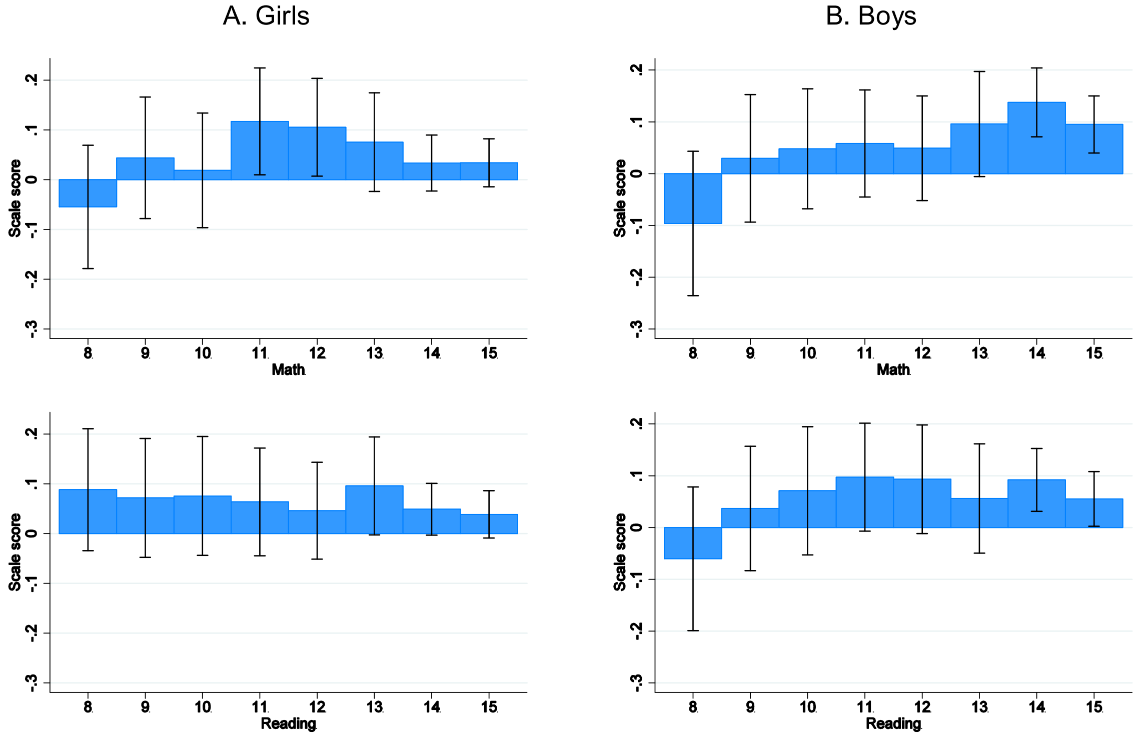
Displays the pre-move achievement trends for the four years leading up to a move of 25 miles or more. Results reported separately for four groups of movers: within CT, within ET, ET to CT, and CT to ET. Coefficients recovered from a regression of test scores on time-until-move dummies, a vector of controls (age-gender dummies, longitude, and school population shares for FRL, male, black, Asian, and Hispanic), and a fixed effect for the period before the move. Standard errors are clustered at the individual level, and included as bars representing 95% confidence intervals.

Figure 1 - 2: Hours of sunlight before school over move, by mover type



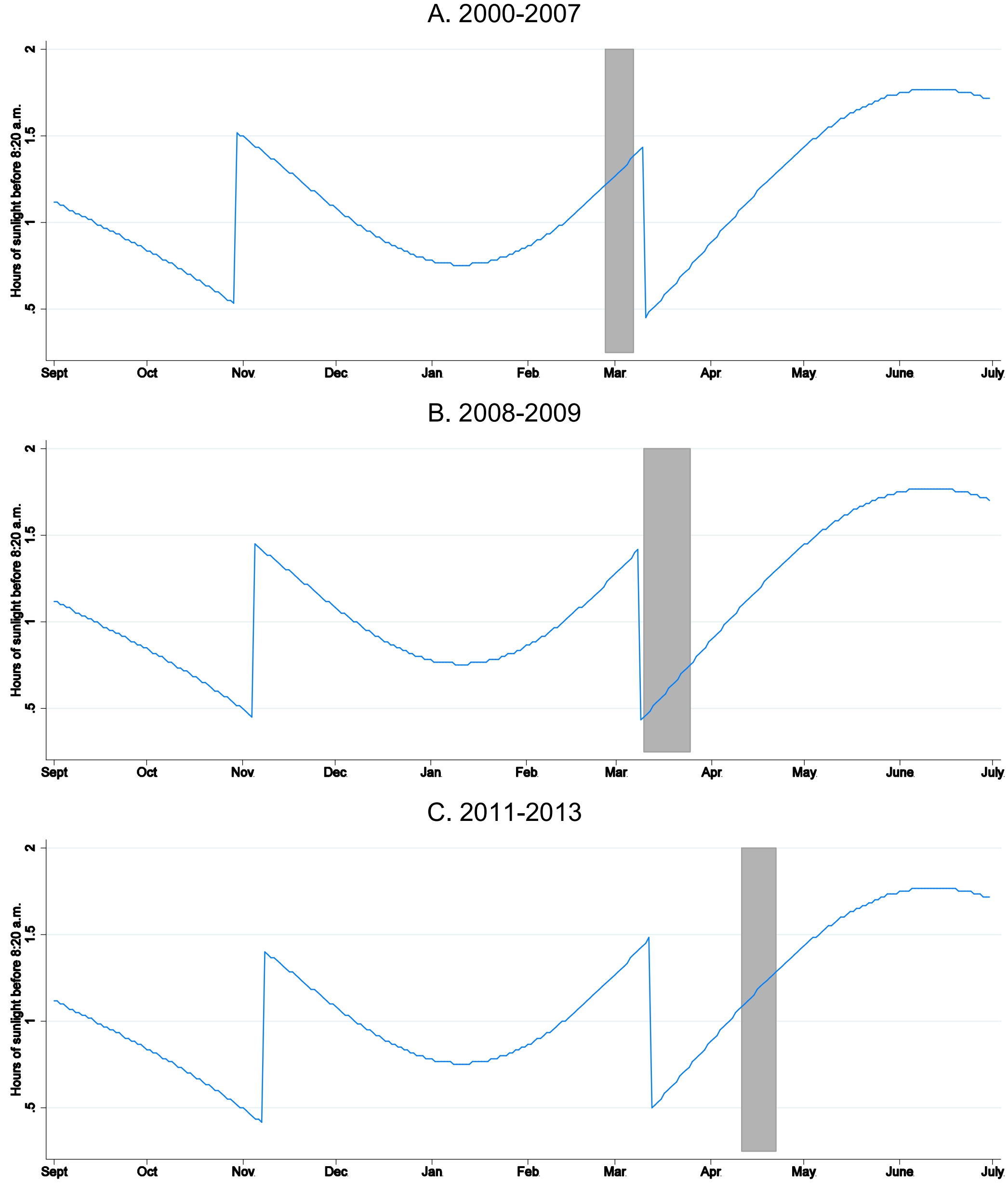
Displays the hours of sunlight before school for four groups: within CT, within ET, ET to CT, and CT to ET. Estimates are from a regression of relative school start time on time relative to move for each mover group, a vector of controls (age-gender dummies, longitude, and school population shares for FRL, male, black, Asian, and Hispanic), and a student-move fixed effect. The year before the move is normalized to be zero; we adjust the level of the coefficients with the group mean of relative start times for one year before the move. Standard errors are clustered at the individual level, and included as bars representing 95% confidence intervals.

Figure 1 - 3: Effect of school start times on academic achievement, by age, gender, and subject



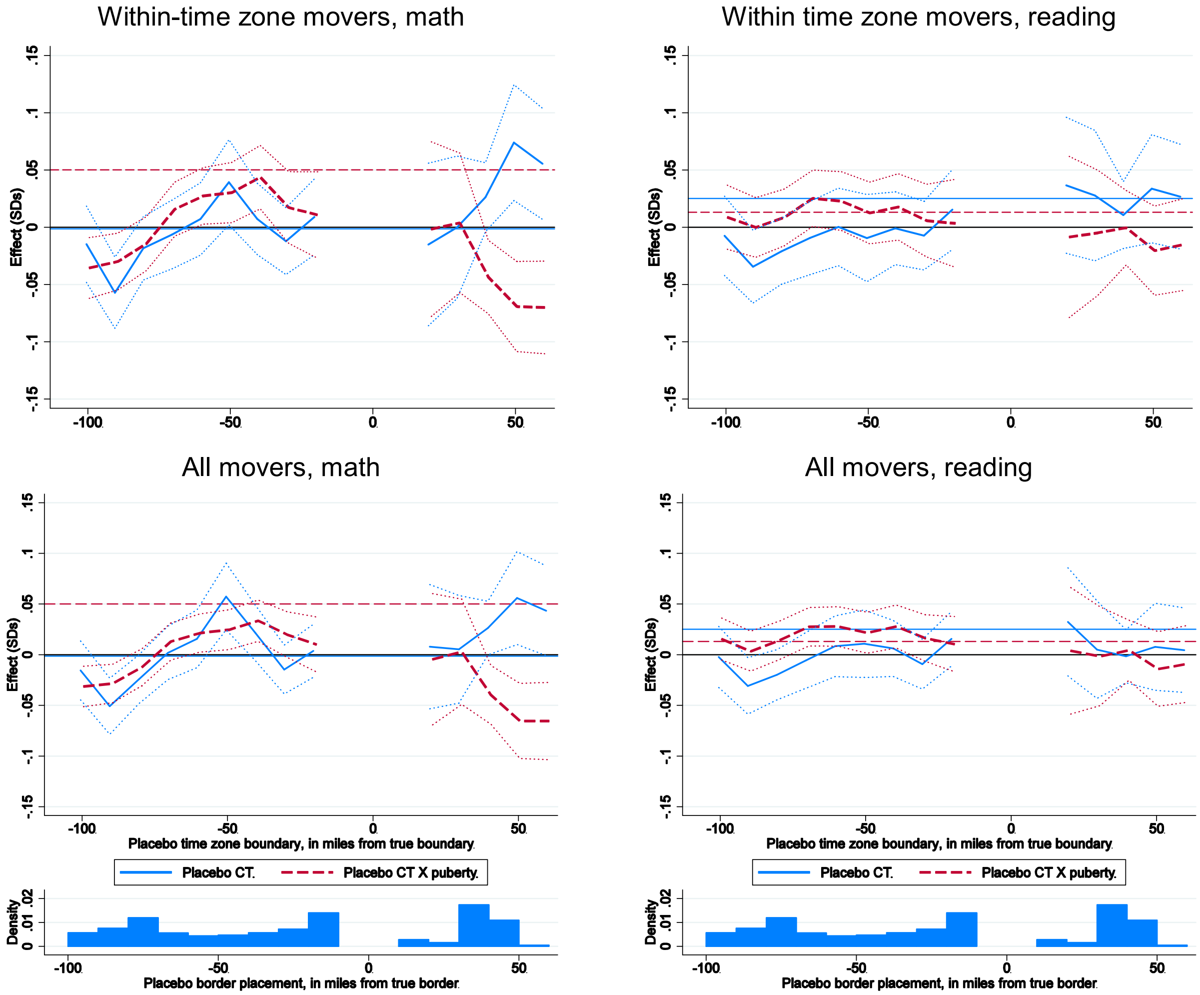
Each subfigure displays the age-gender specific effect of start times on academic achievement. Coefficients are from a regression of scale scores on school start time interacted with age, a vector of controls (age-gender dummies, longitude, and school population shares for FRL, male, black, Asian and Hispanic), and an individual fixed effect. Start time-age interactions are instrumented with time zone-age interactions. Sample is listed in the column headers, dependent variable is noted on the horizontal axis. Standard errors are clustered at the individual level, and included as bars representing 95% confidence intervals.

Figure 1 - 4: Hours of sunlight before 8:20 a.m. start time, by year with testing periods



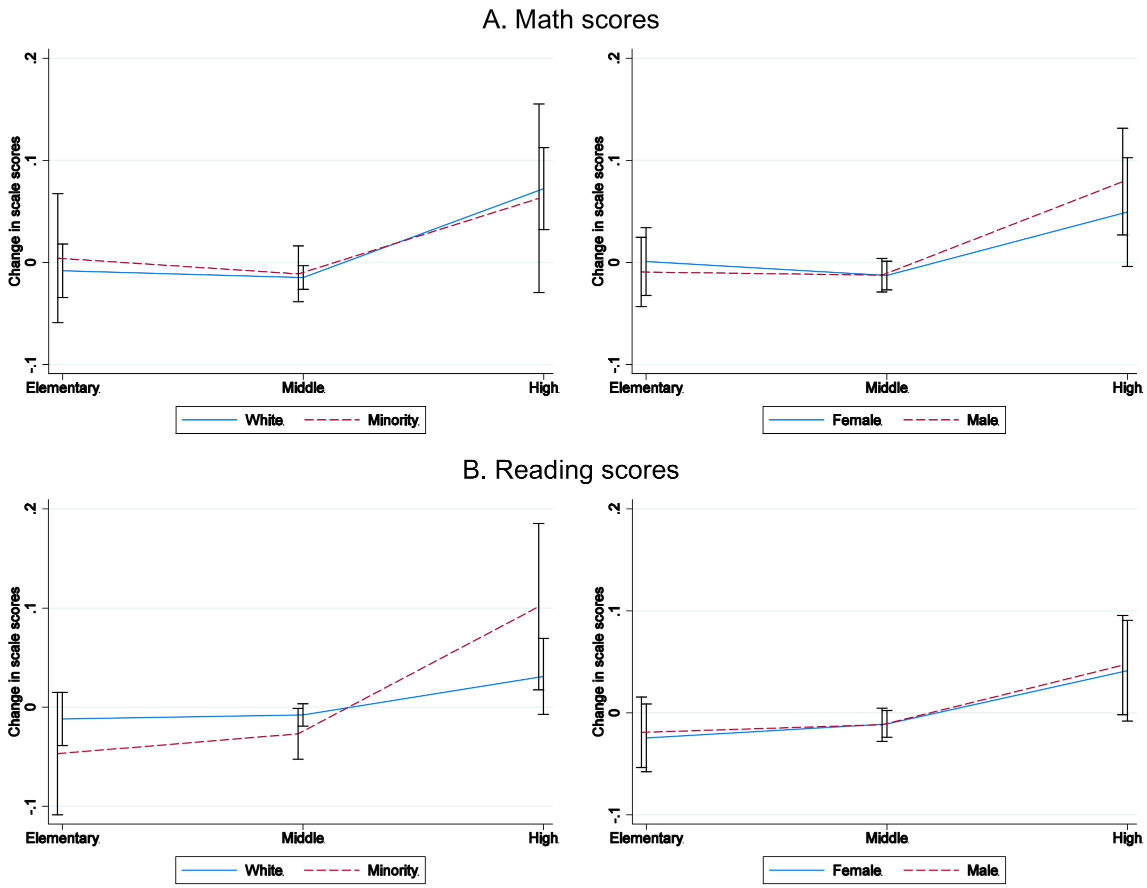
Amount of sunlight before school and testing dates for a hypothetical school for each of the three testing regimes. School location and opening time chosen to match the average test-day relative start time in ET in 2008. Grey areas represent testing periods. The figures display sunlight for 2007, 2008, and 2011, respectively, but all are archetypes of their era.

Figure 1 - 5: Effect of placebo time zones on academic achievement



Dependent variable as noted in panel heading. Test scores measured in SDs normalized at the grade-year level for the entire state. Thin horizontal lines represent baseline coefficient estimates. We generate placebo time zones in ten mile increments from the true time zone boundary. Then, placebo coefficients are calculated from individual regressions of the outcome on the true time zone interacted with puberty, and the placebo time zone interacted with puberty. All specifications include age- gender dummies, longitude controls, school demographic means (FRL, male, black, Asian, and Hispanic) and individual fixed effects. Standard errors clustered at the individual level. We display results including and excluding cross-time zone movers. Sample excludes a 25 mile donut around the time zone boundary due to treatment bleed across the boundary.

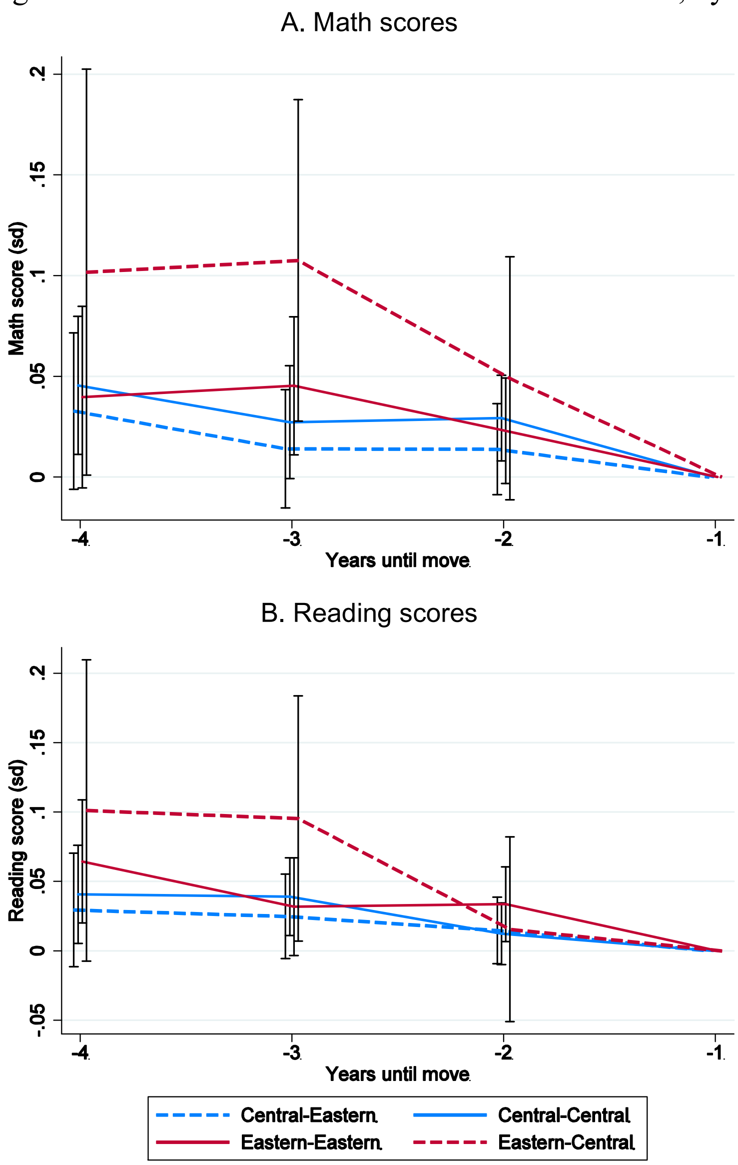
Figure 1 - 6: Counterfactual change in test scores, reordered start times



Estimated test score gains under a counterfactual policy where start times are adjusted to be later for older children. Adjustment is conducted by taking the average start time for each school type in each district (elementary, middle, and high), and swapping them between school types so that elementary schools open first, then middle schools, then high schools. We then adjust the level of all school times so that the mean counterfactual district start time is the same as the true mean start time. This results in bell times 22 minutes earlier for elementary schools, 13 minutes earlier for middle schools, and 44 minutes later for high schools. Gains are then calculated by multiplying the changes in start time for each child with the relevant coefficients from Table 1-3. Bars represent 95% confidence intervals.

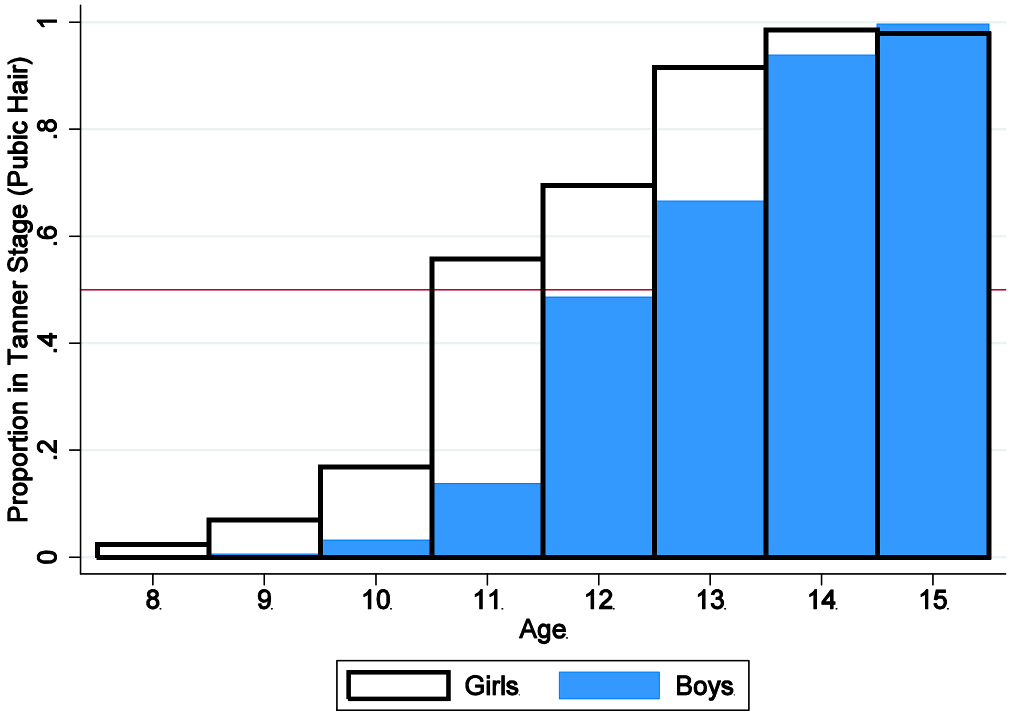
Figures – Study 1 Appendix

Figure A1 - 1: Pre-move trends in academic outcomes, by mover type without additional controls



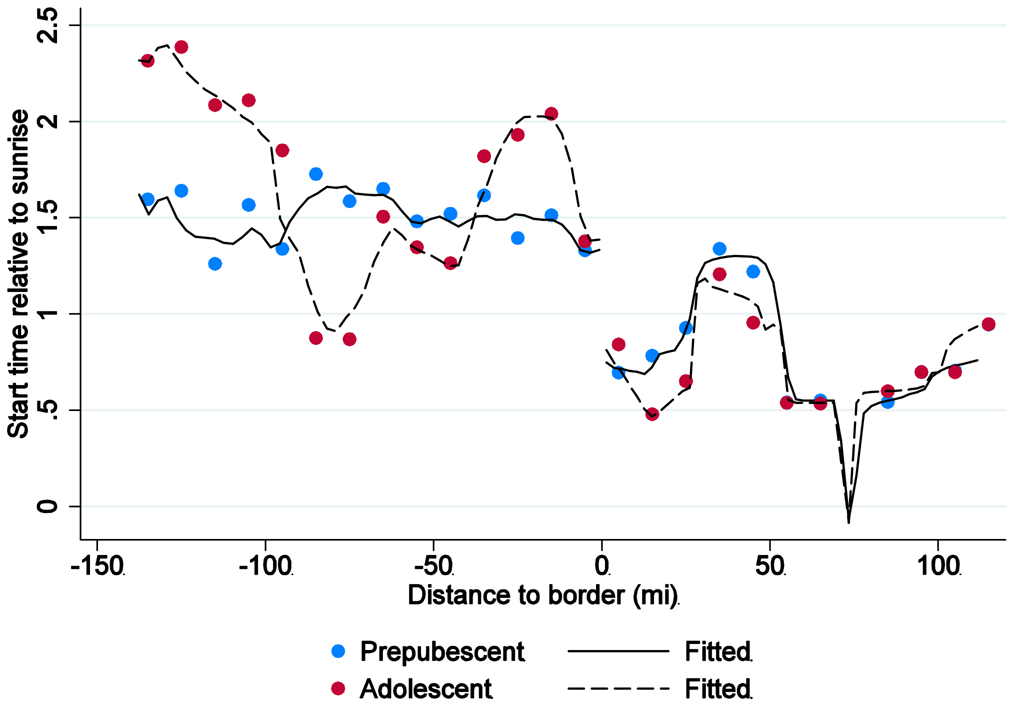
Displays the pre-move achievement trends for the four years leading up to a move of 25 miles. Results reported separately for four groups of movers: within CT, within ET, ET to CT, and CT to ET. Coefficients recovered from a regression of test scores on time- until-move dummies and a fixed effect for the period before the move. Standard errors are clustered at the individual level, and included as bars representing 95% confidence intervals.

Figure A1 - 2: Tanner Stage 3 proportions by age and sex



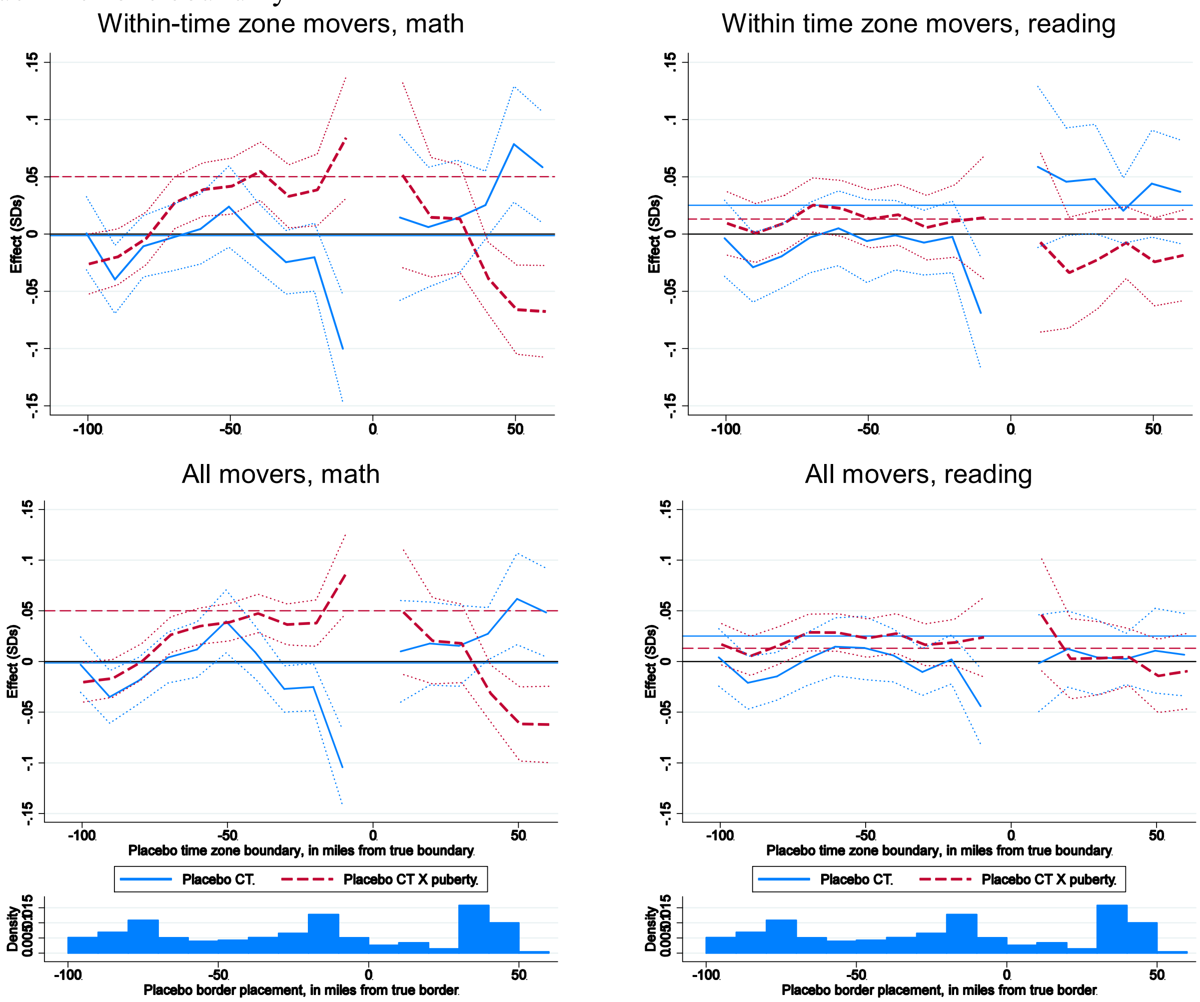
Displays proportion of children who had entered the Tanner Stage for pubic hair development at a given age for males and females. Horizontal line represents median child entering the stage.

Figure A1 - 3: Relative start time near the time zone boundary



Displays a nonparametric regression of relative start time (start time minus sunrise) on distance to the time zone boundary, estimated separately for each time zone. Scatter points are ten mile bin averages.

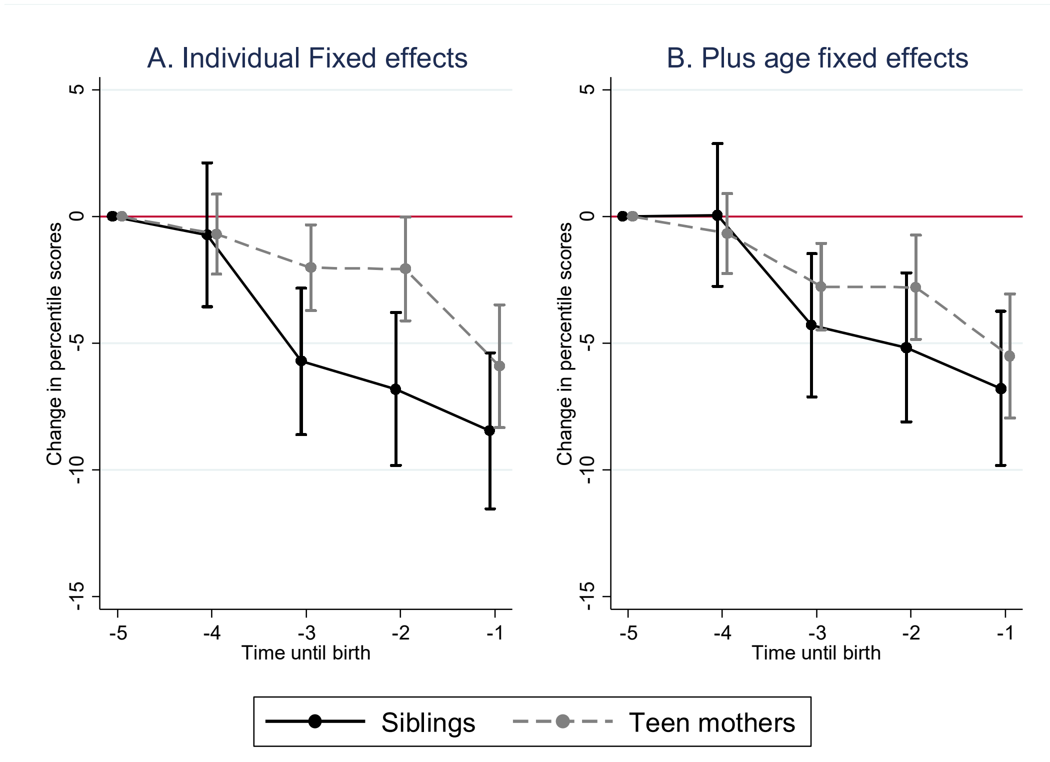
Figure A1 - 4: Effect of placebo time zones on academic achievement, no sample exclusion near true time zone boundary



Dependent variable as noted in panel heading. Test scores measured in SDs normalized at the grade-year level for the entire state. Thin horizontal lines represent baseline coefficient estimates. We generate placebo time zones in ten mile increments from the true time zone boundary. Then, placebo coefficients are calculated from individual regressions of the outcome on the true time zone interacted with puberty, and the placebo time zone interacted with puberty. All specifications include age- gender dummies, longitude controls, school demographic means (FRL, male, black, Asian, and Hispanic) and individual fixed effects. Standard errors clustered at the individual level. We display results including and excluding cross-time zone movers.

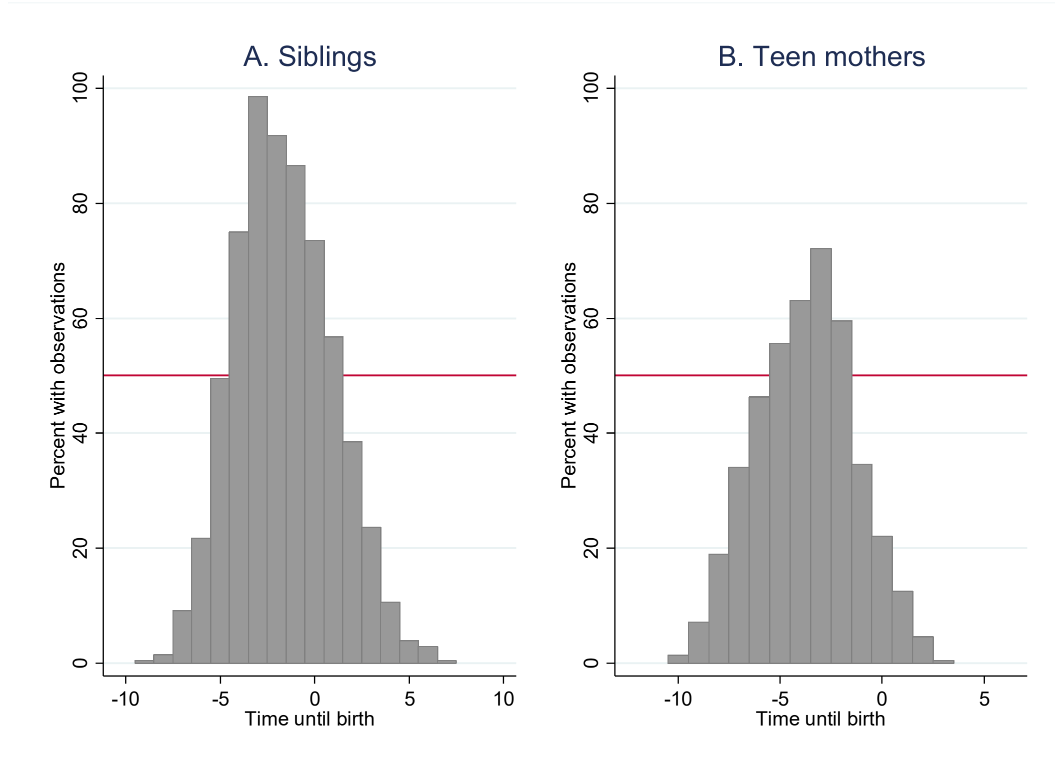
Figures – Study 2

Figure 2 - 1: Trajectory pre-trends for teen mothers and siblings



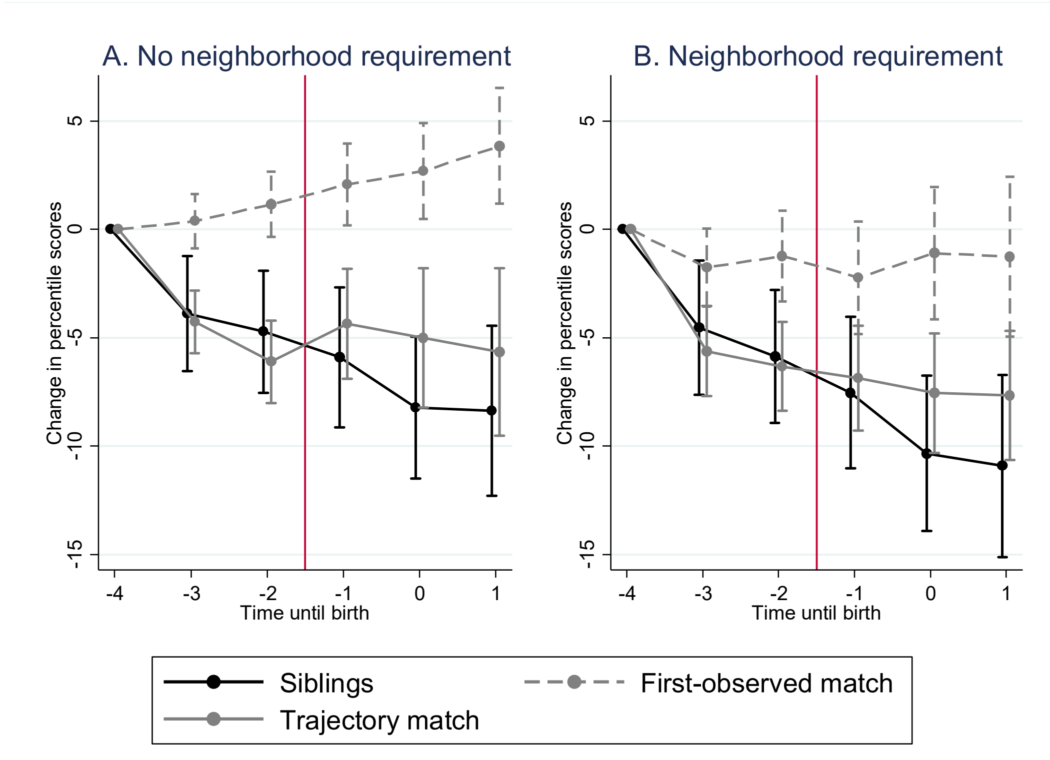
Note: Teen mothers include all teen mothers from families of two or more where the mother gave birth at age 15-17. Siblings include all younger siblings from families where an older sister gave birth at age 15-17. Panel A estimates based on a regression of mean percentile rank (of math and reading test scores) on years until birth with person fixed effects within the noted population, with standard errors clustered by family ID. Panel B adds age fixed effects.

Figure 2 - 2: Available pre- and post-birth grades for teen mothers and siblings



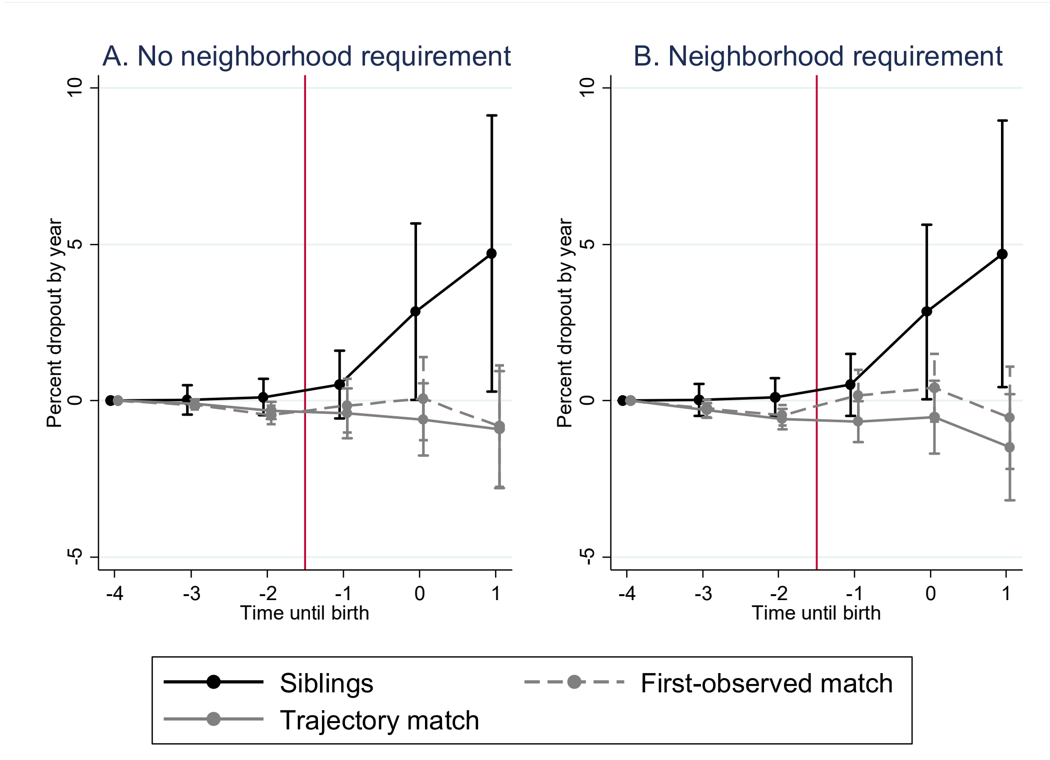
Note: Proportion of siblings and teen mothers with test score data available by years relative to birth for the trajectory matching population.

Figure 2 - 3: National percentile pre- and post-trends, by group for siblings



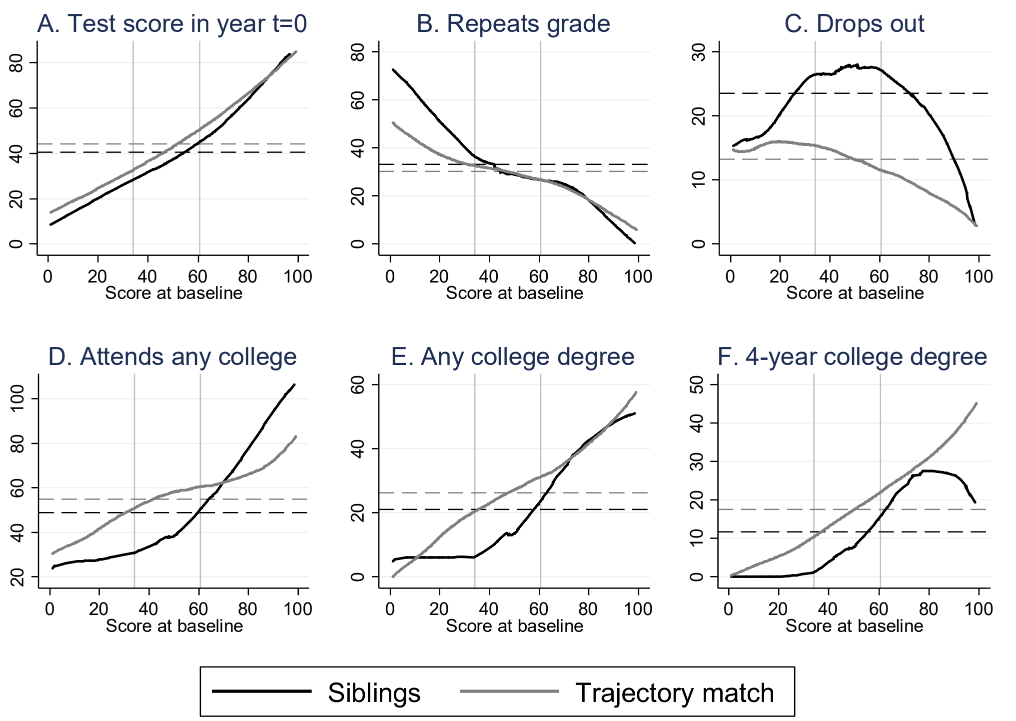
Note: Siblings include all siblings from families where an older sister gave birth at age 15-17. First-observed matches include matches from non-teenage-childbearing families to the siblings based on first-observed characteristics. Trajectory matches include matches from non-teenage-childbearing families to the siblings based on three-year test score trends characteristics and first-observed characteristics. Estimates based on a regression of mean percentile rank on years relative to birth (or time relative to the match year for the matches) with person and age fixed effects within the noted population. Panel B requires the matches to be from the same neighborhood at first observation as the siblings; Panel A does not. The vertical line marks the end of the trajectory matching; points to the right allowed to differ from matches.

Figure 2 - 4: Dropout rates pre- and post-trends, by group for siblings



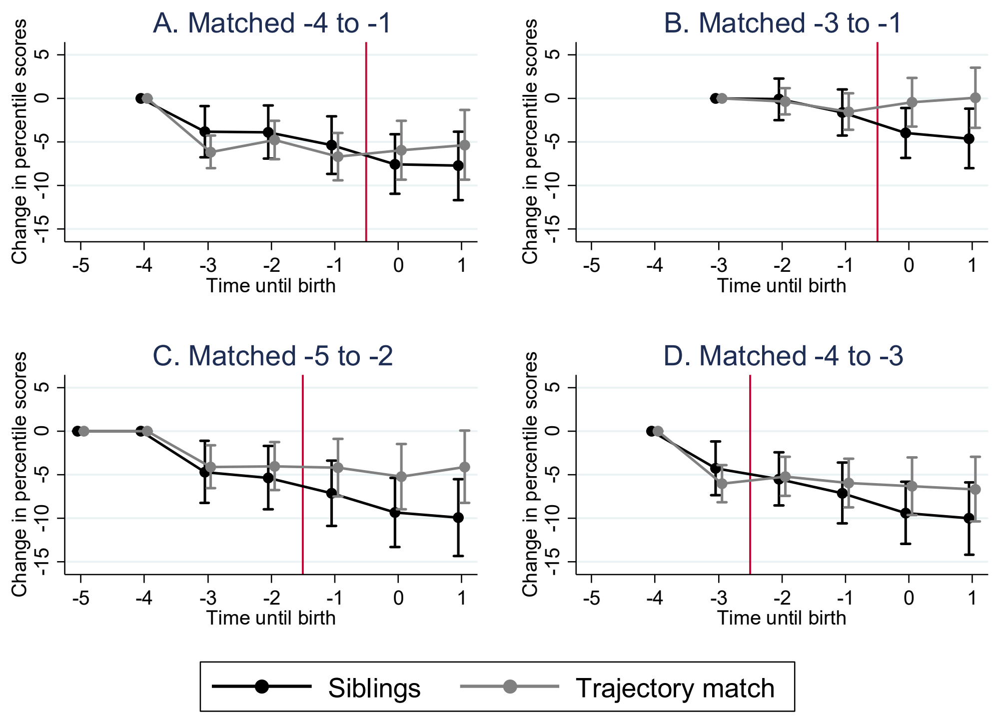
Note: Siblings include all siblings from families where an older sister gave birth at age 15-17. First-observed matches include matches from non-teenage-childbearing families to the siblings based on first-observed characteristics. Trajectory matches include matches from non-teenage-childbearing families to the siblings based on three-year test score trends characteristics and first-observed characteristics. Estimates based on a regression of dropout on years relative to birth (or time relative to the match year for the matches) with person and age fixed effects within the noted population. Panel B requires the matches to be from the same neighborhood at first observation as the siblings; Panel A does not. Vertical line marks the end of the trajectory matching; points to the right allowed to differ from matches.

Figure 2 - 5: Distribution of outcomes following birth for siblings and matched controls



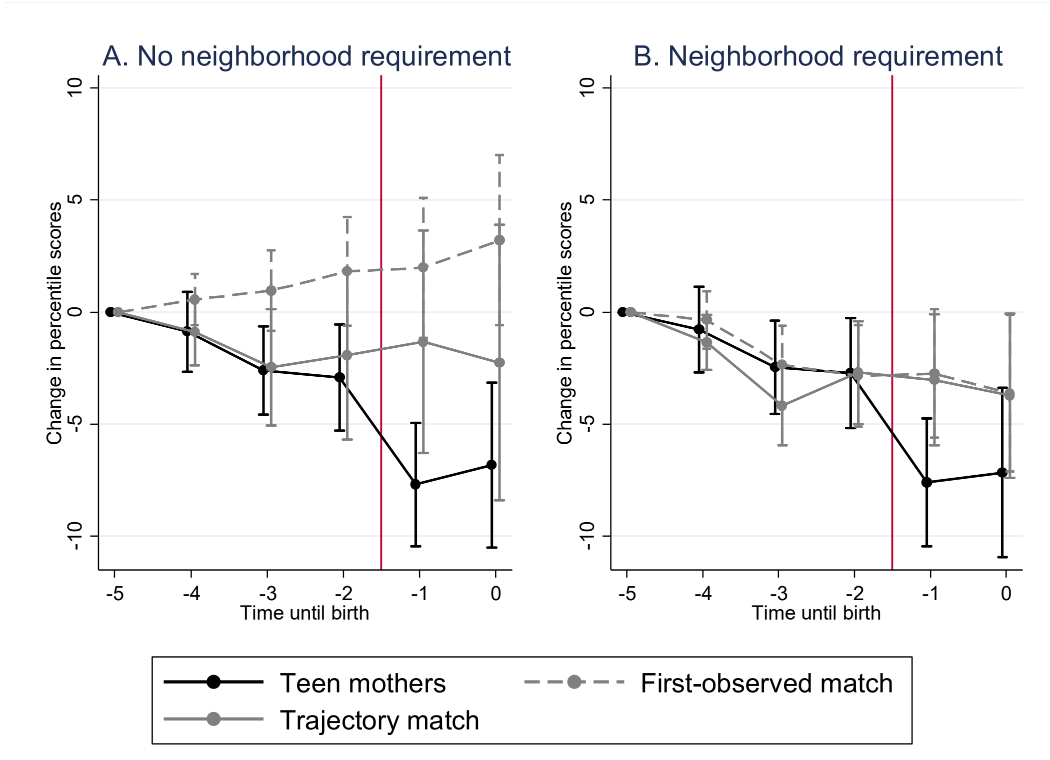
Note: Displays a locally weighted regression of the outcome (y-axis, noted in the panel header) and first-observed test scores (x-axis) using Stata's *lowess* command. Siblings include all siblings from families where an older sister gave birth at age 15-17. Trajectory matches include matches from non-teenage-childbearing families to the siblings based on prior-year test score trends, other observable characteristics, and same-neighborhood requirements. Dashed horizontal lines represent the unadjusted overall mean for the outcome for siblings (black dashed line) and their trajectory matches (gray dashed line). Vertical light gray lines divide first-observed test scores into tertiles.

Figure 2 - 6: Alternative trajectory lengths for national percentile trends, by group for siblings



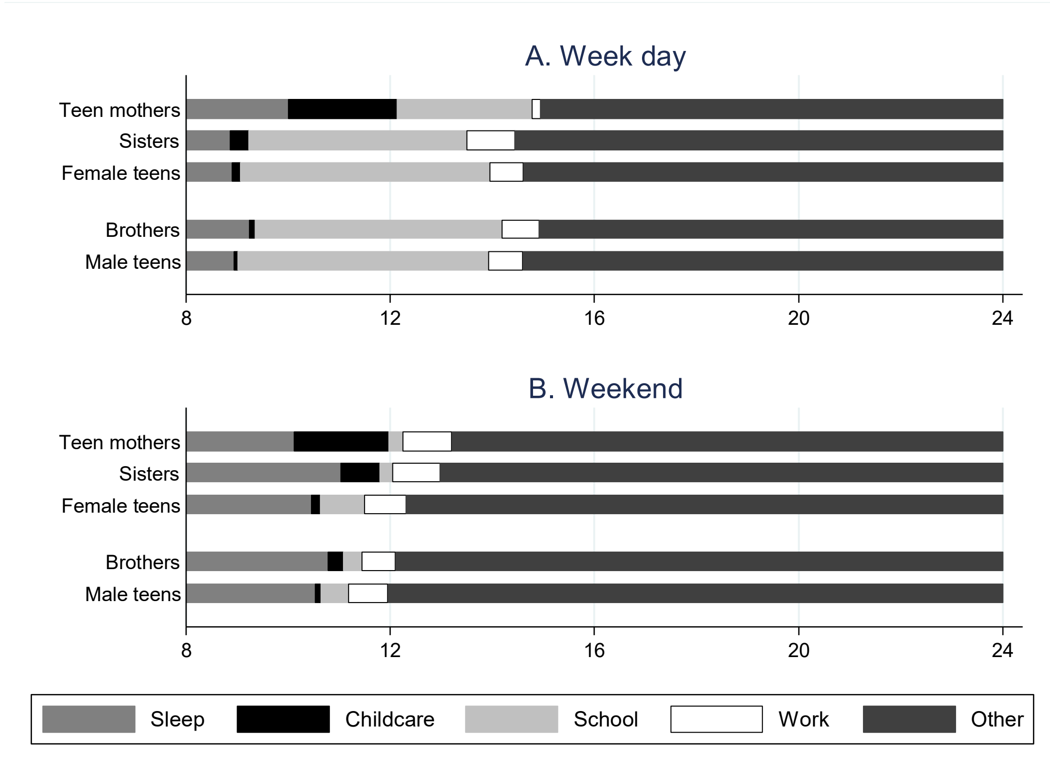
Note: Siblings include all siblings from families where an older sister had a baby at age 15-17. Trajectory matches include matches from non-teenage-childbearing families to the siblings based years of test score trends and other individual and school characteristics; years used in match indicated in panel header. Estimates based on a regression of mean percentile rank on time until birth (or time relative to the match year for the matches) with person and age fixed effects. Vertical line marks the end of the trajectory matching; points to the right allowed to differ from matches.

Figure 2 - 7: National percentile pre- and post-trends, by group for teen mothers



Note: Teen mothers include all females who gave birth at age 15-17. First-observed matches include females from non-teenage-childbearing families to the teen mothers based on first-observed characteristics. Trajectory matches include females from non-teenage-childbearing families to the teen mothers based on four-year test score trends and first-observed characteristics. Estimates based on a regression of mean percentile rank on years relative to birth (or time relative to the match year for the matches) with person and age fixed effects within the noted population. Panel B requires the matches to be from the same neighborhood at first observation as the siblings; Panel A does not. The vertical line marks the end of the trajectory matching; points to the right allowed to differ from matches.

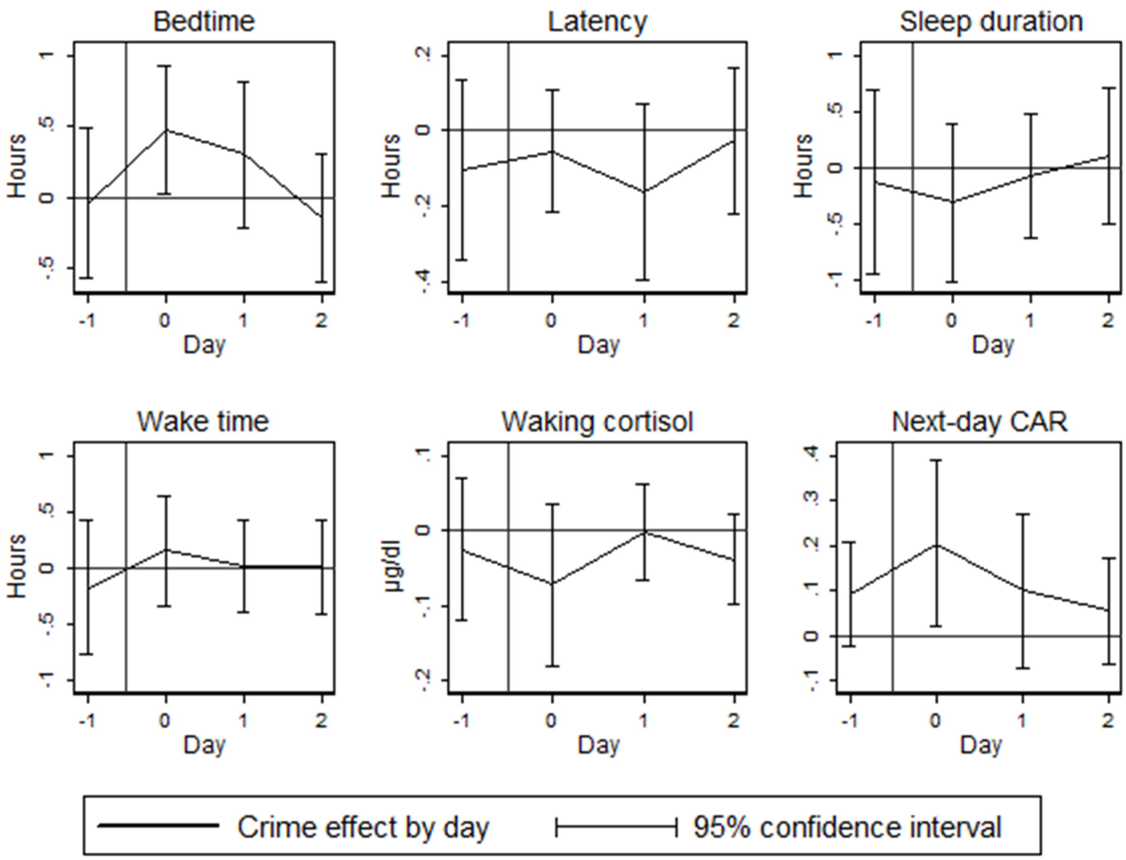
Figure 2 - 8: Unadjusted time use for teenagers by sibling/teen mother status



Note: Outcomes are time use on various activities in hours. Other includes any activity not included as sleep, childcare as a primary activity, school, or work. Categories are exclusive. Population includes 15-, 16-, and 17-year-old American Time Use Survey respondents by group. Teen mothers identified as respondents with own child in their household. Teen fathers excluded due to low N. Sisters and brothers identified as respondents with related, non-own, non-sibling child under the age of five in their household. Other teens are the remaining teens.

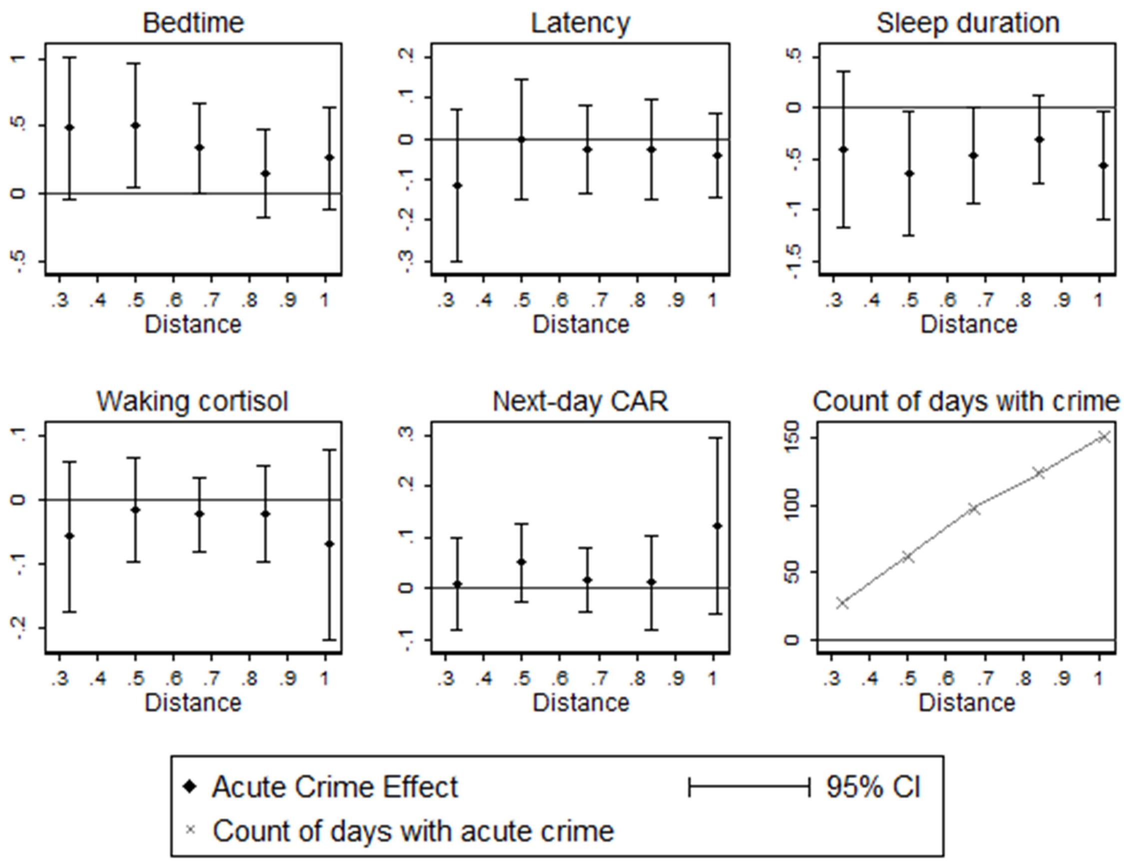
Figures – Study 3

Figure 3 - 1: Effect of crime on sleep and cortisol by day



Effect of crime on sleep and cortisol by day relative to an acute crime event (vertical line). Each chart is derived from a regression that includes interactions between lagged indicators for whether a crime occurred within an individual's police beat during a given time period. All regressions include individual and day-of-the-week fixed effects. Includes 95% confidence intervals.

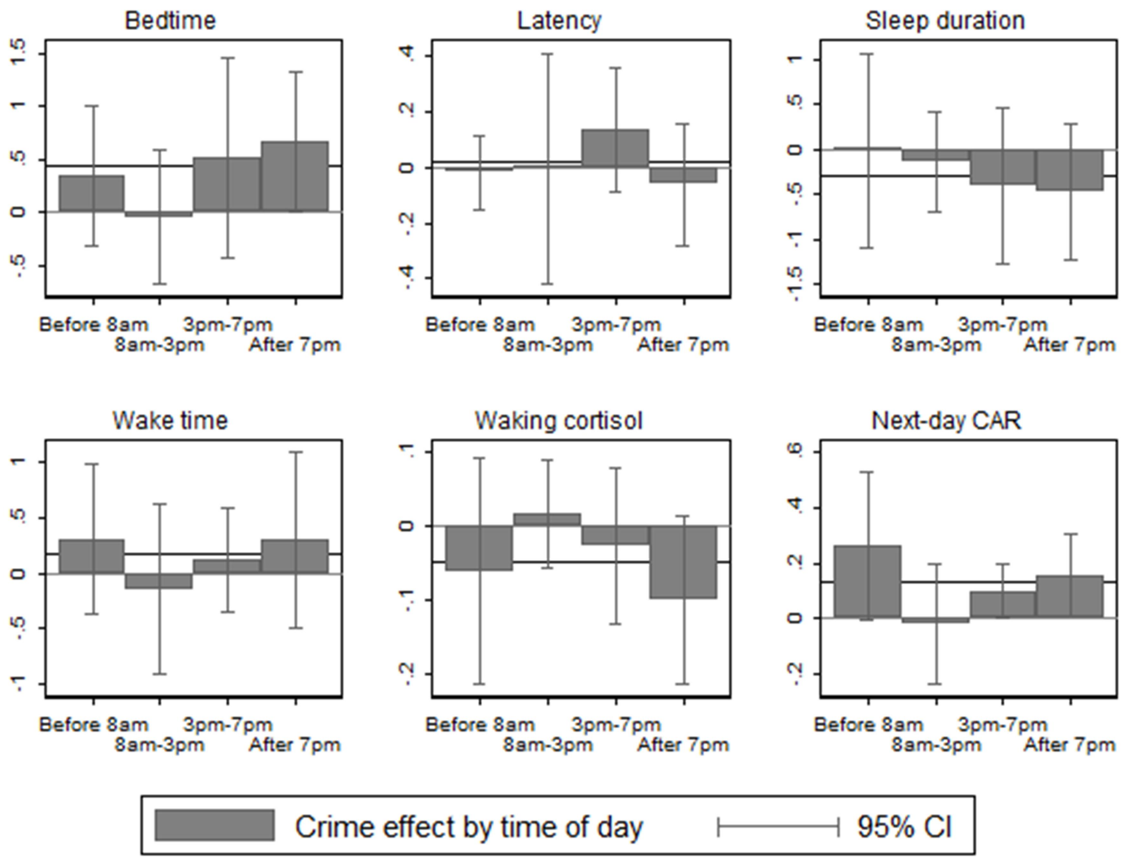
Figure 3 - 2: Effect of acute violent crimes by various distance cutoffs



Effect of acute violent crimes on bedtime, sleep latency, and sleep duration (in hours) and next-day morning cortisol and CAR (in µg/dl) by various crime distance cutoffs, relative to days without crime within the distance cutoff. Each line represents a separate regression that includes individual and day-of-the-week fixed effects. Includes 95% confidence intervals. Also displays a count of individual study-days with a crime for each distance; as the radius expands the count increases.

Figures – Study 3 Appendix

Figure A3 - 1: Effect of acute violent crimes at different time points



Effect of acute violent crimes on bedtime, sleep latency, and sleep duration (in hours) and next-day morning cortisol and CAR (in µg/dl) at different time points, relative to days without crime. Horizontal line represents the main effect from Panel A of the main tables. All time points estimated together for each outcome. Includes 95% confidence intervals.

References

Adam, E. K. (2012). Emotion-cortisol transactions occur over multiple time scales in development: Implications for research on emotion and the development of emotional disorders. *Monographs of the Society for Research in Child Development*, *77*(2), 17–27. https://doi.org/10.1111/j.1540-5834.2012.00657.x

Adam, E. K., Hawkley, L. C., Kudielka, B. M., & Cacioppo, J. T. (2006). Day-to-day dynamics of experience–cortisol associations in a population-based sample of older adults. *Proceedings of the National Academy of Sciences*, *103*(45), 17058–17063. https://doi.org/10.1073/pnas.0605053103

Adam, E. K., Snell, E. K., & Pendry, P. (2007). Sleep timing and quantity in ecological and family context: A nationally representative time-diary study. *Journal of Family Psychology*, *21*(1), 4–19. https://doi.org/10.1037/0893-3200.21.1.4

Adler, N. E., & Rehkopf, D. H. (2008). U.S. disparities in health: Descriptions, causes, and mechanisms. *Annual Review of Public Health*, *29*(1), 235–252. https://doi.org/10.1146/annurev.publhealth.29.020907.090852

Adler, N. E., & Stewart, J. (2010). Health disparities across the lifespan: Meaning, methods, and mechanisms. *Annals of the New York Academy of Sciences*, *1186*(1), 5–23. https://doi.org/10.1111/j.1749-6632.2009.05337.x

Alapin, I., Fichten, C. S., Libman, E., Creti, L., Bailes, S., & Wright, J. (2000). How is good and poor sleep in older adults and college students related to daytime sleepiness, fatigue, and ability to concentrate? *Journal of Psychosomatic Research*, *49*(5), 381–390. https://doi.org/10.1016/S0022-3999(00)00194-X

American Academy of Pediatrics. (2014). School start times for adolescents. *Pediatrics*, *134*(3), 642–649. https://doi.org/10.1542/peds.2014-1697

Arendt, J. (2000). Melatonin, circadian rhythms, and sleep. *New England Journal of Medicine*, *343*(15), 1114–1116. https://doi.org/10.1056/NEJM200010123431510

Ashcraft, A., Fernández-Val, I., & Lang, K. (2013). The consequences of teenage childbearing: Consistent estimates when abortion makes miscarriage non-random. *The Economic Journal*, *123*(571), 875–905. https://doi.org/10.1111/ecoj.12005

Bailey, S. J., Haynes, D. C., & Letiecq, B. L. (2013). “How can you retire when you still got a kid in school?”: Economics of raising grandchildren in rural areas. *Marriage & Family Review*, *49*(8), 671–693. https://doi.org/10.1080/01494929.2013.803009

Barton, P. E., & Coley, R. J. (2009). *Parsing the Achievement Gap II* (pp. 1–38). Princeton, N.J.: Policy Information Center - Educational Testing Service. Retrieved from http://www.ets.org/Media/Research/pdf/PICPARSINGII.pdf

Barton, P. E., & Coley, R. J. (2010). *The Black-White Achievement Gap: When Progress Stopped* (pp. 1–42). Princeton, N.J.: Policy Information Center - Educational Testing Service. Retrieved from http://www.ets.org/Media/Research/pdf/PICBWGAP.pdf

Black, S. E., Breining, S. N., Figlio, D. N., Guryan, J., Karbownik, K., Skyt Nielsen, H., … Simonsen, M. (2016, August 12). *Sibling Spillover*. Northwestern University.

Breining, S. N. (2014). The presence of ADHD: Spillovers between siblings. *Economics Letters*, *124*(3), 469–473. https://doi.org/10.1016/j.econlet.2014.07.010

Breining, S. N., Daysal, N. M., Simonsen, M., & Trandafir, M. (2015). *Spillover Effects of Early-Life Medical Interventions* (SSRN Scholarly Paper No. ID 2615250). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2615250

Buckhalt, J. A. (2011). Insufficient sleep and the socioeconomic status achievement gap. *Child Development Perspectives*, *5*(1), 59–65. https://doi.org/10.1111/j.1750-8606.2010.00151.x

Buckles, K. S., & Hungerman, D. M. (2016). *The Incidental Fertility Effects of School Condom Distribution Programs* (Working Paper No. 22322). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w22322

Bureau of Labor Statistics. (2016, June 24). American Time Use Survey Technical Note. Retrieved March 28, 2017, from https://www.bls.gov/news.release/atus.tn.htm

Burton, L. M. (1999). Teenage childbearing as an alternative life-course strategy in multigeneration black families. *Human Nature*, *1*(2), 123–143. https://doi.org/10.1007/BF02692149

Campbell, I. G., Grimm, K. J., Bie, E. de, & Feinberg, I. (2012). Sex, puberty, and the timing of sleep EEG measured adolescent brain maturation. *Proceedings of the National Academy of Sciences*, *109*(15), 5740–5743. https://doi.org/10.1073/pnas.1120860109

Carrell, S. E., Maghakian, T., & West, J. E. (2011). A's from Zzzz's? The causal effect of school start time on the academic achievement of adolescents. *American Economic Journal: Economic Policy*, *3*(3), 62–81. https://doi.org/10.1257/pol.3.3.62

Carskadon, M. A., Acebo, C., & Jenni, O. G. (2004). Regulation of adolescent sleep: Implications for behavior. *Annals of the New York Academy of Sciences*, *1021*, 276–291. https://doi.org/10.1196/annals.1308.032

Carskadon, M. A., Acebo, C., Richardson, G. S., Tate, B. A., & Seifer, R. (1997). An approach to studying circadian rhythms of adolescent humans. *Journal of Biological Rhythms*, *12*(3), 278–289. https://doi.org/10.1177/074873049701200309

Carskadon, M. A., Vieira, C., & Acebo, C. (1993). Association between puberty and delayed phase preference. *Sleep*, *16*(3), 258–263.

Chase-Lansdale, P. L., Gordon, R. A., Coley, R. L., Wakschlag, L. S., & Brooks-Gunn, J. (1999). Young African American multigenerational families in poverty. In E. M. Hetherington (Ed.), *Coping with divorce, single parenting, and remarriage a risk and resiliency perspective* (pp. 165–191). Mahwah, N.J.: Lawrence Erlbaum Associates.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). The long-term impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633–2679. https://doi.org/10.1257/aer.104.9.2633

Chida, Y., & Steptoe, A. (2009). Cortisol awakening response and psychosocial factors: A systematic review and meta-analysis. *Biological Psychology*, *80*(3), 265–278. https://doi.org/10.1016/j.biopsycho.2008.10.004

Child Trends Data Bank. (2016). *Teen pregnancy: Indicators of child and youth well-being* (pp. 1–10). Bethesda, MD. Retrieved from https://www.childtrends.org/indicators/teen-pregnancy/

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*(6), 673–682. https://doi.org/10.1016/j.econedurev.2007.10.002

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources*, *45*(3), 655–681.

Clow, A., Hucklebridge, F., Stalder, T., Evans, P., & Thorn, L. (2010). The cortisol awakening response: More than a measure of HPA axis function. *Neuroscience & Biobehavioral Reviews*, *35*(1), 97–103. https://doi.org/10.1016/j.neubiorev.2009.12.011

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, *27*(4), 724–750. https://doi.org/10.1002/pam.20375

Crowley, S. J., Acebo, C., & Carskadon, M. A. (2007). Sleep, circadian rhythms, and delayed phase in adolescence. *Sleep Medicine*, *8*(6), 602–612. https://doi.org/10.1016/j.sleep.2006.12.002

Curtin, S. C., Abma, J. C., Ventura, S. J., & Henshaw, S. K. (2013). *Pregnancy rates for U.S. women continue to drop* (National Center for Health Statistics Data Brief No. 138) (pp. 1–8). Washington, D.C.: Center for Disease Control. Retrieved from https://www.cdc.gov/nchs/data/databriefs/db136.pdf

Dahl, G. B., & Lochner, L. (2012). The impact of family income on child achievement: Evidence from the Earned Income Tax Credit. *The American Economic Review*, *102*(5), 1927–1956. https://doi.org/10.1257/aer.102.5.1927

Dahl, R. E. (1996). The regulation of sleep and arousal: Development and psychopathology. *Development and Psychopathology*, *8*(1), 3–27. https://doi.org/10.1017/S0954579400006945

Destin, M., & Oyserman, D. (2009). From assets to school outcomes: How finances shape children's perceived possibilities and intentions. *Psychological Science*, *20*(4), 414–418. https://doi.org/10.1111/j.1467-9280.2009.02309.x

Dressendörfer, R. A., Kirschbaum, C., Rohde, W., Stahl, F., & Strasburger, C. J. (1992). Synthesis of a cortisol-biotin conjugate and evaluation as a tracer in an immunoassay for salivary cortisol measurement. *The Journal of Steroid Biochemistry and Molecular Biology*, *43*(7), 683–692. https://doi.org/10.1016/0960-0760(92)90294-S

East, P. L. (1998). Impact of adolescent childbearing on families and younger sibling: Effects that increase younger siblings' risk for early pregnancy. *Applied Developmental Science*, *2*(2), 62–74. https://doi.org/10.1207/s1532480xads0202_1

East, P. L. (1999). The first teenage pregnancy in the family: Does it affect mothers' parenting, attitudes, or mother-adolescent communication? *Journal of Marriage and the Family*, *61*(2), 306–319. https://doi.org/10.2307/353750

Eaton, D. K., McKnight-Eily, L. R., Lowry, R., Perry, G. S., Presley-Cantrell, L., & Croft, J. B. (2010). Prevalence of insufficient, borderline, and optimal hours of sleep among high school students – United States, 2007. *Journal of Adolescent Health*, *46*(4), 399–401. https://doi.org/10.1016/j.jadohealth.2009.10.011

Edgar, D. M., Dement, W. C., & Fuller, C. A. (1993). Effect of SCN lesions on sleep in squirrel monkeys: Evidence for opponent processes in sleep-wake regulation. *Journal of Neuroscience*, *13*(3), 1065–1079.

Edwards, F. (2012). Early to rise? The effect of daily start times on academic performance. *Economics of Education Review*, *31*(6), 970–983. https://doi.org/10.1016/j.econedurev.2012.07.006

Ellis, B. J., Bates, J. E., Dodge, K. A., Fergusson, D. M., John Horwood, L., Pettit, G. S., & Woodward, L. (2003). Does father absence place daughters at special risk for early sexual activity and teenage pregnancy? *Child Development*, *74*(3), 801–821. https://doi.org/10.1111/1467-8624.00569

Fletcher, J. M. (2011). The effects of teenage childbearing on the short- and long-term health behaviors of mothers. *Journal of Population Economics*, *25*(1), 201–218. https://doi.org/10.1007/s00148-011-0381-9

Fletcher, J. M., & Lehrer, S. F. (2009). *Using Genetic Lotteries within Families to Examine the Causal Impact of Poor Health on Academic Achievement* (Working Paper No. 15148). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w15148

Fletcher, J. M., & Wolfe, B. L. (2009). Education and labor market consequences of teenage childbearing: Evidence using the timing of pregnancy outcomes and community fixed effects. *The Journal of Human Resources*, *44*(2), 303–325. https://doi.org/10.3368/jhr.44.2.303

Fletcher, J. M., & Yakusheva, O. (2016). Peer effects on teenage fertility: Social transmission mechanisms and policy recommendations. *American Journal of Health Economics*, *2*(3), 300–317. https://doi.org/10.1162/AJHE_a_00046

Flinn, M. V., & England, B. G. (1995). Childhood stress and family environment. *Current Anthropology*, *36*(5), 854–866.

Fogel, S. M., & Smith, C. T. (2011). The function of the sleep spindle: A physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neuroscience & Biobehavioral Reviews*, *35*(5), 1154–1165. https://doi.org/10.1016/j.neubiorev.2010.12.003

Fries, E., Dettenborn, L., & Kirschbaum, C. (2009). The cortisol awakening response (CAR): Facts and future directions. *International Journal of Psychophysiology*, *72*(1), 67–73. https://doi.org/10.1016/j.ijpsycho.2008.03.014

Fuller-Thomson, E., Minkler, M., & Driver, D. (1997). A profile of grandparents raising grandchildren In the United States. *The Gerontologist*, *37*(3), 406–411. https://doi.org/10.1093/geront/37.3.406

Geronimus, A. T., & Korenman, S. (1992). The socioeconomic consequences of teen childbearing reconsidered. *The Quarterly Journal of Economics*, *107*(4), 1187–1214. https://doi.org/10.2307/2118385

Gershenson, S., & Tekin, E. (2015). *The Effect of Community Traumatic Events on Student Achievement: Evidence from the Beltway Sniper Attacks* (Working Paper No. 21055). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w21055

Gibson, M., & Shrader, J. (2015). *Time use and productivity: The wage returns to sleep* (Working Paper Series No. 2015–17) (pp. 1–74). Williamstown, MA: Williams College. Retrieved from https://web.williams.edu/Economics/wp/GibsonShrader_Sleep.pdf

Groen, J. A., & Pabilonia, S. W. (2015). *Snooze or lose: High school start times and academic achievement* (Working Paper No. 484). Washington, D.C.: Bureao of Labor Statistics. Retrieved from http://www.bls.gov/osmr/abstract/ec/ec150060.htm

Hale, L., Hill, T. D., Friedman, E., Javier Nieto, F., Galvao, L. W., Engelman, C. D., … Peppard, P. E. (2013). Perceived neighborhood quality, sleep quality, and health status: Evidence from the Survey of the Health of Wisconsin. *Social Science & Medicine*, *79*, 16–22. https://doi.org/10.1016/j.socscimed.2012.07.021

Hansen, M., Janssen, I., Schiff, A., Zee, P. C., & Dubocovich, M. L. (2005). The impact of school daily schedule on adolescent sleep. *Pediatrics*, *115*(6), 1555–1561. https://doi.org/10.1542/peds.2004-1649

Heissel, J. A., & Ladd, H. F. (2016). *The effects of school turnaround in North Carolina: A regression discontinuity approach* (Working Paper No. 156) (pp. 1–54). Washington, D.C.: National Center for Analysis of Longitudinal Data. Retrieved from http://www.caldercenter.org/publications/school-turnaround-north-carolina-regression-discontinuity-analysis

Heissel, J. A., Levy, D. J., & Adam, E. K. (2017). *Stress, Sleep, and Performance on Standardized Tests: Understudied Pathways to the Achievement Gap*. Working paper, Northwestern University.

Heissel, J. A., & Norris, S. (2016). *Rise and Shine: The Effect of School Start Times on Academic Performance from Childhood through Puberty*. Working paper, Northwestern University.

Heissel, J. A., Sharkey, P. T., Torrats-Espinosa, G., Grant, K. E., & Adam, E. K. (2017). *Violence and vigilance: The acute effects of community violent crime on sleep and cortisol*. Working paper, Northwestern University.

Hicken, M. T., Lee, H., Ailshire, J., Burgard, S. A., & Williams, D. R. (2013). "Every shut eye, ain't sleep": The role of racism-related vigilance in racial/ethnic disparities in sleep difficulty. *Race and Social Problems*, *5*(2), 100–112. https://doi.org/10.1007/s12552-013-9095-9

Hinrichs, P. (2011). When the bell tolls: The effects of school starting times on academic achievement. *Education Finance and Policy*, *6*(4), 486–507. https://doi.org/10.1162/EDFP_a_00045

Hotz, V. J., McElroy, S. W., & Sanders, S. G. (2005). Teenage childbearing and its life cycle consequences: Exploiting a natural experiment. *The Journal of Human Resources*, *40*(3), 683–715.

Hotz, V. J., Mullin, C. H., & Sanders, S. G. (1997). Bounding causal effects using data from a contaminated natural experiment: Analysing the effects of teenage childbearing. *The Review of Economic Studies*, *64*(4), 575–603. https://doi.org/10.2307/2971732

Hunfeld, J. A. M., Wladimiroff, J. W., & Passchier, J. (1997). The grief of late pregnancy loss. *Patient Education and Counseling*, *31*(1), 57–64. https://doi.org/10.1016/S0738-3991(97)01008-2

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, *38*(3), 499–534. https://doi.org/10.3102/00028312038003499

Interstate Compact on the Placement of Children. (2017, March 17). Florida Age of Majority. Retrieved March 17, 2017, from http://icpcstatepages.org/florida/ageofmajority/

Jackson, C. K. (2009). Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation. *Journal of Labor Economics*, *27*(2), 213–256. https://doi.org/10.1086/599334

Jacob, B. A. (2007). The challenges of staffing urban schools with effective teachers. *Future of Children*, *17*(1), 129–153.

Jacob, B. A., & Rockoff, J. E. (2011). *Organizing schools to improve student achievement: Start times, grade configurations, and teacher assignments* (Discussion Paper No. 2011–8) (pp. 1–28). Washington, DC: The Hamilton Project. Retrieved from https://www.brookings.edu/wp-content/uploads/2016/06/092011_organize_jacob_rockoff_paper.pdf

Jenni, O. G., & Carskadon, M. A. (2012). Sleep behavior and sleep regulation from infancy through adolescence: Normative aspects. *Sleep Medicine Clinics*, *7*(3), 529–538. https://doi.org/10.1016/j.jsmc.2012.06.002

Joëls, M., Pu, Z., Wiegert, O., Oitzl, M. S., & Krugers, H. J. (2006). Learning under stress: How does it work? *Trends in Cognitive Sciences*, *10*(4), 152–158. https://doi.org/10.1016/j.tics.2006.02.002

Joensen, J. S., & Nielsen, H. S. (2015, November 20). *Spillovers in educational choice*. University of Chicago. Retrieved from http://popcenter.uchicago.edu/events/Joensen%20paper.pdf

Kane, J. B., Morgan, S. P., Harris, K. M., & Guilkey, D. K. (2013). The educational consequences of teen childbearing. *Demography*, *50*(6), 2129–2150. https://doi.org/10.1007/s13524-013-0238-9

Kapinos, K. A., & Yakusheva, O. (2016). Long-term effect of exposure to a friend's adolescent childbirth on fertility, education, and earnings. *Journal of Adolescent Health*, *59*(3), 311–317.e2. https://doi.org/10.1016/j.jadohealth.2016.05.003

Kearney, M. S., & Levine, P. B. (2012). Why is the teen birth rate in the United States so high and why does it matter? *Journal of Economic Perspectives*, *26*(2), 141–166. https://doi.org/10.1257/jep.26.2.141

Kennelly, L., & Monrad, M. (2007). *Approaches to Dropout Prevention: Heeding Early Warning Signs with Appropriate Interventions*. American Institutes for Research. Retrieved from http://eric.ed.gov/?id=ED499009

Krueger, A. B., & Whitmore, D. M. (2002). Would smaller classes help close the Black-White achievement gap? In J. E. Chubb & T. Loveless (Eds.), *Bridging the Achievement Gap.* Washington, D.C.: Brookings Institution Press.

Laberge, L., Petit, D., Simard, C., Vitaro, F., Tremblay, R. E., & Montplaisir, J. (2001). Development of sleep patterns in early adolescence. *Journal of Sleep Research*, *10*(1), 59–67. https://doi.org/10.1046/j.1365-2869.2001.00242.x

Lacoe, J., & Sharkey, P. T. (2016). Life in a crime scene: Stop, question, and frisk activity in New York City neighborhoods in the aftermath of homicides. *Sociological Science*, *3*, 116–134. https://doi.org/10.15195/v3.a7

Lavie, P. (2001). Sleep disturbances in the wake of traumatic events. *New England Journal of Medicine*, *345*(25), 1825–1832. https://doi.org/10.1056/NEJMra012893

Levine, D. I., & Painter, G. (2003). The schooling costs of teenage out-of-wedlock childbearing: Analysis with a within-school propensity-score-matching estimator. *Review of Economics and Statistics*, *85*(4), 884–900. https://doi.org/10.1162/003465303772815790

Lindo, J. M., & Packham, A. (forthcoming). How much can expanding access to long-acting reversible contraceptives reduce teen birth rates? *American Economic Journal: Economic Policy*. Retrieved from https://www.aeaweb.org/articles?id=10.1257/pol.20160039&&from=f

Lok, I. H., & Neugebauer, R. (2007). Psychological morbidity following miscarriage. *Best Practice & Research Clinical Obstetrics & Gynaecology*, *21*(2), 229–247. https://doi.org/10.1016/j.bpobgyn.2006.11.007

Lufi, D., Tzischinsky, O., & Hadar, S. (2011). Delaying school starting time by one hour: Some effects on attention levels in adolescents. *Journal of Clinical Sleep Medicine*, *7*(2), 137–143.

Lupien, S. J., Wilkinson, C. W., Brière, S., Ménard, C., Ng Ying Kin, N. M. K., & Nair, N. P. V. (2002). The modulatory effects of corticosteroids on cognition: Studies in young human populations. *Psychoneuroendocrinology*, *27*(3), 401–416. https://doi.org/10.1016/S0306-4530(01)00061-0

Maldonado, E. F., Fernandez, F. J., Trianes, M. V., Wesnes, K., Petrini, O., Zangara, A., … Ambrosetti, L. (2008). Cognitive performance and morning levels of salivary cortisol and alpha-amylase in children reporting high vs. low daily stress perception. *The Spanish Journal of Psychology*, *11*(1), 3–15. https://doi.org/10.1017/S1138741600004066

Marshall, W. A., & Tanner, J. M. (1970). Variations in the pattern of pubertal changes in boys. *Archives of Disease in Childhood*, *45*(239), 13–23. https://doi.org/10.1136/adc.45.239.13

Martin, D. (1999, August 1). Late to bed, early to rise makes a teen-ager ... tired. *The New York Times*. Retrieved from http://www.nytimes.com/1999/08/01/education/late-to-bed-early-to-rise-makes-a-teen-ager-tired.html

Matthews, K. A., Gallo, L. C., & Taylor, S. E. (2010). Are psychosocial factors mediators of socioeconomic status and health connections? *Annals of the New York Academy of Sciences*, *1186*(1), 146–173. https://doi.org/10.1111/j.1749-6632.2009.05332.x

Meltzer, L. J., & Montgomery-Downs, H. E. (2011). Sleep in the family. *Pediatric Clinics of North America*, *58*(3), 765–774. https://doi.org/10.1016/j.pcl.2011.03.010

Miller, A. R. (2009). The effects of motherhood timing on career path. *Journal of Population Economics*, *24*(3), 1071–1100. https://doi.org/10.1007/s00148-009-0296-x

Miller, B. C., Benson, B., & Galbraith, K. A. (2001). Family relationships and adolescent pregnancy risk: A research synthesis. *Developmental Review*, *21*(1), 1–38. https://doi.org/10.1006/drev.2000.0513

Monstad, K., Propper, C., & Salvanes, K. G. (2011). *Is Teenage Motherhood Contagious? Evidence from a Natural Experiment* (SSRN Scholarly Paper No. ID 1908553).

Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=1908553

Myers, H. F. (2008). Ethnicity- and socio-economic status-related stresses in context: An integrative review and conceptual model. *Journal of Behavioral Medicine*, *32*(1), 9–19. https://doi.org/10.1007/s10865-008-9181-4

Ng, E. P., Ng, D. K., & Chan, C. H. (2009). Sleep duration, wake/sleep symptoms, and academic performance in Hong Kong Secondary School Children. *Sleep and Breathing*, *13*(4), 357–367. https://doi.org/10.1007/s11325-009-0255-5

Nicoletti, C., & Rabe, B. (2014). *Sibling Spillover Effects in School Achievement* (SSRN Scholarly Paper No. ID 2529324). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2529324

Oakley, N. R. (1997). *Validation with polysomnography of the sleepwatch sleep/wake scoring algorithm used by the Actiwatch activity monitoring system*. Bend: Mini Mitter, Cambridge Neurotechnology.

Orsi, J. M., Margellos-Anast, H., & Whitman, S. (2010). Black-White health disparities in the United States and Chicago: A 15-year progress analysis. *American Journal of Public Health*, *100*(2), 349–356. https://doi.org/10.2105/AJPH.2009.165407

Penman-Aguilar, A., Carter, M., Snead, M. C., & Kourtis, A. P. (2013). Socioeconomic Disadvantage as a Social Determinant of Teen Childbearing in the U.S. *Public Health Reports*, *128*(Supplement 1), 5–22.

Prettyman, R. J., Cordle, C. J., & Cook, G. D. (1993). A three-month follow-up of psychological morbidity after early miscarriage. *British Journal of Medical Psychology*, *66*(4), 363–372. https://doi.org/10.1111/j.2044-8341.1993.tb01762.x

Pruessner, J. C., Wolf, O. T., Hellhammer, D. H., Buske-Kirschbaum, A., von Auer, K., Jobst, S., … Kirschbaum, C. (1997). Free cortisol levels after awakening: A reliable biological marker for the assessment of adrenocortical activity. *Life Sciences*, *61*(26), 2539–2549. https://doi.org/10.1016/S0024-3205(97)01008-4

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. J. Duncan & R. J. Murnane (Eds.), *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances* (pp. 91–116). New York, NY: Russell Sage Foundation.

Reardon, S. F., & Robinson, J. F. (2007). Patterns and trend in racial/ethnic and socioeconomics academic achievement gaps. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of Research in Education Finance and Policy* (1st ed.). New York, NY: Routledge.

Repetti, R. L., Taylor, S. E., & Seeman, T. E. (2002). Risky families: Family social environments and the mental and physical health of offspring. *Psychological Bulletin*, *128*(2), 330–366. https://doi.org/10.1037/0033-2909.128.2.330

Romeo, R. D. (2010). Pubertal maturation and programming of hypothalamic–pituitary–adrenal reactivity. *Frontiers in Neuroendocrinology*, *31*(2), 232–240. https://doi.org/10.1016/j.yfrne.2010.02.004

Sadeh, A. (1996). Stress, trauma, and sleep in children. *Child and Adolescent Psychiatric Clinics of North America*, *5*(3), 685–700.

Sadeh, A., Gruber, R., & Raviv, A. (2003). The effects of sleep restriction and extension on school-age children: What a difference an hour makes. *Child Development*, *74*(2), 444–455. https://doi.org/10.1111/1467-8624.7402008

Sapolsky, R. M., Romero, L. M., & Munck, A. U. (2000). How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. *Endocrine Reviews*, *21*(1), 55–89. https://doi.org/10.1210/er.21.1.55

Schanzenbach, D. W. (2006). What Have Researchers Learned from Project STAR? *Brookings Papers on Education Policy*, *9*, 205–228.

Schwabe, L., & Wolf, O. T. (2010). Learning under stress impairs memory formation. *Neurobiology of Learning and Memory*, *93*(2), 183–188. https://doi.org/10.1016/j.nlm.2009.09.009

Sharkey, P. T. (2010). The acute effect of local homicides on children's cognitive performance. *Proceedings of the National Academy of Sciences*, *107*(26), 11733–11738. https://doi.org/10.1073/pnas.1000690107

Sharkey, P. T., Tirado-Strayer, N., Papachristos, A. V., & Raver, C. C. (2012). The effect of local violence on children's attention and impulse control. *American Journal of Public Health*, *102*(12), 2287–2293. https://doi.org/10.2105/AJPH.2012.300789

Smith, A. C. (2016). Spring forward at your own risk: Daylight Saving Time and fatal vehicle crashes. *American Economic Journal: Applied Economics*, *8*(2), 65–91. https://doi.org/10.1257/app.20140100

Stanford CEPA. (2016). *The Educational Opportunity Monitoring Project: Racial and Ethnic Achievement Gaps*. Retrieved from http://cepa.stanford.edu/educational-opportunity-monitoring-project/achievement-gaps/race/#first

Stroud, L. R., Foster, E., Papandonatos, G. D., Handwerger, K., Granger, D. A., Kivlighan, K. T., & Niaura, R. (2009). Stress response and the adolescent transition: Performance versus peer rejection stressors. *Development and Psychopathology*, *21*(1), 47–68. https://doi.org/10.1017/S0954579409000042

Talbot, L. S., McGlinchey, E. L., Kaplan, K. A., Dahl, R. E., & Harvey, A. G. (2010). Sleep deprivation in adolescents and adults: Changes in affect. *Emotion*, *10*(6), 831–841. https://doi.org/10.1037/a0020138

Tavernier, R., Heissel, J. A., Sladek, M. R., Grant, K. E., & Adam, E. K. (forthcoming). Adolescents' technology and face-to-face time use predict objective sleep outcomes. *Sleep Health*.

U.S. Census Bureau. (2016). Table FM-3. Average Number of Own Children Under 18 Per Family, By Type of Family: 1955 To Present. Retrieved April 5, 2017, from https://www.census.gov/hhes/families/data/families.html

U.S. Department of Education. (2012). *Average start time for public high schools and percentage distribution of start times in public high schools, by selected school characteristics: 2011–12*. Washington, D.C.: National Center for Education Statistics. Retrieved from https://nces.ed.gov/surveys/sass/tables/sass1112_201381_s1n.asp

U.S. Department of Health and Human Services. (2015). Trends in Teen Pregnancy and Childbearing. Retrieved April 4, 2017, from https://www.hhs.gov/ash/oah/

U.S. Department of Housing and Urban Development. (2016). Neighborhoods and Violent Crime. Retrieved April 4, 2017, from https://www.huduser.gov/portal/periodicals/em/summer16/highlight2.html

U.S. Department of Justice. (2016). *Uniform Crime Report: Crime in the United States, 2015*. Federal Bureau of Investigation. Retrieved from https://ucr.fbi.gov/crime-in-the-u.s/2015/crime-in-the-u.s.-2015/offenses-known-to-law-enforcement/violent-crime

Vedhara, K., Hyde, J., Gilchrist, I. ., Tytherleigh, M., & Plummer, S. (2000). Acute stress, memory, attention and cortisol. *Psychoneuroendocrinology*, *25*(6), 535–549. https://doi.org/10.1016/S0306-4530(00)00008-1

Vigil, J. M., Geary, D. C., Granger, D. A., & Flinn, M. V. (2010). Sex differences in salivary cortisol, alpha-amylase, and psychological functioning following Hurricane Katrina. *Child Development*, *81*(4), 1228–1240. https://doi.org/10.1111/j.1467-8624.2010.01464.x

Vrshek-Schallhorn, S., Doane, L. D., Mineka, S., Zinbarg, R. E., Craske, M. G., & Adam, E. K. (2012). The cortisol awakening response predicts major depression: Predictive stability over a 4-year follow-up and effect of depression history. *Psychological Medicine*, *43*(3), 483–493. https://doi.org/10.1017/S0033291712001213

Wahlstrom, K. L., Wrobel, G., & Kubow, P. (1998). *Minneapolis Public Schools start time study executive summary*. Minneapolis, MN: University of Minnesota, Center for Applied Research and Educational Improvement. Retrieved from http://conservancy.umn.edu/handle/11299/3902

Walker, M. P., & Stickgold, R. (2006). Sleep, memory, and plasticity. *Annual Review of Psychology*, *57*(1), 139–166. https://doi.org/10.1146/annurev.psych.56.091103.070307

Yi, J., Heckman, J. J., Zhang, J., & Conti, G. (2015). Early health shocks, intra-household resource allocation and child outcomes. *The Economic Journal*, *125*(588), F347–F371. https://doi.org/10.1111/ecoj.12291

Appendices

Appendices for Study 1

Robustness checks for mover definition

The identifying variation comes from students who move between schools in different time zones in the Florida panhandle. Most of these moves are long-distance; the median move is 83 miles. The disruption inherent in such a move may have an independent effect on achievement, which is important to control for in the present context. To help identify the effect of moving, as well as the effect of other school-level covariates, I include in the sample students who move within a time zone. This requires defining what constitutes a move by setting a threshold distance between the schools the student attended. Otherwise, graduating from middle school to high school would constitute a move. A high threshold has the advantage of making the move more likely to match a cross-time zone move in terms of disruptiveness; a low threshold increases sample size and precision.

I settled on a threshold of 25 miles, but the results are robust to other threshold choices. Table A1-1 presents estimates for 15, 20, 25, and 30 mile thresholds for math and reading outcomes. I also consider defining a move as any move between different school districts, although this will include students who move less disruptive distances, such as when families move to a nearby suburb that happens to be in a different district. Across all definitions, the results are broadly consistent. In math, the effect for prepubescent children ranges from 0.009 to 0.037 SDs; the effect for adolescents ranges from 0.067 to 0.084 SDs. In reading, the range is 0.034 to 0.061 for younger children and 0.044 to 0.057 for adolescents. The effects statistically differ from zero for adolescents for both math and reading across all distances.

Specification robustness checks

I include two sets of control variable robustness checks. First, in Table A1-2, I consider different levels of aggregation for the demographic share controls (FRL, male, black, Asian, and Hispanic). Instead of aggregating at the school-year level, as in the main results, I consider district-year, district third graders-year,[71] school-year, and school-grade-year. All specifications include age-gender dummies and an individual fixed effect. For each level of aggregation, I present one specification with no other controls, one that adds urban dummies and log income controls, and a final model that includes school size and student/teacher ratio.

Comparing across the rows of Table A1-2, the results are largely unchanged. In Panel A, all specifications show an effect size in math of 0.003-0.037 SDs for prepubescents, and 0.062-0.096 for adolescents. The effect is statistically significant at the 1% level or better for adolescents but null for younger students. In reading, the estimates are also similar across specifications: 0.046-0.087 SDs for prepubescents, and 0.044-0.074 SDs for adolescents. The prepubescent effect is occasionally significant at the 5% level; the adolescent effect has a *p*-value of about 1%.

For absences, the inclusion of demographics (but not the level of aggregation) makes a substantive difference in the results. Comparing Columns 1-3 with Columns 4-15, the inclusion of demographic controls (at any level of aggregation) reduces the size of the suspension effect from about 1.5 percentage points and significant at the 1% level to about 0.8 percentage points and significant at the 10% level for prepubescents. The adolescent effects are generally null once I control for demographics. Since there may be significant between-school differences in policies for counting absences (and these may be correlated with school demographics), I think that the

[71] District third graders-year is the demographic means for the third graders in the given district-year.

results with demographic controls are more trustworthy. It is therefore reassuring that they are the same regardless of the level of demographic aggregation.

Table A1-3 contains the second control robustness check. Columns 1 and 3 restate the baseline results for math and reading. Columns 2 and 5 include controls for latitude; average sunrise times over the school year vary by about a minute over the north-south range of the panhandle and this could conceivably have some effect on sleep (in contrast, the east-west variation in sunrise times from longitude is nearly 20 minutes, excluding the time zone change).[72] The addition of latitude has a moderately sized but statistically insignificant effect on the prepubescent coefficients. The change in the adolescent coefficients is smaller.

In Columns 3 and 6 I test whether the inclusion of third grade district test scores as control variables affects the results. Third grade test scores are appealing as a summary measure of district quality, but may be endogenous if start times affect performance for children in kindergarten to third grade. For this reason I do not include them in the main specification, but it is reassuring that they have little effect on the results.

Changes in school characteristics over the move

A potential threat to the identification strategy is changes in school and peer characteristics as students move between time zones. If students moving from CT to ET move to significantly worse schools, while ET-CT movers moved to better schools, it would not be surprising that student achievement declined upon entering ET and rose upon exiting. Because, on average, there is less sunlight before school in ET than in CT, this could generate a spurious positive relationship between relative school start times and academic achievement.

[72] The average disguises some larger differences over the year; but it is never larger than three minutes.

I consider this question directly in Table A1-4. I take the years directly before and after each move, and term these pairs of years a *moving episode*.[73] I then regress school- and zip-level characteristics on moving episode fixed effects and move indicators for the four different types of movers: Eastern-Eastern, Central-Central, Eastern-Central, and Central-Eastern. Each coefficient is a measure of the change in characteristics over the move. As outcomes, I consider the five school-level demographic share controls included in the preferred specification (percent FRL, male, black, Asian, and Hispanic), as well as school student/teacher ratio and zip code-level median income as a measure of school and community resources.

The first two rows of Table A1-4 show that peer quality changed slightly over the move for within-time zone movers. ET-ET movers had 4.5 percentage points fewer FRL classmates; CT-CT movers had 1.7 percentage points fewer. School quality as measured by the student/teacher ratio increased slightly for both groups. Median income rose by $1,000 for within-ET movers and fell by $430 for within-CT movers. These differences are statistically significant, but none are particularly large or striking.

The cross-time zone movers tell a slightly different story. Eastward movers generally ended up in a richer area – 4.5 percentage points fewer FRL classmates and $5,700 higher median income –and had 14.0 percentage points more black classmates and 0.5 percentage points more Hispanic classmates. School quality as measured by the student/teacher ratio was unchanged. ET-CT movers saw approximately the opposite changes in medium income and percent of black students. The economic and peer changes may work in opposite directions in this case, making it unclear in which direction the overall bias goes. However, neither the inclusion of demographic controls (in Table 1-2) or income controls (in Table A1-2)

[73] Since occasionally a student will move in consecutive years, a small number of observations are repeated.

substantively changes the results, suggesting that changes in peer characteristics have only a moderate effect on outcomes over the move, and do not significantly affect the results.

Performance trend before move

Figure 1-1 shows that test score trends are similar for all groups of movers in the years before the move. However, math scores trend *up*, which is somewhat surprising since the disruption of the upcoming move would be expected to reduce scores. Figure A1-1 show results from a regression of scale scores on time-until-move dummies and a fixed effect for the period until the move. This is identical to the regression displayed in Figure 1-1, but without controls. The figure confirms that unconditionally, test scores trend down in both math and reading before a move. This is largely a result of removing the age-gender fixed effects, which soak up any time trend. Comparing across different groups of movers, the trends are slightly further apart than in the version with controls, but are still generally statistically indistinguishable.

Robustness checks for puberty definition

One of the main interests in this paper is how the effect of relative school start times varies with pubertal status. This requires a working definition of puberty, and there are several defensible alternatives. Pubertal development is typically measured with the Tanner Scale. There are two versions: one that uses levels of pubic hair to define the stages and another that uses breast and genital development. I rely on the pubic hair version of the Scale, which Campbell et al. (2012) indicate is more closely associated with pubertal changes in sleep patterns. They also note that changes in sleep patterns begin during Stage 3, so I use the age of median attainment (by gender) of Stage 3 as the definition of puberty.

Table A1-5 shows the main results with three alternative definitions of puberty: pubic hair Stage 2, pubic hair Stage 4, and breast/genital Stage 3. These changes typically shift the age of puberty by at most a year, and not necessarily for both genders. The results are largely unchanged, although slightly attenuated in some specifications. Because this definition of puberty is a worse fit for the underlying biological processes, this is unsurprising.

Estimates without interactions

Table A1-6 displays a version of the baseline model without an interaction between relative start time and pubertal status. Allowing for heterogeneity by pubertal status is important, but for completeness I have included this specification.

Across the rows, the change in sunlight is about 30 minutes over the time zone border. For both math and reading, the effect of moving start times one hour later is about the average of the child and adolescent effects from Table 1-2. In math, the estimated effect is 0.043 SD per hour by the final column, and the estimates are only occasionally statistically significant. In reading, the effect is 0.059 SD per hour by the final column, and the effect sizes are all significant to at least the 5% level in all estimates. The attendance results vary, with a decrease of 0.7 percentage points in absence per hour of sunlight by the final column.

PSID data definitions

The present paper demonstrates that students treated with later relative start times have higher academic achievement. However, I do not directly observe sleep levels in the academic outcomes dataset. To more concretely link changes in start times to changes in sleep, I use the Child Development Supplement of the Panel Study of Income Dynamics (PSID) to estimate the effect of the time zone boundary on sleep. The survey collected time use diaries for students on a weekend day and a weekday in the years 1997, 2002, and 2007. I include all states with a single time zone,[74] and all children who were 6-19 during the survey and within 400 miles of the ET-CT time zone boundary. The aim is descriptive, so I regress daily hours of sleep on a fully interacted set of dummies for puberty, CT, and whether the night was a weekend. In the preferred specification, I also include controls for gender, black/non-black, and FRL status. I expect that children in CT will have more sleep on weekdays when they face earlier relative start times, and those in ET will compensate with more sleep on weekends.

Table A1-7 contains the results. As discussed in the mechanisms section, children in CT get 6 minutes more sleep per night during the week than children in ET; during puberty they get 17 minutes more. On the weekend, children in ET compensate for low levels of sleep during the week by sleeping 10 minutes more per night in the years before puberty and 19 minutes more while in puberty. I conservatively cluster at the state level. The coefficient for the difference in sleep between adolescents in CT and ET is significant at the 10% level; most others are not. Including student fixed effects suggests a slightly larger difference between the time zones: the decrease in sleep during puberty is 15 minutes smaller for adolescents in CT than in ET. This set of results corresponds to a pass-through rate of about 40-50% from school start times to sleep if Florida panhandle school start times are representative of the rest of the US near the ET-CT time zone boundary. This number is close to the 46% pass-through reported by Wahlstrom (1998).

Treatment bleed for schools near the time zone boundary

In the placebo analysis, I study how test scores change when students move east-west or west-east but *not* across the true time zone boundary. Ideally, I would examine within-time zone

[74] The CDS does not geocode individuals at a sub-state level in the publicly available version, which precludes analysis using observations in states with multiple time zones – including Florida.

moves to and from the region directly adjacent to the boundary, to help test whether there are unobservable changes in the school or community environment that occur nearby, but not exactly at, the time zone boundary.

This approach will be problematic if there is an effect of being near the time zone boundary on school start times – then, moving from directly beside the boundary in CT to a city fifty miles west could increase relative start times, directly increasing test scores. Figure A1-3 displays a nonparametric regression of relative start times on distance to the time zone boundary, estimated separately for each time zone. In the region directly adjacent to the boundary, start times veer towards the other time zone's norm, particularly for adolescents. I interpret this as the synchronization of start times across time zones, which allows parents to help their children prepare for school before going to work, whether or not they are commuting across time zones. This also means that start times are later for students moving west either from the region directly beside the boundary in CT, or *to* the region directly beside the boundary in ET.

In the main placebo results, I account for the treatment bleed across time zones by taking out a 25 mile "donut" around the time zone boundary. However, in the interest of completeness I include the unexcised version in Figure A1-4. The difference with Figure 1-5 is most stark in the puberty-time zone coefficient for math, where there is a consistent effect above the size of the true coefficient. Comparing between figures, removing the donut around the time zone boundary reduces the size of *all* placebo coefficients. The placebo effect is coming largely from individuals moving between the area close to the true time zone boundary and the rest of the study area, not individuals moving between areas far from the time zone boundary.

Appendix for Study 2

Patterns for teen mothers by race and FRL status

This appendix explores the subgroup analysis for the own effects on teen mothers. One concern with the results might be that they are driven by a particular subgroup; for instance, perhaps lower-income mothers have fewer supports. Table A4 repeats the main analysis by subgroup. To be conservative, these estimates include the neighborhood requirement, as they account for unobserved neighborhood differences. Column 1 contains the preferred overall estimates for reference. There are no statistical differences in outcomes by subgroup, though if anything non-black and non-FRL teen mothers have worse than FRL teen mothers outcomes following the birth.

Appendix for Study 3

I include several preliminary tests of potential differences in effects in Table A3-1. These should be interpreted with caution given the low power in the small sample.

Differences by chronic neighborhood violence

I begin by examining whether the effect sizes differed for those living in high-crime areas. I defined high-crime areas as those that had an above-median number of violent crimes in their police beat in the months of the study. When interacting acute exposure to crime ($Crime_{it}$) with residence in high-crime areas, I found no difference in the estimated change in sleep or cortisol across adolescents living in high- and low-crime areas. If anything, the directions indicate that the effect of crime is higher in low-crime areas. The estimates are 38 minutes later bedtime for low-crime areas versus 20 minutes for high-crime areas, 47 minutes less sleep duration for low-crime areas versus one minute less sleep duration for high-crime areas, 0.142 μg/dl lower waking cortisol for low-crime areas versus 0.001 μg/dl lower waking cortisol for high-crime areas, and 0.227 μg/dl higher CAR for low-crime areas versus 0.124 μg/dl higher CAR for high-crime areas. Given the small sample size, these tests should be interpreted with caution, but do point to important work for future work.

Differences by pubertal status

I use the mean of the responses to the four-point Peterson pubertal scale to measure pubertal status for each adolescent. Lower scores indicate earlier pubertal stages. With this strategy, there are no estimated differences in the effect size by pubertal status for any of the sleep or cortisol outcomes, though this may be due to low power. The directions of some of the estimated effects indicate this may be an important area for future study. As an example, the

CAR estimate is 0.046 μg/dl for participants at the lowest end of the pubertal scale (meaning the pubertal transition hasn't started) and 0.142 μg/dl for those at the top end of the scale (meaning the pubertal transition has completed). Although the difference between these estimates is not statistically significant, the difference is intriguing for future research, especially given prior research indicating that adolescence is associated with increased cortisol reactivity relative to children (e.g., Stroud et al., 2009). Conversely, sleep effects are larger for those at the lower end of the pubertal scale, with 52 minutes later bedtime and 41 minutes less sleep duration for participants at the lowest end of the pubertal scale, compared to no change in bedtime and four minutes more sleep for participants at the highest end of the pubertal scale. Again, these tests should be interpreted with caution given the small sample size, but point to important future work.

Sensitivity to time of day

Crime may have different effects at different times of the day. Figure A3-1 displays preliminary results at different time points: (1) before 8:00am (approximately before school), (2) between 8:00am and 3:00pm (approximately during school hours), (3) between 3:00pm and 7:00pm (after school/afternoon), and (4) after 7:00pm (approximately during bedtime/winding-down time). These divisions roughly break crime into timing quartiles. The horizontal line represents the effect estimated in the main table. For bedtime, the largest effect occurs during the bedtime period – but it's barely different (and statistically indistinguishable) from the afternoon crime. In fact, I cannot statistically distinguish near-bedtime crime from any of the other effects given the sample size. Similarly, the sleep duration estimates are largest for near-bedtime crime, followed closely by afternoon crime, but I cannot distinguish between the periods.

An alternative model approximately bisects the data, defining an early period as before 4:00 p.m. The child was likely at school during much of this period, while they were more likely to be home after this period. An indicator for early crime was then interacted with crime (Panel C of Table A3-1). Again, the later crime had larger effects on bedtime (35 minutes later, *p*-value=0.041) and sleep duration (30 minutes less, *p*-value=0.096) than earlier crime (both statistically null at 17 minutes later bedtime and five minutes shorter sleep duration), though the difference is not statistically significant.

Overall, this section provides suggestive evidence that near-bedtime crime may matter more for sleep, but it is not definitive given the small sample size. Future research in this area may help distinguish whether the effect is driven by the opponent process model of sleep, greater knowledge of the crime that occurs when the adolescent is home, or noise and disruption from the crime and expanded police presence (Edgar et al., 1993; Lacoe & Sharkey, 2016).

Vita

Jennifer A. Heissel
Institute for Policy Research
2040 Sheridan Road • Evanston, IL 60208
jheissel@u.northwestern.edu
https://sites.google.com/site/jheissel/

Education

Expected 2017	Ph.D. in Human Development and Social Policy, Northwestern University *Committee:* David Figlio, Emma Adam, Jonathan Guryan
2012	Master of Public Policy, Sanford School of Public Policy, Duke University *Committee:* Helen Ladd, Seth Sanders, Jacob Vigdor
2007	B.A., *magna cum laude*, Economics and Sociology, University of Notre Dame

Publications

** Indicates author ordering by contribution*

Forthcoming	**Heissel, J. A.**, & Norris, S. Rise and shine: The effect of school start times on academic performance from childhood through puberty. *Journal of Human Resources*.
Forthcoming	**Heissel, J. A.** Teenage motherhood and sibling test scores. *American Economic Review Papers and Proceedings*.
Forthcoming	***Heissel, J.A.**, Sharkey, P., Torrats-Espinosa, G., Grant, K., & Adam, E.K. Violence and vigilance: The acute effects of community violent crime on sleep and cortisol. *Child Development*.
Forthcoming	*Tavernier, R., **Heissel, J.A.**, Sladek, M., Grant, K., & Adam, E.K. Adolescents' technology and face-to-face time use predict objective sleep outcomes. *Sleep Health*.
2016	**Heissel, J. A.** *The relative benefits of live versus online delivery: Evidence from* virtual Algebra I in North Carolina. *Economics of Education Review*, 53, 99-115. Media coverage: *Education Week, NPR Education, The 74*
2016	*Levy, D. J., **Heissel, J. A.**, Richeson, J. A., & Adam, E. K. Psychological and biological responses to race-based social stress as pathways to disparities in educational outcomes. *American Psychologist*, 71, 455-473. Media coverage: *The Nation, The Atlantic, Fortune*

2015 *Adam, E. K., **Heissel, J. A.**, Zeiders, K. H., Richeson J. A., Brodish, A., Ross, E. C., Ehrlich, K. B., Levy, D. J., Kemeny, M. E., Malanchuk, O., Peck, S., Fuller-Rowell, T. and Eccles, J. S. Developmental histories of perceived racial discrimination and diurnal cortisol profiles in adulthood: A 20-year prospective study. *Psychoneuroendocrinology*, 62, 279-291.
Media coverage: Pacific Standard, The Louisiana Weekly, Seattle Medium, Boston Globe, Mother Jones

Manuscripts Submitted for Publication

*__Heissel, J.A.__, Levy, D. J., & Adam, E.K. Stress, sleep, and performance on standardized tests: Understudied pathways to the achievement gap. Revised and resubmitted at *AERA Open*.

Heissel, J.A., & Ladd, H. F. School turnaround in North Carolina: A regression discontinuity analysis. Invited to be revised and resubmitted at *Economics of Education Review*.
Media coverage: The 74, Education Next, The News & Observer

Works in Progress

Spillover effects within families: Evidence from teenage motherhood and sibling performance from high school through college

High-stakes testing, stress, and performance: Biological pathways (with Emma Adam, Jennifer Doleac, David Figlio, and Jonathan Meer)

Home and school location, wind direction, and traffic patterns: The effect of roadway pollution on academic performance (with Claudia Persico and David Simon)

Water main replacement, lead exposure, and academic performance in Chicago (with Claudia Persico and David Simon)

Urban-rural divide: Distribution of sample selection across disciplinary journals (with Sarah Cannon)

Professional, Invited, & Upcoming Presentations

2017 American Economic Association; Naval Postgraduate School; University of Alabama; University of Chicago (Crime Lab New York, Poverty Lab); University of Illinois; University of Notre Dame; Association for Education Finance and Policy; Aarhus University (upcoming)

2016 Southern Economic Association; University of Notre Dame; Association for Psychological Sciences

2015 Association for Public Policy Analysis and Management; Association for Education Finance and Policy; Leadership, Policy, and Educational Opportunity: New Directions; Summer School on Socioeconomic Inequality (poster)

2014	Association for Public Policy Analysis and Management; Association for Education Finance and Policy
2012	North Carolina State Board of Education

Selected Awards and Honors

2016-17	Dissertation Year Funding, Northwestern University ($24,354)
2012-17	Sybil N. Heide Fellowship, Northwestern University ($18,700)
2015	School of Education and Social Policy Global Initiative Award, "Aarhus University-Northwestern University Workshop on Academic Performance" ($3,000)
2012	Outstanding Master's Project at the Sanford School of Public Policy
2012	United States Presidential Management Fellowship Finalist
2010-11	Joel L. Fleishman Endowment Funding, Duke University ($31,360)
2007	John Sheehan Prize for best senior economics thesis, University of Notre Dame
2005	Undergraduate Research Opportunities Program grant, "The Development of Political Ideologies at Notre Dame" ($500)

Teaching Assistantships

Winter 2017	Economics of Inequality; undergraduate course at Northwestern University
Spring 2016	Quantitative Methods III: Empirical Tools for Causal Quantitative Analysis; graduate course at Northwestern University
Fall 2015	Quantitative Methods I: Probability and Statistics; graduate course at Northwestern University
Spring 2013	Advanced Research Methods; undergraduate course at Northwestern
Spring 2012	Policy Analysis II; graduate course at Duke University
Fall 2011	Advanced Microeconomics and Public Policy; graduate course at Duke University
Spring 2011	Microeconomic Policy Tools; undergraduate course at Duke University

Advising

 # Indicates undergraduate student advisee

2016-17	Graduate advisor for #Abigail Durgan's undergraduate senior thesis, "The relationship between parental experiences of violence and current relationships with partners and children"
2016-17	Graduate advisor for undergraduate project, "Anticipating success: The relationship between emotions, perceived readiness, and perceived test performance on high-stakes tests" (with #Julie Blumenfeld, #Isabel Hoffman, and Emma Adam)

Professional Service

Ad hoc	Reviewer for Journal of Human Resources; American Journal of Education; PLOS ONE
Mar. 2017	Panel chair, Association for Education Finance and Policy Conference
Nov. 2015	Panel chair, Association for Public Policy Analysis and Management Conference
Spring 2015	Prospective and admitted student outreach, Human Development & Social Policy
2014-15	Organizer, Economics Working Group, School of Education and Social Policy
Oct. 2014	Panel participant, Arts & Letters: Thinking about graduate or professional school, University of Notre Dame
2011-12	Professional and Management Courses Chair, Sanford Academic Committee
2011	Secretary/Honor Board Representative, Sanford MPP Student Council

Community Service

2016-17	Notre Dame Fan Council
Summer 2013	Notre Dame Chicago Student-Alumni Mentoring Program
2011-12	Young Alumni Representative, Notre Dame Club of Eastern North Carolina

Professional Positions

2013-16	Research Assistant; Institute for Policy Research, Northwestern University, Evanston, IL
2007-10, 2012	Project Consultant (2012); Associate (2010); Analyst (2007-09); Huron Consulting Group – Higher Education, Chicago, IL
2011-12	Consultant; New Teacher Center, Durham, NC
2011	Financial and Business Services Research Intern; North Carolina Department of Public Instruction, Raleigh, NC
2010-11	Research and Policy Intern; Action for Children North Carolina, Raleigh, NC

Professional Affiliations

American Economic Association, Association for Public Policy Analysis and Management, Association for Education Finance and Policy, Association for Psychological Science