NBER WORKING PAPER SERIES

THE PSYCHOSOCIAL EFFECTS OF THE FLINT WATER CRISIS ON SCHOOL-AGE CHILDREN

Sam Trejo
Gloria Yeomans-Maldonado
Brian Jacob

Working Paper 29341
http://www.nber.org/papers/w29341

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2021

This work has been supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1656518 and by the Institute of Education Sciences under Grant No. R305B140009 and R305B170015. We are grateful to Sean Reardon, Carolyn Hoxby, Eric Bettinger, Ben Domingue, Jeremy Freese, Tom Dee, David Rehkopf, Eli Ben-Michael, Avi Feller, Jesse Rothstein, Emily Morton, Marissa Thompson, and Paula Lantz for helpful comments. We also wish to thank Jasmina Camo-Biogradlija, Kyle Kwaiser, Jonathan Hartman, Nicole Wagner, and the Education Policy Initiative staff for assisting with access to Michigan's restricted-use educational data. This research used data structured and maintained by the MERI-Michigan Education Data Center (MEDC). MEDC data is modified for analysis purposes using rules governed by MEDC and are not identical to those data collected and maintained by the Michigan Department of Education (MDE) and/or Michigan's Center for Educational Performance and Information (CEPI). Results, information, and opinions solely represent the analysis, information and opinions of the authors and are not endorsed by, or reflect the views or positions of, grantors, MDE and CEPI or any employee thereof. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Psychosocial Effects of the Flint Water Crisis on School-Age Children
Sam Trejo, Gloria Yeomans-Maldonado, and Brian Jacob
NBER Working Paper No. 29341
October 2021
JEL No. I10,I21,I28,I30,J01,J18

ABSTRACT

Lead poisoning has well-known impacts for the developing brain of young children, with a large literature documenting the negative effects of elevated blood lead levels on academic and behavioral outcomes. In April of 2014, the municipal water source in Flint, Michigan was changed, causing lead from aging pipes to leach into the city's drinking water. In this study, we use Michigan's universe of longitudinal, student-level education records, combined with home water service line inspection data containing the location of lead pipes, to empirically examine the effect of the Flint Water Crisis on educational outcomes of Flint public school children. We leverage parallel causal identification strategies, a between-district synthetic control analysis and a within-Flint difference-in-differences analysis, to separate out the direct health effects of lead exposure from the broad effects of living in a community experiencing a crisis. Our results highlight a less well-appreciated consequence of the Flint Water Crisis – namely, the psychosocial effects of the crisis on the educational outcomes of school-age children. These findings suggest that cost estimates which rely only on the negative impact of direct lead exposure substantially underestimate the overall societal cost of the crisis.

Sam Trejo
Princeton University
106 Wallace Hall
Princeton, NJ 08544
samtrejo@princeton.edu

Gloria Yeomans-Maldonado
Children's Learning Institute
University of Texas Medical Center
7000 Fannin, Suite 2300
Houston, TX 77030
Gloria.Yeomans@uth.tmc.edu

Brian Jacob
Gerald R. Ford School of Public Policy
University of Michigan
735 South State Street
Ann Arbor, MI 48109
and NBER
bajacob@umich.edu

1. Introduction

In January of 2016, the eyes of America became fixed firmly upon Flint, Michigan. National news outlets reported that Flint's water supply had been contaminated with high levels of lead. After months of state officials insisting the tap water was safe to drink, then-Michigan Governor Rick Snyder declared a state of emergency and called in the National Guard to distribute bottled water. Within weeks, the Flint Water Crisis was classified as a federal disaster and the Environmental Protection Agency took over management of the town's water supply (EPA 2016). By then, the roughly 100,000 citizens of Flint had been exposed to polluted water for over a year and a half. The majority Black, industrial city quickly became a national symbol for governmental negligence and racial injustice.

A substantial medical and social scientific literature documents the relationship between lead poisoning in early childhood and future cognitive and behavioral challenges. Lead exposure during childhood is associated with a host of negative outcomes, including increased anxiety (Winter and Sampson 2017), increased behavioral problems (Washerman et al. 1998), lower levels of self-regulation and executive functioning (Canfield, Gendle, and Cory-Slechta 2004), and decreased academic achievement (Amato et al. 2012). In adulthood, individuals exposed to lead as a child have decreased brain volume (Cecil et al. 2008), higher rates of criminal offending (Beckley et al. 2018), and decreased social mobility (Reuben et al. 2017).

In this study, we document another, less well appreciated, consequence of the Flint Water Crisis – namely, the psychosocial effects of the crisis on the educational outcomes of school-age children. Previous research has found that while the Flint Water Crisis led to only modest increases in child blood lead levels *on average* (Zahran, McElmurry, and Sadler 2017)[1], it

[1] The authors find an average increase of roughly 0.5 μg/dL, which they describe as similar to that of the annual change in lead exposure experienced by children in Flint from winter to summer

resulted in substantial increases among a small proportion of children. Hanna-Attisha et al. (2016) find that the percentage of Flint children with elevated blood lead levels doubled, from roughly 2.5% to 5%, as a result of the crisis. Given the low levels of exposure on average, and the fact that the consequences of lead contamination are known to be most severe for infants and toddlers, it is not clear if one would expect the lead exposure caused by the crisis to have had a noticeable impact on average educational outcomes in Flint in the short run. However, the crisis also had a profoundly negative psychosocial impact on Flint residents, who reported increased anxiety, depression, post-traumatic stress, sleep problems and worries about physical health (Brooks and Patel 2020).

To quantify the impact of the Flint Water Crisis on several important educational outcomes, we utilize the universe of longitudinal, student-level education records for the State of Michigan matched to address-level information on the water service lines in Flint collected during inspections conducted shortly after the crisis.[2] We first employ a between-district analysis using synthetic control methods (Abadie 2021; Ben-Michael, Feller, and Rothstein 2021; Arkhangelsky et al. 2018) to compare educational outcomes in Flint to outcomes in observationally similar school districts throughout Michigan. Such an analysis captures both the direct health impact of lead exposure as well as any psychosocial effects resulting from the upheaval of the crisis on individual children.

Second, we employ difference-in-differences methods to compare the academic trajectories of Flint children living in homes with dangerous service lines (i.e. lead or galvanized steel) to children in the same neighborhoods living in homes with safer service lines (i.e. copper),

[2] Public education records may well be the *only* data on virtually all children living in Flint, which has very low rates of private school attendance that begins before the crisis and follows them longitudinally.

which present low risk of lead exposure.[3] Past research has shown that Flint children living in homes with lead pipes consumed 4.5 times the amount of lead per day than children living in homes with copper pipes (Zahran et al. 2020). Even after controlling for a detailed set of house characteristics, there is substantial variation in the presence of lead service lines across houses within census blocks. Moreover, there is no association between the presence of lead service lines in a child's home and an extensive set of student demographic variables and baseline academic measures, suggesting that lead exposure may have been quasi-randomly distributed across children within a census block. This within-Flint analysis isolates the direct health effect of lead exposure resulting from lead service lines on the academic outcomes of Flint children.

The results of our between-district analysis suggest that the Flint Water Crisis induced a 0.14 standard deviation decrease in math achievement and a 9% increase in the number of students with a qualified special educational need. We find limited or no evidence for effects on reading achievement or daily attendance. In our within-Flint analysis, however, we find little to no difference in the academic outcomes of children living in homes with lead pipes compared to observationally similar children living in the same neighborhoods in homes with copper pipes.

Together, these results suggest that, for school-age children, the broad psychosocial effects of the Flint Water Crisis were larger than the direct health effects of lead poisoning on these educational outcomes. This finding is consistent with a large extant literature on the psychosocial consequences of adverse community events. From a policy perspective, our

[3] In Flint, lead service lines, galvanized steel service lines, and a small fraction of service lines made of non-standard materials were all considered dangerous (and scheduled for replacement) because they could leach lead into the water supply. For the sake of parsimony, we henceforth use the term "lead" pipes to refer collectively to housing with any pipe material (lead, galvanized steel, or non-standard materials) that increased the risk of lead contamination. Nonetheless, just 2% of home that we include in our lead variable are comprised only of other dangerous pipes (i.e., galvanized steel or non-standard materials). Likewise, we refer to pipes that constituted minimal risk of exposure as "copper" pipes. In Table A8 of the appendix, we probe the robustness of our results with varying definitions of definitions of lead and copper (i.e. dangerous and not dangerous) service lines.

findings suggests that existing estimates of the effects of lead exposure on child outcomes may substantially underestimate the overall cost of crises like the one that occurred in Flint.

Our research contributes to the literature in several ways. First, we quantify the educational costs of a famous case of government mismanagement. Second, we provide the first quasi-experimental study of a lead poisoning event due to lead plumbing in contemporary times. Third, we explore of the effects of lead on children who are exposed when they are above the age of five. Fourth, we highlight the indirect ways in which experiencing a community crisis can itself profoundly affect the educational development of school-age children.

At the same time, the present study has important limitations. Most importantly, because we observe only school-age children in our data, we cannot speak to the educational consequences of the Flint Water Crisis on infants or young children. Moreover, our analysis is limited to specific educational outcomes captured in state administrative data, such as standardized test scores and special needs classification. Hence, we are therefore unable to detect impacts on a variety of meaningful behaviors and skills. While our parallel identification strategies help us to separate the direct health effects of lead from the psychosocial effects of the ensuing crisis, we cannot identify precisely pathways through which any indirect psychosocial effects may have operated. Finally, community organizations in Flint responded to the crisis by dramatically expanding the set of social, medical and educational services available to children in Flint, suggesting that even our between-district estimates understate the negative effects of the crisis alone.[4]

The remainder of the paper proceeds as follows. In Section 2, we discuss the prior literature on lead poisoning and the psychosocial impacts of crises, with an eye toward their

[4] See, for example, the work of the newly formed Flint Pediatric Public Health Initiative, https://msuhurleypphi.org

effects the cognitive and emotional development of children. In Section 3, we provide relevant background information on the Flint Water Crisis. In Section 4, we describe our between-district data, analysis, and results. In Section 5, we describe our within-Flint data, analysis, and results. We interpret and discuss our findings and conclude the paper in Section 6.

2. Background Literature

2.i. The Effects of Lead Poisoning

Lead is a powerful neurotoxin with no known safe level of exposure (Cecil et al. 2008; Reuben et al. 2020).[5] While lead exposure has fallen dramatically over the past 40 years, estimates suggest that over 500,000 children under the age of five in the United States today still have elevated blood lead levels. Lead exposure is particularly dangerous for young children, who are comparatively small in size (and therefore more susceptible to low amounts of lead) and in the midst of a critical period for brain development (Lidsky and Schneider 2003). Lead exposure during childhood is associated with a host of negative outcomes, including increased anxiety (Winter and Sampson 2017), increased behavioral problems (Washerman et al. 1998), lower levels of self-regulation and executive functioning (Canfield, Gendle, and Cory-Slechta 2004), and decreased academic achievement (Amato et al. 2012). In adulthood, individuals exposed to lead as a child have decreased brain volume (Cecil et al. 2008), higher rates of criminal offending (Beckley et al. 2018), and decreased social mobility (Reuben et al. 2017). Unfortunately, correlational analyses may suffer from confounding due to omitted variables,

[5] The CDC has repeatedly lowered the threshold level of concern for children's blood lead, from 20 µg/dL to 10 µg/dL to 5 µg/dL (CDC 2013). Substantial evidence suggests that even amounts of lead below the current threshold, 5 µg/dL, can lead to intellectual and behavioral impairment (Canfield et al. 2003; Winter and Sampson 2017; Reuben et al. 2017).

such as socioeconomic status. In addition, the extent to which lead exposure is associated with negative outcomes is less well understood in children who are exposed at ages greater than five.

A growing quasi-experimental literature documents the causal effects of lead on downstream health and human development.[6] Much of the extant research leverages exogenous variation in exposure to lead resulting from public health programs that test children's blood and screen homes for exposed lead paint.[7] Billings and Schnepel (2018) and Aizer et al. (2018) combine individual-level blood lead measures with administrative school data to show a negative relationship between childhood exposure to lead and future math and reading test scores. Sorensen et al. (2019) utilize publicly available aggregate data for the entire United States and leverage the timing of Department of Housing and Urban Development lead abatement grants to show that abatement programs reduce lead poisoning incidents by roughly 70%, increase overall academic achievement, and decrease racial disparities in achievement.

Another strategy that has been used to identify the causal effects of lead is to exploit large, regional changes in exposure to lead. For example, Reyes (2007; 2015) shows that the removal of lead from gasoline in the late 1970s as a result of the Clean Air Act explains part of the decrease of antisocial and risky behavior in adolescents and violent crime in adulthood. Aizer and Currie (2017) use a similar approach to show that lead exposure in childhood increases future school suspensions and incarceration. Historically, the expansion of lead pipes in the early 20th century corresponded with increases in infant mortality (Troesken 2008), decreases in

[6] While compelling, existing quasi-experimental studies using of blood lead measurements of children are not without their challenges. Measurement of lead is complicated by the fact that the amount of lead in the blood can fluctuate wildly, dissipates quickly after exposure, and is often not a reliable signal of the amount of lead in an individual's body (Lidsky and Schneider 2003), leading to attenuation bias.

military test scores (Ferrie, Rolf, and Troesken 2012), and increases in homicide rates (Feigenbaum and Muller 2016).

Table A1 of the appendix uses estimates from previous studies to benchmark the magnitude of the relationship between lead exposure and cognitive ability. A one microgram per deciliter (μg/dL) increase in blood lead is associated with approximately 0.03 SD reduction in cognitive ability, math achievement, and reading achievement. Thus, moving from no lead exposure to the CDC threshold of 5 μg/dL would translate to an approximately 0.15 SD effect on cognitive outcomes. Similarly, an increase of 0.5 μg/dL, the estimated change among children living in Flint during the crisis, would be expected to correspond to a 0.015 SD effect on cognitive outcomes.

2ii. Lead Contamination via Drinking Water

Lead has been a common material used in plumbing for over a thousand years.[8] For meaningful amounts of lead to contaminate tap water, there must both be lead pipes supplying the water *and* water corrosive enough to break down the interior of the pipes and cause leaching. Lead in water is uniquely difficult to contain; while most contaminants can be filtered out at the water plant, lead typically gets into drinking water at the end of the system through lead service lines, which run beneath the ground and connect individual residences to city water mains.[9]

Today, 30% of the community water systems use some lead pipes (Cornwell, Brown, and Via 2016). The most recent surveys suggest that 6.7 million homes, serving approximately 19

[8] In 1900, of the forty-six largest cities in the U.S. for which data are available on piping material, thirty-nine used lead pipes (Troesken 2006). Lead pipes were preferred because, while costly, they lasted longer and were more malleable than alternative materials, making them easier to bend around existing structures. However, lead particles from these pipes can leach into drinking water.
[9] As such, there is typically both a publicly-owned and privately-owned portion of each house's service line.

million Americans, are supplied by lead service lines across the country. Furthermore, though most homes have their water supplied through “lead-free” pipes, they are not entirely without risk. Lead-free pipes can contain up to 8% lead, which has caused lead poisoning in cities that never had lead pipes (Renner 2010). Moreover, lead in home faucets and other fixtures can contaminate drinking water.

2iii. The Psychosocial Impacts of Community Crisis

The neurological effect of lead poisoning is not the only pathway through which the events of the Flint Water Crisis may have negatively impacted children. A large literature suggests that social and psychological processes also have an important role to play in shaping the outcomes of individuals living in crisis-affected communities. For example, in assessing the long-term effects of the Chernobyl nuclear accident in Ukraine, a United Nations report concluded that the negative effects on mental health outcomes as a result of fear, anxiety, and trauma actually surpassed the negative physical health effects of exposure to radiation (Chernobyl Forum 2006).

There are a range of psychosocial pathways by which a crisis like Flint’s may affect child development. Traumatic events, such as terrorist attacks and natural disasters, are associated with negative psychological consequences to entire communities. For example, using interviews of individuals following major events such as the 1995 Oklahoma City bombing (Pfefferbaum et al. 1999; 2000), the September 11, 2001 terrorist attacks (Neria, DiGrande, and Adams 2011), and Hurricane Katrina (Spell et al. 2008), researchers have documented persistent psychological distress and trauma to residents of the affected regions. Importantly, this is often true even among those who were themselves not exposed to the direct impacts of the crises themselves (for example, children who were not present for the events of the Oklahoma City bombing and lived many miles away).[10]

A more recent, a causal literature leverages quasi-experimental variation in violent events across place and time to study the psychosocial effects of community-level trauma. Sharkey (2010) and Rossin-Slater et al. (2020) show that community exposure to deadly shootings negative affect children's academic performance and mental health, respectively. The effects on academic performance emerge almost immediately, and the effects on mental health persist for many years. Similarly, Gershenson and Tekin (2018) show schools near the locations of the 2002 DC sniper attacks experienced decreases in academic achievement, with the effects largely being driven by effects on children in high poverty schools.

Internationally, we observe similar trends with respect to the broad effects of adverse community shocks. A study of civil war and genocide in Cambodia (Omoeva et al. 2018) found that disruptions to primary education during civil conflicts decreased educational attainment and earnings decades later. In Colombia, increased regional exposure to violence resulting from the emergence of drug cartels led to negative impacts on the academic achievement and educational attainment of children, with effects on fetal and child physical and mental health as a probable mechanism for the observed effects.

In addition, racialized events within a community like the Flint Water Crisis can produce a sense of social marginalization and subsequent civil unrest, which has been shown to impact academic performance. For example, Gershenson and Hayes (2018) show that the weeks of protests in Ferguson, Missouri, following the 2014 police killing of 18-year-old unarmed Black man Michael Brown caused decreases in math and reading achievement, partly mediated through

[10] Notably, Spell et. al 2008 found that parental stress and mental health served as a mediator of the effects of Hurricane Katrina on children.

decreases in student attendance. Such events may limit instructional time and learning by increasing absences, school closures, and disruption to daily routines.

Finally, there has been a concern that children in Flint will be stigmatized as a result of the public perception of how the water crisis negatively affected children (Green 2019; Gómez and Dietrich 2018). Such stigmatization may negatively impact development through processes like interpersonal discrimination (Goosby, Cheadle, and Mitchell 2018) and stereotype threat (Spencer, Logel, and Davies 2016).

3. The Flint Water Crisis

3.i. Timeline

Flint, a mid-sized industrial city in east-central Michigan, has experienced severe economic decline over the last half-century. Since the 1960s, Flint residents had been supplied water from Lake Huron, provided by Detroit's Water and Sewerage Department. In 2011, with Flint's government bankrupt, then-Michigan governor Rick Snyder appointed the first of a long string of emergency city managers tasked with balancing the city's budget. To reduce costs, the City of Flint ordered that the flow of water from Detroit be shut off and replaced by the Flint River, a small river which runs through its downtown.[11] This raised concerns because the Flint River's water presents a challenge to treat due to high levels of bacteria and carbon concentrations (Masten, Davies, and McElmurry 2016). Moreover, this change shifted the responsibility of treating Flint drinking water from Detroit to Flint's own Water Service Center,

[11] The switch to the Flint River as the city's municipal water source was intended to be only a temporary measure while Flint and neighboring municipalities in Genesee County could form a new water system, the Karegnondi Water Authority. While this system began operation in 2017, Flint would never connect to it.

a small facility traditionally maintained as a backup facility that was ill-prepared to conduct Flint's water treatment.

On April 25, 2014, Flint's residents began receiving drinking water from the Flint River. Figure A1 in the appendix provides a timeline of the events surrounding the Flint Water Crisis. As is the case in many older cities, a large fraction of Flint's service lines were made of lead. The water from the Flint River was both corrosive and improperly treated, causing lead from the pipes to leach into the tap water (Pieper, Tang, and Edwards 2017). Almost immediately, Flint residents began to complain about the color, taste, and odor of their drinking water. While some Flint residents opted to begin consuming bottled water, many continued to drink the tap water as city and state officials insisted on its safety.[12]

In early 2015, local researchers and some employees in federal EPA began to raise concerns about elevated lead levels in the tap water of Flint residents as well as increased blood lead levels among Flint children. In October 2015, the city switched back to water from Detroit. However, the city's tap water remained unsafe to drink as a protective mineral film needed to develop over time inside the pipes to prevent further leaching. In January 2016, almost 18 months after the April 2014 switch to water from the Flint River, Michigan Governor Rick Snyder declared a state of emergency and called in the Michigan National Guard to distribute clean water. In August 2020, the state of Michigan reached a $600 million settlement with the victims of the crisis, with the bulk of the money going to those who were children. Today, the vast majority of lead service lines in Flint have been inspected and, if necessary, replaced.

[12] Famously, the Mayor of Flint appeared on local on local news to drink the tap water himself and publicly demonstrated its supposed safety (*WNEM* 2015).

3.ii. Prior Research on Flint Water Crisis

Prior research has found that blood lead levels in Flint children increased by roughly 0.5 µg/dL (0.2 SD) as a result of the crisis (Zahran, McElmurry, and Sadler 2017). In addition, the fraction of children identified with blood lead levels above the CDC's acceptable threshold (5 µg/dL) roughly doubled, rising from 2.5% to 5% (Hanna-Attisha et al. 2016), with the greatest increases in neighborhoods with highest water lead levels. While any increase in lead exposure can be damaging, the average blood lead increase experienced by Flint children in prior research (an approximately 20% increase from the pretreatment period mean of approximately 2.5 µg/dL), was relatively modest—similar to that of the annual change in lead exposure experienced by children in Flint from winter to summer (Laidlaw et al. 2016; Zahran, Mcelmurry, and Sadler 2017).[13] According to this research, even in the midst of the Flint Water Crisis, the fraction of child blood lead attributable to non-water sources (for example, paint, soil dust, and airplane fuel) was larger than water lead sources (Zahran et al. 2020). However, others in the medical and public health community in Flint argue that the blood level measures are not reliable, and likely understate the true extent of the exposure that children experienced during the crisis.[14]

There is somewhat mixed evidence on the causal effects of the Flint Water Crisis on neonatal outcomes. Grossman and Slusky (2019) examine the impact of the Flint Water Crisis on fertility and infant health using difference-in-differences by comparing childbirths in Flint to other Michigan cities before and after the water crisis. They find that the Flint Water Crisis

[13] Children's blood levels peak in summer and fall and retreat during winter and spring period. There are many potential explanations for this phenomenon, including seasonal changes to meteorological factors like precipitation, wind, humidity that suspend lead dust in soil, the fact that a greater amount of lead leaching has been shown to occur at higher water temperatures, and that opening and closing windows releases lead paint particulates.
[14] Personal communication with Mona Hanna-Attisha (September 10, 2021).

decreased fertility by 12% (which they attribute to a corresponding increase in rates of preterm pregnancy loss) and that the overall health of Flint newborns decreased. However, Gómez et al. (2019) found that lead levels in Flint females of childbearing age did not increase during the Flint Water Crisis and subsequent 18-month time period, casting doubt on lead as a pathway for the observed effects.

Regardless of the physical role of lead, residents of the Flint viewed the crisis as having a profoundly negative impacting their wellbeing. A recent review of 11 predominantly cross-sectional studies highlights the ways in which Flint residents perceived the crisis as increasing anxiety, depression, post-traumatic stress, sleep problems and worries about physical health (Brooks and Patel 2020). In addition, studies in this review found that residents reported that coping with these negative mental health consequences lead to increases in risky health behaviors such as smoking and alcohol misuse. Some studies also found that, among Flint residents, lower perceived tap water quality was associated with poorer mental and physical health, and that the negative psychological consequences of the Flint Water Crisis continued even after the state of emergency was lifted.

4. Between-District Analysis

In this section, we leverage a between-district analysis to determine the total impact of the Flint Water Crisis on student outcomes. That is, we compare the academic trajectories of students living in Flint to those of students living in observationally similar districts using data spanning 2006 through 2019. This strategy allows us to estimate the total impact of the Flint Water Crisis on student outcomes, including both direct health and psychosocial pathways.

4.i. Data

Our primary data source is student-level administrative educational records from the state of Michigan spanning 2006-2019. This data was provided by the Michigan Education Data Center and links students across time with a unique identifier. It contains annual information on all students in Michigan public schools (including charter schools), from pre-kindergarten through high school, and includes demographics, enrollment information (including attendance and mobility), and outcomes such as academic achievement. This data also contains United States Census and American Community Survey data taken from IPUMS that characterizes the demographic and socioeconomic characteristics of the census block groups in which students live.

We focus on four key educational outcomes: math achievement, reading achievement, special needs status, and daily attendance. Math and reading achievement are measured using annual state-administered educational assessments, which are given to students in grades 3-8 starting in 2007. We standardize these test scores at the grade-subject-year level using the distribution of all students in Michigan. Special needs status and daily attendance are observed for all students regardless of grade level (i.e. K-12), with the former available since 2006 and the latter available starting in 2009. We choose these four outcomes because (a) they are well-measured beginning many years prior to treatment and continuing through the Flint Water Crisis and (b) they are good theoretical candidates to be impacted by lead exposure and community crisis.

To construct a panel of Michigan districts, we first must assign each student to a school district. School choice is widespread in Michigan, and during the 2013-2014 school year only 45% of public-school students living in Flint city limits attended Flint Community Schools, Flint's zoned school district (the remainder of students living in Flint attended charter schools or neighboring districts and are still captured in our data). However, the Flint Water Crisis impacted all children living in Flint regardless of where they attended school. For this reason, we define our treatment group to include all students living within Flint, or identically, the geographic boundaries of the Flint Community Schools. Similarly, our comparison districts are defined based on students' zoned school district (i.e. their geographic district) as opposed to the district they actually attend (i.e. their administrative district).[15] For the remainder of the paper, we will use the term district to refer to a geographic school district. To facilitate the synthetic control analysis, we collapse the student-level data into a geographic district-year panel from 2006-2019.

In 2013-2014, Flint's 16,210 zoned students in K-12 made it the 9th largest residential district in the Michigan. Flint also has an exceptionally high fraction of economically disadvantaged students (89%) and fraction of Black students (76%). To ensure we are comparing Flint to similar districts, we first exclude very small school districts (specifically, the 185 districts with fewer than 1,000 students in the 2013-2014 school year) from our set of potential comparison districts. The exclusion of very small districts also helps reduce year-to-year fluctuations in outcomes that can occur with a small number of observations. Relative to these remaining Michigan districts with at least 1,000 students, Flint is at the 99th percentile in terms of fraction Black and fraction economically disadvantaged. Thus, we restrict our control sample to districts whose student composition place them in the top 10% of either of those two characteristics which make Flint an outlier. These criteria leave us with 54 potential control districts. See Appendix Figure A3 for a graphical illustration of the sample selection process and Appendix Table A2 for the complete list of potential comparison districts.

[15] In Column 3 of Figure A5 in the appendix, we explore whether our results differ depending on whether students attended their "home" school district or not.

Table 1 displays descriptive statistics for all Michigan geographic districts for 2013-2014. While the 54 potential control districts more closely mirror Flint in terms of size and demographics, Flint has notably lower academic achievement than this group. We now turn to our synthetic control algorithm to identify the weighted average of these 54 potential control districts that best approximates Flint's outcome trends in the pre-treatment period.

4.ii. Trends in Educational Outcomes in Flint

We begin by plotting academic outcomes over time in Flint (Figure 1). For math achievement, we observe a positive trend in the pretreatment period from 2007-2014 and then a drop of roughly 0.15 standard deviations in the first year following the water crisis. The negative trend that began in 2015 trend largely continues through 2019. The descriptive trend in reading achievement is similar but smaller in magnitude. With respect to special needs, we observe a positive trend before the crisis that appears to quicken following the water crisis. Turning to K-12 student attendance, we observe a steady decrease in attendance before the crisis that grows in 2017 and 2018.

The graphs in Appendix Figure A2 explore several additional features of the Flint trends over time. To explore whether a change in the fraction of students tested could explain any of the observed achievement trends, we plot the fraction of students tested but find no change around the time of the water crisis.[16] To examine whether selective out-migration may confound the picture, we plot the enrolment in the Flint district as well as the fraction of Flint students who left the district each year. Again, we see no sharp changes coincident with the water crisis.

[16] Some students with a severe disability or who have recently immigrated to the United States are not subject to state testing. In addition, students who do not come to school on the test day or make-up day will have missing test score information.

The trends shown in Figure 1 provide suggestive evidence that the Flint Water Crisis negatively impacted student outcomes. However, it is possible that other factors may have played a role, ranging from changing economic conditions in Michigan to changes in state education policies. For this reason, we turn to synthetic control methods to estimate how educational outcomes in Flint would have evolved in the absence of the water crisis.

4.iii. Empirical Strategy: Synthetic Control

To isolate the causal effect of the Flint Water Crisis, we utilize a synthetic control methodology (SCM) that compares Flint to other Michigan districts over time. SCM was first developed by Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmuller (2010). It is a matching estimator designed to be used in situations with a single treated unit and a small number of potential control units. Synthetic control methods use pre-treatment outcome data to identify the weighted average of control districts that most closely approximates the treated unit. This approach can reduce bias through improved pretreatment fit and allow for a more transparent counterfactual selection process.

We implement a recent extension of synthetic control methods known as demeaned, or intercept-shift, synthetic control (Doudchenko and Imbens 2017), which we summarize here and discuss in greater detail in Section A2 of the appendix. This approach involves first subtracting each treatment and control unit's pretreatment outcome mean from all pre- and post-treatment observations (conceptually similar to a unit fixed effect), and then fitting a classic synthetic control model on those residuals. In doing so, demeaned synthetic control methods match control units to the treatment unit using only data on pretreatment trends rather than on pretreatment means (i.e. levels). Put differently, the demeaned synthetic control, like the classic difference-in-differences model, compares only changes over time (rather than average cross-sectional differences) between treated and control units. While Flint's mean academic outcomes are extreme outliers in the overall state distribution, its trends resemble many other districts that lie closer to the area of common support (i.e. convex hull) of control outcomes; thus, demeaning drastically increases our common support between Flint and potential control units. At the same time, demeaning does not stretch our relatively short panel of pretreatment outcomes too thin, as would more complicated extensions, such as incorporating time weights, outcome modeling, or machine learning.

The demeaned synthetic control that we implement is identical to several synthetic control methods extensions recently suggested in the literature. For example, our approach is equivalent to using augmented synthetic control methods in the specific case where each unit's pretreatment mean is used as covariate with a coefficient constrained to be equal to 1 (Ben-Michael, Feller, and Rothstein 2021). Our approach is also identical to fitting synthetic difference-in-differences with uniform time weights and including a unit intercept (Arkhangelsky et al. 2018). As such, our demeaned synthetic control estimator can be identically expressed as a weighted difference-in-differences estimator. While early versions of synthetic control methods balanced both lagged outcomes and covariates, we follow the recent literature (Doudchenko and Imbens 2017; Arkhangelsky et al. 2018; Ben-Michael, Feller, and Rothstein 2021) and choose to balance only lagged outcomes.[17]

One potentially unsatisfying aspect of existing synthetic control methods is that in the standard approach each implementation is specific to only a single outcome. Because of this, the

[17] Covariates have been shown to be redundant in synthetic control when used alongside all lagged outcomes (Kaul et al. 2021). Thus, our demeaned synthetic control estimates the causal effect of the Flint Water Crisis on a given outcome using a demeaned weighted average of only that outcome for the selected control units.

synthetic control weights for a given treated unit might different substantially across the multiple outcomes considered (Jardim et al. 2017). For example, it might be odd if the synthetic Flint for math achievement is comprised of an entirely different set of control districts that the synthetic Flint for reading achievement. We address this multiple outcome issue by finding a single set of synthetic control weights that simultaneously balances all four of our educational outcomes in the pretreatment period. In addition, using more pretreatment data to identify our synthetic control weights helps maximize their stability and thereby increase the statistical power of our analysis.

4.iv. Results

Figure 2 presents the synthetic control estimates described above. The thick blue line measures the difference between student outcomes in Flint versus the synthetic control group. Starting in 2015, the blue line reflects the treatment effect of the Flint Water Crisis and the shaded grey area corresponds to the confidence interval around the estimate. Panel A of Table 2 lists the ten control districts assigned a non-zero weight in the construction of the Flint counterfactual. Four districts play a large role in the comparison: Oak Park, Lincoln, Dowagiac and River Rouge. River Rouge and Oak Park are small urban districts near Detroit with very high percentages of both low-income and Black students. Dowagiac is a predominantly white district (roughly 65%) located in the Southwest part of the state with very high poverty levels (75%). Lincoln is an urban district located near Ypsilanti, Michigan, roughly 45 minutes southwest of Detroit. While the selected comparisons are similar to Flint in these demographic characteristics, the demeaned synthetic control model selects comparison districts whose pre-2014 educational *trends* most closely match Flint's. Hence, the most important test of the SCM model is how closely the outcome trends of the controls match Flint's prior to the water crisis. As one would hope, the blue line prior to 2014 is nearly zero, suggesting a very good match. The root mean square error in the pretreatment period (pre-RMSE) provides a formal measure of model fit, and are shown in the Panel B of Table 2. The scaled pre-RMSE (.07 SD) is generally low,[18] showing that even while balancing four outcomes simultaneously, we are able to provide reasonably good fit in the pretreatment period.[19]

Our results suggest the Flint Water Crisis negatively impacted student performance along several dimensions. Math achievement in Flint closely tracks the comparison districts from 2006 through 2014, but drops notably starting in 2015. The SCM estimates indicate a 0.14 standard deviation decrease in student performance in math (Table 2). The 0.14 SD effect on math achievement falls within the range of "medium" effect size according to recently introduced standards for educational interventions (Kraft 2020). However, when you consider both the large size of the treated group and the fact that the effect is negative, math results might be considered within the "large" effect size category. We did not observe a significant detectable effect on reading achievement in our main analysis (though our robustness checks yielded some suggestive results). This pattern of results is consistent with many other studies that suggest student math performance is more malleable to interventions than reading performances (Alexander, Pitcock, and Boulay 2016). We find no impact on student attendance, which is somewhat surprising given the social and community upheaval associated with the crisis.

[18] The unit of the pre-RMSE represents the standard deviation of the outcomes among the final 55 districts provided to the synthetic control software. Math achievement and special needs, for example, have SD of .2 and .02, respectively, in our 54 district synthetic control sample (See Table 1). Thus, rescaling the pre-RMSE back to the original units of our outcome variable's yields pre-RMSEs of .014 SD *of the statewide distribution* and 0.14 percentage points, respectively.

[19] Note that the individual RMSE values in Table 2 are displayed in the specific units of each outcome (e.g., standard deviation of the overall Michigan distribution of achievement, proportion for special needs, etc.) while the overall RMSE value is scaled to each variable in our data's standard deviation.

Finally, the graph in the bottom left of Figure 2 suggests that the Flint Water Crisis led to a 1.3 percentage point (9%) increase in the proportion of students with a documented special educational need. It is possible that the heightened attention on Flint schoolchildren during the crisis led to greater screening, which in turn increased special needs diagnosis rates while leaving underlying population characteristics unchanged.[20] On the other hand, the concurrent reduction in math performance, which is not subject to bias due to increased screening, suggests that underlying needs in the population likely increased as well. We suspect the observed increase reflects some combination of both forces, though it is difficult to know the exact magnitude of this increase.

Section A3 of the appendix presents the results from a series of alternative specifications meant to assess the sensitivity of our primary estimates. In particular, one might be concerned about the slight uptick in special needs classifications in Flint that is evident *prior* to the water crisis in the primary results (see Panel C of Figure 2). As described in in the prior section, we estimated a single synthetic control model simultaneously for all four outcomes. A downside of this approach is that it will not maximize the pre-treatment fit for each outcome. For this reason, we fit alternative versions of the synthetic control that consider each outcome individually as well as models that include two of the four outcomes together (one model that includes math and reading achievement and another model that includes attendance and special need services together).[21] The alternative specifications for special needs, which have improved fit and display

[20] Lawsuits brought by the ACLU and Education Law Center against the city of Flint spurred efforts to identify children in need of academic support. See, for example, https://www.aclumich.org/en/press-releases/flint-students-secure-groundbreaking-gains-settlement-special-education-class-action.

[21] A number set of the models considering only a single outcome attain perfect fit in the pretreatment period. Perhaps counterintuitively, perfect fit is problematic for synthetic control methods; it implies that the models do not converge to a sufficiently unique solution and that we need either a longer pretreatment panel or more outcomes to balance. We therefor omit all results that attain perfect fit.

no pre-treatment trend (see Figure A4), show that the potential pre-treatment trend is not driving our results.

We also estimate models using a restricted set of potential comparison districts. Rather than using the 54 districts that are in the top 10% in terms of fraction Black or fraction economically disadvantaged, we use only the 26 districts that are in the top 5% in terms of at least one of these characteristics. Estimates based on this restricted sample are comparable to our baseline estimates.

Finally, we test the robustness of our results to student residential mobility. The key concern is that the treatment might have led to compositional changes in Flint, which could bias our estimates. Because *geographic* school districts are our unit of observation, a student who transfers to a school in a neighboring district will remain in our Flint sample, unless they change their residential address to outside Flint. The mobility trends for Flint shown in Appendix Figure A5 suggest that this is not a concern in practice. As a further check (see Tables A6 and A7), we re-estimate our SCM using a panel of data based on only students living in Flint or the comparison districts during the 2013-2014 school year. Using this "fixed sample" approach, a student that leaves Flint following the Water Crisis will nonetheless be contribute to the Flint average in these later years. Results based on this "fixed" sample are comparable to our primary estimates.[22]

In order to explore heterogeneity in the effects of the Water Crisis, we decompose our effect estimates by subgroup, based on gender, grade level and administrative district (see Appendix Table A5).[23] The effects on math achievement were larger for children who were

[22] Very few students from any district leave the state or transfer to private schools so, in practice, attrition from our statewide database of public school students is not a concern.
[23] We describe this decomposition further in Section A2.iv of the appendix.

younger at the time of the crisis and for students who attended a school in the Flint administrative district (rather than a charter school or a school in a neighboring district). The effects on special needs status were larger for boys than for girls.

5. Within-Flint Analysis

In this section, we conduct a within-Flint analysis to better isolate the effects of lead exposure itself on student outcomes. In particular, we compare the educational trajectories of Flint children living in housing units that were exposed to different levels of lead contamination during the water crisis. We ask: How did Flint children who lived in homes with lead pipes fare compared to Flint children who lived in homes with copper pipes before and after the crisis? We begin by describing our student-level panel construction. Next, we discuss the difference-in-differences methods we utilize to estimate the causal effect living in home with lead pipes during the Flint Water Crisis. Finally, we present event study and difference-in-difference results.

5i. Data

To identify households with the greatest risk of exposure to lead contamination, we use data on the materials of the water service lines running to individual buildings (we call this the "pipes data"). This data comes from service line inspections conducted by Flint's Fast Action and Sustainability Program (FAST Start), a team of city- and state-appointed officials who were tasked with managing lead service line replacement following the Flint Water Crisis. [24] In total,

[24] The data was generously provided to us by academic researchers at the University of Michigan and Georgia Tech University who partnered with the FAST Start team to help with data management and refine the prediction of service line material (Abernethy et al., 2018). This team developed a web and mobile application where on-site contractors as well as state and local officials filled in essential information about service line work accomplished at

this data includes 24,646 unique parcels with valid inspection data. Notably, this set of parcels contains most but not all occupied properties in Flint, which restricts the number of home addresses, and in turn Flint children, that we are able to include in our within-Flint analysis.[25]

The inspection data indicates the material used in both the public service line (i.e., the pipes running from the city water supply to the private home) and the private service line (i.e., the pipes running from the public lines into the specific unit and to the faucets within the home). In addition to lead, galvanized steel is considered a dangerous material for service lines that have at one point contained lead in the system because they have been known to capture small pieces of lead in the corrosion in their inner walls. Even if the lead piping was removed from a water system long ago, galvanized steel pipes can still release trapped lead into the water. In contrast, copper service lines are considered very safe and are the standard service line material used in new pipe construction today. For our main analysis, we follow the approach used by the FAST Start team and consider any home with lead, galvanized steel, or an unknown material in either the public or private service lines to be a danger in terms of lead exposure. In Table A8, we show that our results are robust to alternative definitions, such as only considering lead lines to be dangerous and copper service lines to be not dangerous. In addition to information on service line materials, the inspection data includes a host of other variables describing the housing unit, including: the use type (residential, commercial, industrial), rental or owner-occupied, year built, the condition of the house unit in 2014 (good, fair, poor, or structurally deficient), assessed building value, and assessed land value.

each site. These researchers have since formed the company BlueConduit to continue their municipal lead service line prediction work. Our data was extracted for their database on January 8, 2020.

[25] According to the city of Flint's 2016 tax parcel data, there were 31,685 residential properties in Flint that were not vacant. However, at the time of the crisis, only 26,642 properties had active water accounts and met certain minimum criteria regarding age of house and lead risk were eligible to be inspected and potentially replaced. In addition, not all of the eligible properties have been replaced; only 24,646 of the eligible parcels had valid inspection data as of 2020, when the latest update was provided to our research team.

Past research utilized the same service line inspections data, combined with home water testing results, to show that service lines were a key source of lead exposure in Flint. During the peak of the crisis, children living in homes with lead service lines consumed 4.5 times the amount of lead per day than children residing in homes with copper pipes (Zahran et al. 2020).[26] However, by 2017, two years after switching back to Detroit's water, children in homes with lead service lines were consuming virtually no lead in their home drinking water and there were no longer meaningful differences in lead exposure across homes with varying service line material types.[27] By matching data on residential water tests with data on pipe material for a subset of students (see Section A4 of the appendix for more details), we were able to confirm that similar patterns applied in our analytic sample. In particular, in our student sample, homes with lead or galvanized steel service lines were 4.7 percentage points (P<.01), or 77%, more likely to exhibit water lead levels above the EPA threshold of 15 ppb.

To create our analysis sample, we start with the 17,024 students who were living in Flint and enrolled in Michigan public schools during the 2013-2014 school year (the final academic year before the Flint Water Crisis).[28] Using a probabilistic matching algorithm based on street number, street direction, street name and street type, we were able to identify the service line material for 10,245 students, or 60.18% of the initial sample.[29] A small fraction of the non-

[26] The same study found that during the Flint Water Crisis, galvanized steel service lines, which we include in our primary "lead" treatment definition, were associated with 1.8 times more childhood lead consumption with copper pipes.
[27] It is important to recognize that lead service lines were not the only source of lead release. Premise plumbing (plumbing inside a home) has lead and children consumer water from multiple locations. For example, Flint schools had significant water lead levels despite the fact that schools did not have lead service lines.
[28] This analytic sample is slightly larger than the sample used in our between-district analysis. There, we focus only on children in kindergarten through 12th grade because districts across the state may vary in their pre-K enrollment and high school matriculation policies. In our within-Flint analysis, where our identification strategy relies on within-student comparisons over time, we elect to include observations of students in pre-K and in ungraded classrooms.
[29] This address-level match was executed on our behalf by staff at the University of Michigan, who obtained special permission from the Michigan's Department of Education to use personally identifiable information (in particular,

matches, roughly 1%, were due to missing or invalid address information in either the pipes data or the student education records. A manual review conducted on a random subsample of the data suggests that our matching algorithm failed to match roughly 18% of true matches. Thus, the primary reason for unmatched students is likely the limited coverage of the pipes data.

While the match rate will not affect the internal validity of our estimates, it may limit the generalizability of our findings. To help understand how well results from the matched sample may generalize to the full population of Flint students, Table 3 compares the full set of Flint students to the set whose addresses matched the pipes data. The matched students appear extremely similar to the unmatched students, particularly when we compare students within the same census blocks (Column 5). For example, 74% of matched students are Black compared with 79% of unmatched students, but if we limit our comparison to students within the same census block, this modest difference completely disappears. There are modest but noticeable differences in terms of the schools attended, even within census blocks. Students who matched to pipes data were less likely to be attending charter schools (27% vs. 35%) and more likely to be attending traditional public schools in either Flint or a neighboring school district. Matched students were living in census blocks with slightly lower poverty rates (40% vs. 45%). The overall similarity of students in the matched and unmatched samples suggests that our results should generalize well to the broader population of Flint children.

5ii. Empirical Strategy: Difference-in-Differences

Using the language of program evaluation, our within-Flint analysis considers children living in housing units with lead pipes in 2014 as the treatment group and their peers living in

---

historical address data from all Michigan students). Students' home addresses, as well as all other personally identifiable information, were stripped from the matched data before it was returned to our research team.

housing units with copper pipes as the control group. If children were randomly distributed across housing units in Flint, one could estimate the impact of potential lead exposure simply by comparing outcomes of children living in homes with and without lead pipes.

However, given Flint's historical growth patterns and the decreases over time in the installation of lead pipes, there is reason to believe that homes in certain neighbourhoods are more likely to have lead piping than others.[30] Consistent with this intuition, we find that 37.5% of the variation in the presence of lead pipes occurs between (as opposed to within) census blocks. To examine the determinants of lead pipes in our analytic sample, Table 4 presents results from several regression models that include both household-level and neighborhood-level predictors. Column 1 shows that a small set of house characteristics are strong predictors of the presence of lead piping. For example, older houses are more likely to contain lead pipes; all else equal, a house that is 10 years older is 20 percentage points more likely have lead pipes, a large effect given that just 40% of all houses have lead pipes. Conditional on age, rental status, and house condition, a $10,000 increase in the assessed home value is associated with a 0.6 percentage point reduction in the likelihood of lead service lines. Column 2 shows that, conditional on house characteristics, units located in higher poverty census blocks are more likely to have lead piping, although the magnitude of this relationship is modest; a 10 percentage point increase in the poverty rate is associated with a 2.5 percentage point increase in the likelihood of lead piping.[31] When we include census block fixed effects in Column 3, the housing unit characteristics remain significant predictors. However, they only explain about 33%

[30] An 1897 Flint city ordinance required "all connections with any water mains be made with lead pipe" (Masten, Davies, and McElmurry 2016), and the use of lead pipes as service lines was slowly phased out over the 21st century.
[31] Interestingly, the magnitude of the coefficient on the neighborhood poverty variable does not change much if one omits the other census block characteristics.

of within-block variation, suggesting that considerable variation in the presence of dangerous piping even among observationally identical housing units located in the same census block. This is consistent with a quasi-random distribution of pipe materials across houses of similar vintage in similar neighborhoods, which is unsurprising given that the city and homeowners alike largely did not know which houses had lead services lines prior to the inspections in the wake of the crisis. This fact forms the basis for our within-Flint empirical strategy, described below.

We next examine the relationship between pipe material and student and school characteristics. Table 5 compares the characteristics of students living in homes with and without lead pipes. Looking at Column 4, we see students living in homes with lead versus copper pipes have virtually identical test scores and virtually identical rates of disability and daily attendance. Consistent with the results shown earlier, students in homes with lead pipes are more likely to be Black (7.9 percentage points) and economically disadvantaged (3.1 percentage points). Columns 5 and 6 present the difference across groups from regressions model that includes census block fixed effects, ensuring comparisons between students living in the same census blocks. Additionally, the specification reported in Column 6 controls for the house characteristics listed in Table 4. Once we condition on neighborhood and house characteristics, the small to modest differences between students living in homes with lead versus copper pipes diminish substantially. Indeed, none of the 11 student characteristics in the top panel are statistically significant at the 5% level in Columns 5 or 6. Across 13 individual demographic characteristics and measures of academic performance, there are only 3 differences that are statistically significant at the 5% level. The only school characteristics that remain significant are school location (city versus suburb), magnet school, and racial composition of the school.

Children living in homes that were most susceptible to lead exposure were observationally similar to children in homes with copper pipes that presented no lead danger. Nonetheless it is possible that the two groups of children differ in unobservable ways that influence their educational outcomes after the water crisis. To account for unobservable time-invariant child and family characteristics, we estimate difference-in-differences models.

To this end, we construct a student-year panel for the 10,245 matched students that runs from the 2009-10 through 2018-19 academic years. This allows us to examine up to five years for each child prior to the water crisis and up to four years following the crisis. Note that this is an unbalanced panel because not all children appear in all years (e.g., many of the older children in 2014 will have graduated or dropped out prior to 2019, and the very young children in 2014 were not yet enrolled in public school in 2010).

In order to generate difference-in-difference estimates of the average treatment effect of having lead pipes during the Flint Water Crisis, it is standard to estimate a two-way fixed effects model for outcome y for student i in year t:

$$y_{it} = \lambda_i + \delta_t + P_t^{t>2014} D_i + \varepsilon_{it}$$

(5.ii.a)

λ_i: Student fixed effects

δ_t: Year fixed effects

$P_t^{t>2014}$: Post-treatment period dummy variable

D_i: Lead service line dummy variable

The identifying assumption of the difference-in-differences model is commonly referred to as "parallel trends," meaning that in the absence of the exposure, the outcomes of the treatment group would follow the same path as that of the control group. A standard approach to testing the plausibility of this assumption is to estimate an event study model, which is simply an extension of the two-way fixed effects model above that allows the effect of the exposure to differ by year relative to the start of the treatment. This corresponds to the following regression model:

$$y_{it} = \lambda_i + \delta_t + \sum_{T=2011}^{2014} \gamma_t^{t=T} D_i + \sum_{T=2015}^{2019} \gamma_t^{t=T} D_i + \varepsilon_{itk}$$

(6.ii.b)

$\gamma_t^{t=T}$: Year dummy variable

The coefficients γ_t trace out the differences by year in outcomes between students living in homes with and without lead service lines, with 2010 serving as the omitted year. The estimates of γ_t from 2011 through 2014 serve as a specification check, whereas the estimates from 2015 to 2019 reflect the impact of greater exposure to lead.

However, a recent and growing literature highlights several important limitations of the standard two-way fixed effect differences in differences approach (de Chaisemartin and D'Haultfœuille 2020; Borusyak, Jaravel, and Spiess 2021; Goodman-Bacon 2021; Callaway and Sant'Anna 2020). Perhaps most importantly, these papers point out that the canonical regression difference-in-differences estimate relies on the strong assumption of treatment effect homogeneity. This is because the two-way fixed effects estimators produce a weighted average of many different comparisons between never treated, yet to be treated, and already treated units. In certain cases, these weights can even be negative, producing a situation in which the estimate has a different sign than the true parameter.

For this reason, we utilize the novel imputation estimator proposed by Borusyak, Jaravel, and Spiess (2021), henceforth BJS, which has the dual advantages of being intuitively clear and straightforward to implement. We first estimate the two-way fixed effects model described above, using only the untreated observations, which includes all observations of the control students (i.e., those living in homes without lead pipes) as well as the pre-treatment observations for treatment students (i.e., years 2010 through 2014). Using these estimates, we generate predicted (imputed) values of the outcome for treatment students in the post-treatment period. For each student-year observation, we then calculate the difference between their observed and predicted outcome (which each correspond to students' potential outcome in the treated state and untreated state, respectively). These student-year "effects" can then be aggregated in various ways. To recover the canonical difference-in-differences estimate, we calculate a simple (unweighted) average of all treatment group students in all post-treatment years. To estimate the year-specific "event study" effects shown in equation 6.ii.b, we generate separate averages for each year. While visual inspection of these pre-exposure estimates can be useful, we also conduct formal tests of the null hypothesis that the pre-exposure effects are jointly significant. Standard errors that account for heterogeneity and serial correlation within students over time are calculated via a bootstrap procedure outlined in Borusyak, Jaravel, and Spiess (2021).[32]

[32] We implement this procedure using the Stata command ***did_imputation*** written by Borusyak, Jaravel, and Spiess (2021).

In our baseline models, we only include student and year fixed effects. To increase our statistical power and test the sensitivity of our results, we estimate additional models that include a variety of additional controls. We first add school-grade fixed effects to limit our comparison to students in comparable educational settings. We also add a host of further controls. We include race-gender-year fixed effects to allow the outcomes for different demographic groups to evolve differently over time. To allow outcomes to evolve differently for students in different grade cohorts, we include interactions between a student's grade in the year before the crisis (2013-2014) and year dummies. Finally, to allow outcomes for students in different neighborhoods to evolve differently over time, we include interactions between census block poverty in 2013-2014 and year dummies.

One final concern involves differential attrition. If the water crisis caused Flint students living in homes with lead pipes to exit the public school system differently than other Flint students, our difference-in-differences estimates could be biased.[33] Crucially, our data includes all Michigan public school students, so students that merely leave the Flint Public Schools do not exit our sample. To determine whether students living in homes with lead pipes in 2014 were more likely to leave the Michigan public schools, we estimated a series of OLS regression models. The coefficient on lead pipes was consistently very small and not close to statistical significance, regardless of the control variables included (including none) or the year examined. This suggests that attrition bias is not a concern in the results discussed below.

[33] General attrition from the public schools among all Flint students would not lead to bias.

5iii. Results

Figure 3 displays connected scatterplots of our four main academic outcomes, with the trends broken out separately by home service line material.[34] In it, we see only small differences between the trends of students with lead pipes and students with copper pipes before and after the crisis. These descriptive patterns foreshadow the largely small or null within-Flint results in our difference-in-differences and event study models.

The results from our event study models are displayed in Figure 4. Beginning with our baseline two-way fixed effect model in the left panel and focusing on the coefficients prior to treatment in 2014, we see little evidence for differential trends across students with and without lead pipes. In three of the four outcomes, a test of the joint significance of the pre-exposure effects is not significant. In the case of math achievement, however, we do reject the null hypothesis that treatment student trends were equal to those of control students. While there is not a clear pattern to the pre-exposure trend, math achievement appears to dip in the final pretreatment year.

As described earlier, because students were not randomly assigned to homes with lead pipes, it is likely that there may be some differences between our treatment and control groups. To account for some of the observable differences, we re-estimate our event study models including the large set of controls described above (shown in the right panel of Figure 4). While the joint test of pre-trends for math is no longer statistically significant (p=0.14), a similar observational pattern emerges. However, because we would a priori expect negative effects of dangerous pipes on math achievement, this potential violation merely entails that our math

[34] Appendix Figure A5 presents the analogous scatterplots for four student mobility outcomes.

affects are biases to be *larger* in magnitude. Figure A6 of the appendix displays the event study plots from our two regression specifications with intermediate levels of additional controls.

When viewing these figures, it is important to keep in mind that the composition of the treatment and control groups are changing across time periods as students graduate, drop out, or (quite infrequently) leave the public school system. This is particularly important when considering the achievement outcomes because we only observe test scores for students in grades 3-8.

We now turn to Table 6, which displays results from four different specifications of difference-in-differences imputation models. In the baseline two-way fixed effects models, dangerous services lines have a -0.03 SD effect on math achievement, a -0.02 SD effect on reading achievement, a 0.5 percentage point effect on special needs status, and a 0.02 percentage point effect on daily attendance. In our most saturated models, with school-grade fixed effects and a variety of other controls, dangerous services lines have a -0.02 effect on math achievement, a -0.01 effect on reading achievement, a 0.26 percentage point effect on special needs status, and a 0.07 percentage point effect on daily attendance. None of the coefficients on our four educational outcomes are statistically significant at the 95% level in any of the four specifications.

In general, the small, statistically insignificant effects that we observe within-Flint stand in stark contrast to the dramatic decline in math achievement and increase in special needs status that we observe for all Flint students in our between-district analysis. Importantly, our difference-in-differences analyses are not powered to detect small effects. For example, using a 95% confidence interval around the estimates from our most saturated model, we can only rule out an increase special needs status greater than in 0.60 percentage points and decreases greater than 0.060 SD for math achievement, 0.063 SD for reading achievement, and 0.25% for attendance.

Our difference-in-differences results are robust to various alternative specifications. Table A8 displays results from classic OLS difference-in-differences models, which are quite comparable the imputation results. Table A9 shows results from a model that uses difference definitions of lead and copper service lines. Specifically, one might be concerned that our results may be sensitive to the way in which we defined various service line materials connoting lead risk. To address this concern, we construct two additional definitions for lead service lines and copper service lines, each more restrictive than the last; our results are not sensitive to differences in how we define lead treatment. Table A10 in the appendix shows difference-in-differences models split into three dimensions of heterogeneity: different grade cohorts (testing whether effects may vary by age at treatment), sex, and administrative district (whether a student attended Flint Community Schools); again, we come up with small, insignificant effects.

6. Discussion

In the first quasi-experimental study of a lead poisoning event due to lead plumbing in contemporary times, we find substantial negative effects of the Flint Water Crisis on the academic outcomes of children living in Flint. However, our analysis reveals a somewhat unexpected pattern of results. Comparing Flint to observationally equivalent school districts in Michigan, we find compelling evidence that the Flint Water Crisis reduced student math achievement and increased the proportion of students with special needs. When we look within Flint and compare children living in homes with lead service lines to their neighborhood peers living in homes with copper service lines, however, we do not find evidence for meaningful differences in academic outcomes.

What might explain the differing results in our between-district and within-Flint analyses? At first glance, it may seem a potential explanation may be that our within-Flint analysis suffers from a bias towards zero that our between-district analysis does not. For example, whereas it is easy to observe our treatment variable in the between-district analysis (where a child lives), it is more difficult to accurately measure treatment in the within-Flint analysis (how much lead a child was exposed to). The within-Flint treatment variable that we use—a child's home service line material—likely serves as a noisy proxy for their true risk of lead exposure during the Flint Water Crisis. Children consume a portion of their water outside the home (for example, at school and community organizations), and service lines are not the only source of exposure to lead in water systems (lead fixtures and lead solder also play a role). In addition, if children consume water at their neighbors' houses or if one child's educational outcomes affect the outcomes of their peers, treatment spillovers are potentially introduced. The existence of measurement error in our treatment variable and positive treatment spillovers could attenuate the estimated effects from our difference-in-difference analysis. Thus, it may well be that there were small effects of lead exposure during the Flint Water Crisis on academic outcomes among the children in our study that went largely undetected.

However, by focusing on specifically math achievement, we can see theoretically that the lead pathway is insufficient to explain the overall academic effects of the Flint Water Crisis estimated in our between-district analysis. Past findings help benchmark the anticipated impact on math achievement of children in Flint as a direct result of lead exposure. As we show in Section A1 of the appendix, we would expect only a .017 SD decrease in math scores as a result of lead effects observed during the Flint Water Crisis, almost an order of magnitude smaller than the .14 SD decrease in math achievement estimated using our synthetic control models. Moreover, because only 40% of children living in Flint were exposed to lead service lines at their residence, lead effects would become attenuated in the between-district analysis. If in fact within-Flint lead effects were driving the observed between-district effects, they would need to be substantially *larger* than .14 SD. Thus, both theoretically and empirically, lead explains little of the between-district results.

While the precise mechanisms of the observed effects remain largely unknown, they are unlikely to be driven primarily by lead poisoning. At first blush, the Flint Water Crisis may have appeared as predominantly an environmental health catastrophe, but we believe our results show that there is more to the story. In particular, the complex community-level psychosocial process that created a crisis surrounding the actual lead exposure played a substantial role in additional to whatever direct health effects lead did have. That is, we would expect the effects of *only* lead to be substantially smaller than the *overall* effects of the Flint Water Crisis. The Flint Water Crisis affected *all* children living in Flint, not just those children living in homes with lead service lines. While it is unclear why we see achievement effects in math but not reading, math scores have been shown to be more sensitive to other short-term mechanisms in middle childhood, such as summer learning loss (Alexander, Pitcock, and Boulay 2016).

What, then, may help explain the null results within-Flint? One possibility is that many Flint residents stopped drinking tap water at home immediately after the switch. Perhaps more likely is that children in houses lead pipes did consume high levels of lead in their water, but that this exposure was not with sufficiently high to cause measurable academic impairments in school-age children. Lead is more likely to affect the developing rather than the mature brain because the "blood-brain" barrier is less effective in young children compared to adults, and it is possible that school-age children are largely beyond their critical period (Lidsky and Schneider 2003). Previous quasi-experimental studies, Aizer et al. (2018) and Billings and Schnepel (2018), focus on children exposed to lead around 1-3 years of age. Because students are only tested in grades 3-8, the average age of exposure of children for whom we observe math and reading achievement is about 6 years old.

Should we expect the large psychosocial effects observed in Flint to generalize well to other lead-in-water crises? There are reasons to be cautious. Unlike other water crises, the change in Flint happened very discretely and was accompanied by a change in taste and discoloration, making it very noticeable to the city's populace. In addition, the magnitude of cover-up and ensuing scandal was large and received prolonged national attention. For this reason, we would expect people to substitute away from contaminated tap water at greater rates than other crises, which would decrease the lead effects. At the same time, Flint citizens were likely more aware of the crisis than people exposed to lead in tap water elsewhere in the United States, thereby heightening the social and psychological mechanisms.

Nonetheless, we believe that documenting Flint case is important in and of itself, and it also serves as an important demonstration that the ways in which such events may affect citizens are not necessarily straightforward; we cannot treat these crises as merely strictly medical phenomena. In addition, our results suggest that we currently substantially underestimate the costs of the Flint Water Crisis. Existing estimates, ranging from 50-400 million dollars (Muennig 2016; Zahran, Mcelmurry, and Sadler 2017), use only the lead effects, not the overall effects stemming from the psychosocial consequences as well. We hope this work draws attention to the huge potential costs of these water crises, and motivates preventative measures, which by comparison are cheap. Tragically, the Flint water switch was intended to save just 5-7 million dollars (about 2 million dollars each year from April 2014 until the completion of the Karegnondi Water Authority).

We must emphasize that our study has important limitations. First and foremost, we utilize school records that can only identify children living in Flint when they enter school. Therefore, we cannot speak to the educational consequences of the Flint Water Crisis on infants or young children. Infants or toddlers in homes with lead pipes were likely more substantially impacted than the school-age children we observe. It is possible, if not probable, that the groups most affected by lead in Flint had yet to enter school and went unobserved in our analysis, which suggests the importance of continued monitoring of and support for infants and toddlers at the time of the crisis. Another limitation is that we do not observe child blood lead levels directly.[35] Our analysis is also limited to specific educational outcomes captured in state administrative data. Hence, we are unable to detect impacts on a variety of potentially important behaviors and skills. While our parallel identification strategies help us to separate the direct health effects of lead from the psychosocial effects of the resulting crisis, we cannot say precisely what pathways any indirect effects may operate through. Finally, it is worth noting that the substantial public health response to the crisis likely mitigated potential negative effects, suggesting what we find in both the within and between Flint analyses underestimate the impact of the crisis alone. Local organizations in Flint, supported by state, federal and foundation funding, responded to the crisis by dramatically expanding the set of social, medical and educational services available to

[35] One way to explore these hypotheses would be to test the individual children living in these homes to get actual measures of the lead level in their blood. However, even if pre and post blood tests were available for all children in these homes, one might be concerned because lead is hard to measure in blood. This is because a single blood lead measure often does not capture true lead risk because lead exists in the blood only temporarily before it is deposited into the hair, bones, brain, and other organs.

children in Flint, ranging from positive messaging campaigns to free childcare to literacy programs (for example, the work of the recently formed Flint Pediatric Public Health Initiative). Flint is just one piece of a broader puzzle; researchers must continue exploring the ways in which lead plumbing underlays complex processes that produce health and social burden in the United States. As the risk of exposure from lead paint continues to fall with the success of many public health interventions, policymakers may do well to begin to shift their attention back towards understanding lead plumbing, and the community crises that it helps create, as an important health and social burden in America today.

References

Abadie, Alberto. 2020. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." *Journal of Economic Literature* Forthcomin.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490): 493–505. https://doi.org/10.1198/jasa.2009.ap08746.

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93 (1): 113–32. https://doi.org/10.1257/000282803321455188.

Aizer, Anna, Janet Currie, Peter Simon, and Patrick Vivier. 2018. "Do Low Levels of Blood Lead Reduce Children's Future Test Scores?" *American Economic Journal: Applied Economics* 10 (1): 307–41. https://doi.org/10.1257/app.20160404.

Alexander, Karl, Sarah Pitcock, and Matthew C Boulay. 2016. "The Summer Slide: What We Know and Can Do about Summer Learning Loss." *Teachers College Press*.

Amato, Michael S., Colleen F. Moore, Sheryl Magzamen, Pamela Imm, Jeffrey A. Havlena, Henry A. Anderson, and Marty S. Kanarek. 2012. "Lead Exposure and Educational Proficiency: Moderate Lead Exposure and Educational Proficiency on End-of-Grade Examinations." *Annals of Epidemiology*. https://doi.org/10.1016/j.annepidem.2012.07.004.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. 2018. "Synthetic Difference in Differences." http://arxiv.org/abs/1812.09970.

Beckley, Amber L., Avshalom Caspi, Jonathan Broadbent, Honalee Harrington, Renate M. Houts, Richie Poulton, Sandhya Ramrakha, Aaron Reuben, and Terrie E. Moffitt. 2018.

“Association of Childhood Blood Lead Levels with Criminal Offending.” *JAMA Pediatrics*. https://doi.org/10.1001/jamapediatrics.2017.4005.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. 2018. “The Augmented Synthetic Control Method,” no. November. http://arxiv.org/abs/1811.04170.

Billings, Stephen B., and Kevin T. Schnepel. 2018. “Life after Lead: Effects of Early Interventions for Children Exposed to Lead.” *American Economic Journal: Applied Economics*. https://doi.org/10.1257/app.20160056.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. “Revisiting Event Study Designs: Robust and Efficienct Estimation.” *Work in Progress*, 1–48.

Brooks, Samantha K., and Sonny S. Patel. 2020. “Psychological Consequences of the Flint Water Crisis : A Systematic Review.” *MedRxiv*, 1–31.

Callaway, Brantly, and Pedro H.C. Sant’Anna. 2020. “Difference-in-Differences with Multiple Time Periods.” *Journal of Econometrics*, no. xxxx: 1–31. https://doi.org/10.1016/j.jeconom.2020.12.001.

Canfield, Richard L., Mathew H. Gendle, and Deborah A. Cory-Slechta. 2004. “Impaired Neuropsychological Functioning in Lead-Exposed Children.” *Developmental Neuropsychology* 26 (1): 513–40. https://doi.org/10.1207/s15326942dn2601_8.

Canfield, Richard L., Charles R. Henderson, Deborah A. Cory-Slechta, Christopher Cox, Todd A. Jusko, and Bruce P. Lanphear. 2003. “Intellectual Impairment in Children with Blood Lead Concentrations below 10 Mg per Deciliter.” *New England Journal of Medicine* 348 (16): 1517–26. https://doi.org/10.1056/NEJMoa022848.

CDC. 2013. “Morbidity and Mortality Weekly Report: Blood Lead Levels in Children Aged 105 Year - United States 1999-2010.”

Cecil, Kim M., Christopher J. Brubaker, Caleb M. Adler, Kim N. Dietrich, Mekibib Altaye, John C. Egelhoff, Stephanie Wessel, et al. 2008. “Decreased Brain Volume in Adults with Childhood Lead Exposure.” *PLoS Medicine*. https://doi.org/10.1371/journal.pmed.0050112.

Chaisemartin, Clément de, and Xavier D’Haultfœuille. 2020. “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects.” *American Economic Review* 110 (9): 2964–96. https://doi.org/10.1257/aer.20181169.

Chernobyl Forum. 2006. “Chernobyl’s Legacy: Health, Environmental and Socio-Economic Impacts and Recommendations to the Governments of Belarus, the Russian Federation and Ukraine.” *International Atomic Energy Agency*.

Cornwell, David A., Richard A. Brown, and Steve H. Via. 2016. “National Survey of Lead Service Line Occurrence.” *Journal - American Water Works Association* 108 (4): E182–91. https://doi.org/10.5942/jawwa.2016.108.0086.

Doudchenko, Nikolay, and Guido W. Imbens. 2017. “Difference-In-Differences and Synthetic Control Methods : A Synthesis ∗.” *NBER Working Paper*.

EPA. 2016. “Proceedings Pursuant To Section 1431 of the Safe Drinking Water Act. In the Matter of: City of Flint, Michigan; Michigan Department of Enviormental Quality; and the State of Michigan.”

Feigenbaum, James J., and Christopher Muller. 2016. “Lead Exposure and Violent Crime in the Early Twentieth Century.” *Explorations in Economic History* 62: 51–86. https://doi.org/10.1016/j.eeh.2016.03.002.

Ferrie, Joseph P., Karen Rolf, and Werner Troesken. 2012. “Cognitive Disparities, Lead Plumbing, and Water Chemistry: Prior Exposure to Water-Borne Lead and Intelligence Test Scores among World War Two U.S. Army Enlistees.” *Economics and Human Biology*.

https://doi.org/10.1016/j.ehb.2011.09.003.

Gershenson, Seth, and Michael S. Hayes. 2018. "Police Shootings, Civic Unrest and Student Achievement: Evidence from Ferguson." *Journal of Economic Geography*. https://doi.org/10.1093/jeg/lbx014.

Gershenson, Seth, and Erdal Tekin. 2018. "The Effect of Community Traumatic Events on Student Achievement: Evidence from the Beltway Sniper Attacks." *Education Finance and Policy*. https://doi.org/10.1162/edfp_a_00234.

Gómez, Hernán F., Dominic A. Borgialli, Mahesh Sharman, Annika T. Weber, Anthony J. Scolpino, James M. Oleske, and John D. Bogden. 2019. "Blood Lead Levels in Females of Childbearing Age in Flint, Michigan, and the Water Crisis." *Obstetrics and Gynecology*. https://doi.org/10.1097/AOG.0000000000003416.

Gómez, Hernán F., and Kim Dietrich. 2018. "The Children of Flint Were Not 'Poisoned.'" *The New York Times*, 2018. https://www.nytimes.com/2018/07/22/opinion/flint-lead-poisoning-water.html.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*, no. xxxx. https://doi.org/10.1016/j.jeconom.2021.03.014.

Goosby, Bridget J., Jacob E. Cheadle, and Colter Mitchell. 2018. "Stress-Related Biosocial Mechanisms of Discrimination and African American Health Inequities." *Annual Review of Sociology*. https://doi.org/10.1146/annurev-soc-060116-053403.

Green, E. L. 2019. "Flint's Children Suffer in Class After Years of Drinking the Lead-Poisoned Water." *New York Times*, 2019. https://www.nytimes.com/2019/11/06/us/politics/flint-michigan-schools.html.

Grossman, Daniel S, and David J G Slusky. 2019. “The Impact of the Flint Water Crisis on Fertility,” no. December: 2005–31.

Jardim, Ekaterina, Mark Long, Robert Plotnick, Emma van Inwegen, Jacob Vigdor, and Hilary Wething. 2017. “Minimum Wage Increases, Wages, and Low-Wage Employment: Evidence from Seattle.” *National Bureau of Economic Research*. https://doi.org/10.3386/w23532.

Kaul, Ashok, Stefan Klößner, Gregor Pfeifer, and Manuel Schieler. 2018. “Synthetic Control Methods: Never Use All Pre-Intervention Outcomes Together with Covariates.” *Working Paper*, no. 83790: 1–20.

Kraft, Matthew A. 2020. “Interpreting Effect Sizes of Education Interventions.” *Educational Researcher*. https://doi.org/10.3102/0013189x20912798.

Laidlaw, Mark A.S., Gabriel M. Filippelli, Richard C. Sadler, Christopher R. Gonzales, Andrew S. Ball, and Howard W. Mielke. 2016. “Children’s Blood Lead Seasonality in Flint, Michigan (USA), and Soil-Sourced Lead Hazard Risks.” *International Journal of Environmental Research and Public Health*. https://doi.org/10.3390/ijerph13040358.

Lidsky, Theodore I., and Jay S. Schneider. 2003. “Lead Neurotoxicity in Children: Basic Mechanisms and Clinical Correlates.” *Brain*. https://doi.org/10.1093/brain/awg014.

Masten, Susan J., Simon H. Davies, and Shawn P. McElmurry. 2016. “Flint Water Crisis: What Happened and Why?” *Journal - American Water Works Association* 108 (12): 22–34. https://doi.org/10.5942/jawwa.2016.108.0195.

Muennig, Peter. 2016. “The Social Costs of Lead Poisonings.” *Health Affairs* 35 (8): 1545. https://doi.org/10.1377/hlthaff.2016.0661.

Neria, Yuval, Laura DiGrande, and Ben G. Adams. 2011. “Posttraumatic Stress Disorder

Following the September 11, 2001, Terrorist Attacks: A Review of the Literature Among Highly Exposed Populations." *American Psychologist* 66 (6): 429–46. https://doi.org/10.1037/a0024791.

Omoeva, Carina, Wael Moussa, Rachel Hatch, Friedrich Affolter, Bosun Jang, Anna Azaryeva-Valente, Daniel Kelly, Dominic Richardson, Hiroyuki Hattori, and Nikola Balvin. 2018. "The Effects of Armed Conflict on Educational Attainment and Inequality," no. 18.

Pfefferbaum, Betty, Sara Jo Nixon, Phebe M. Tucker, Rick D. Tivis, Vern L. Moore, Robin H. Gurwitch, Robert S. Pynoos, and Heather K. Geis. 1999. "Posttraumatic Stress Responses in Bereaved Children after the Oklahoma City Bombing." *Journal of the American Academy of Child and Adolescent Psychiatry* 38 (11): 1372–79. https://doi.org/10.1097/00004583-199911000-00011.

Pfefferbaum, Betty, Thomas W. Seale, Nicholas B. Mcdonald, Edward N. Brandt, Scott M. Rainwater, Brian T. Maynard, Barbara Meierhoefer, and Peteryne D. Miller. 2000. "Posttraumatic Stress Two Years after the Oklahoma City Bombing in Youths Geographically Distant from the Explosion." *Psychiatry* 63 (4): 358–70. https://doi.org/10.1080/00332747.2000.11024929.

Pieper, Kelsey J., Min Tang, and Marc A. Edwards. 2017. "Flint Water Crisis Caused by Interrupted Corrosion Control: Investigating 'Ground Zero' Home." *Environmental Science and Technology*. https://doi.org/10.1021/acs.est.6b04034.

Renner, Rebecca. 2010. "Exposure on Tap: Drinking Water as an Overlooked Source of Lead." *Environmental Health Perspectives*. https://doi.org/10.1289/ehp.118-a68.

Reuben, Aaron, Avshalom Caspi, Daniel W. Belsky, Jonathan Broadbent, Honalee Harrington, Karen Sugden, Renate M. Houts, Sandhya Ramrakha, Richie Poulton, and Terrie E. Moffitt.

2017. “Association of Childhood Blood Lead Levels with Cognitive Function and Socioeconomic Status at Age 38 Years and with IQ Change and Socioeconomic Mobility between Childhood and Adulthood.” *JAMA - Journal of the American Medical Association*. https://doi.org/10.1001/jama.2017.1712.

Reuben, Aaron, Maxwell L. Elliott, Wickliffe C. Abraham, Jonathan Broadbent, Renate M. Houts, David Ireland, Annchen R. Knodt, et al. 2020. “Association of Childhood Lead Exposure with MRI Measurements of Structural Brain Integrity in Midlife.” *JAMA - Journal of the American Medical Association*. https://doi.org/10.1001/jama.2020.19998.

Rossin-Slater, Maya, Molly Schnell, Hannes Schwandt, Sam Trejo, and Lindsey Uniat. 2020. “Local Exposure to School Shootings and Youth Antidepressant Use.” *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.2000804117.

Sharkey, Patrick. 2010. “The Acute Effect of Local Homicides on Children’s Cognitive Performance.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (26): 11733–38. https://doi.org/10.1073/pnas.1000690107.

Sorensen, Lucy C., Ashley M. Fox, Heyjie Jung, and Erika G. Martin. 2019. “Lead Exposure and Academic Achievement: Evidence from Childhood Lead Poisoning Prevention Efforts.” *Journal of Population Economics*. https://doi.org/10.1007/s00148-018-0707-y.

Spell, Annie W., Mary Lou Kelley, Jing Wang, Shannon Self-Brown, Karen L. Davidson, Angie Pellegrin, Jeannette L. Palcic, Kara Meyer, Valerie Paasch, and Audrey Baumeister. 2008. “The Moderating Effects of Maternal Psychopathology on Children’s Adjustment Post-Hurricane Katrina.” *Journal of Clinical Child and Adolescent Psychology* 37 (3): 553–63. https://doi.org/10.1080/15374410802148210.

Spencer, Steven J., Christine Logel, and Paul G. Davies. 2016. “Stereotype Threat.” *Annual Review of Psychology* 2016-Augus (5916): 32. https://doi.org/10.1146/annurev-psych-073115-103235.

Troesken, Werner. 2006. *The Great Lead Water Pipe Disaster*. Cambridge: MIT Press.

———. 2008. “Lead Water Pipes and Infant Mortality at the Turn of the Twentieth Century.” *Journal of Human Resources* 43 (3): 553–75. https://doi.org/10.1353/jhr.2008.0015.

Washerman, Gail A., Beatriz Staghezza-Jaramillo, Patrick Shrout, Dusan Popovac, and Joseph Graziano. 1998. “The Effect of Lead Exposure on Behavior Problems in Preschool Children.” *American Journal of Public Health*.

Winter, Alix S., and Robert J. Sampson. 2017. “From Lead Exposure in Early Childhood to Adolescent Health: A Chicago Birth Cohort.” *American Journal of Public Health* 107 (9): 1496–1501. https://doi.org/10.2105/AJPH.2017.303903.

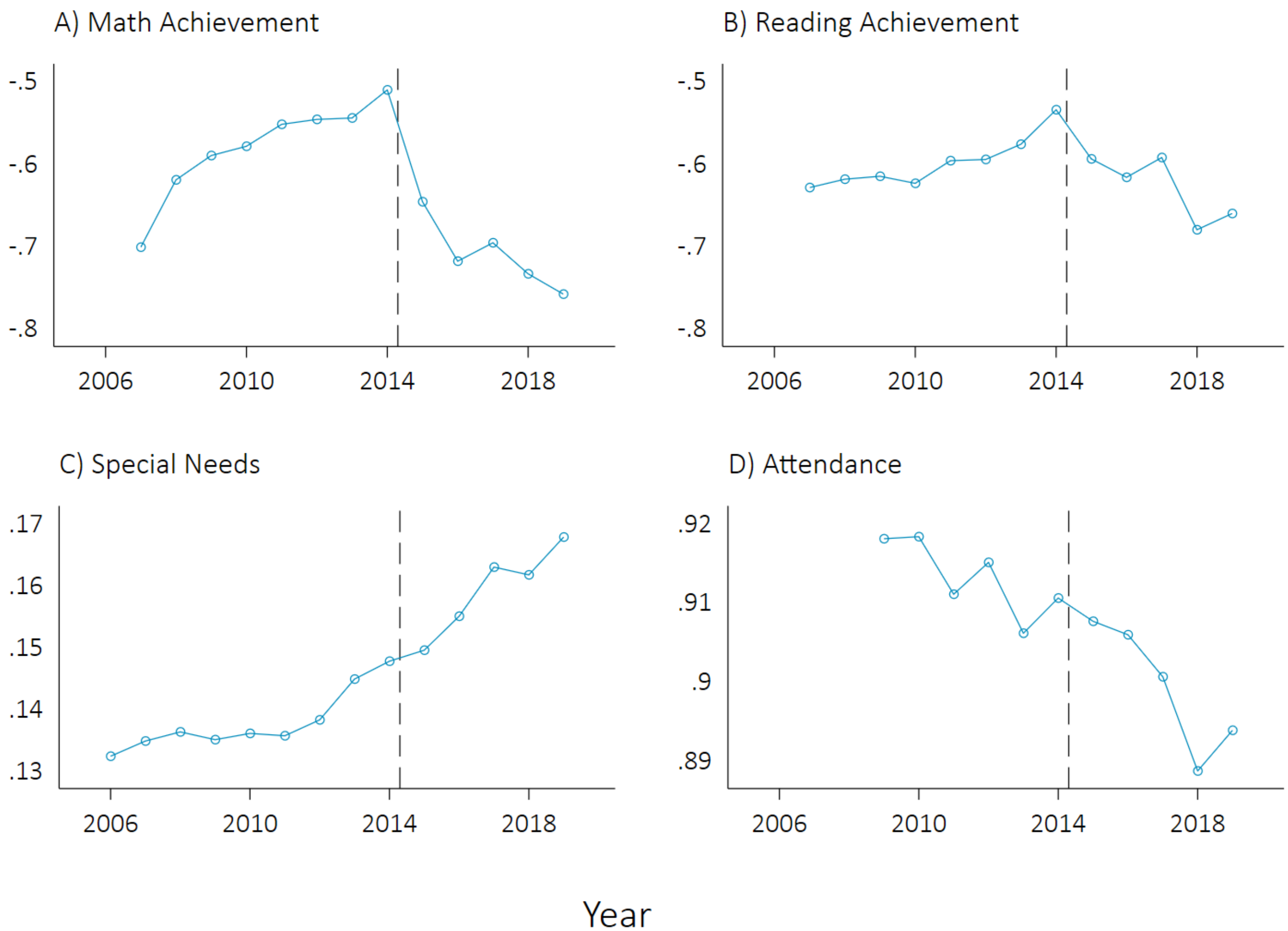
WNEM. 2015. “No Title,” 2015.

Zahran, Sammy, Shawn P. McElmurry, and Richard C. Sadler. 2017. “Four Phases of the Flint Water Crisis: Evidence from Blood Lead Levels in Children.” *Environmental Research*. https://doi.org/10.1016/j.envres.2017.05.028.

Zahran, Sammy, Shawn P Mcelmurry, and Richard C Sadler. 2017. “Four Phases of the Flint Water Crisis : Evidence from Blood Lead Levels in Children.” *Environmental Research* 157 (February): 160–72. https://doi.org/10.1016/j.envres.2017.05.028.

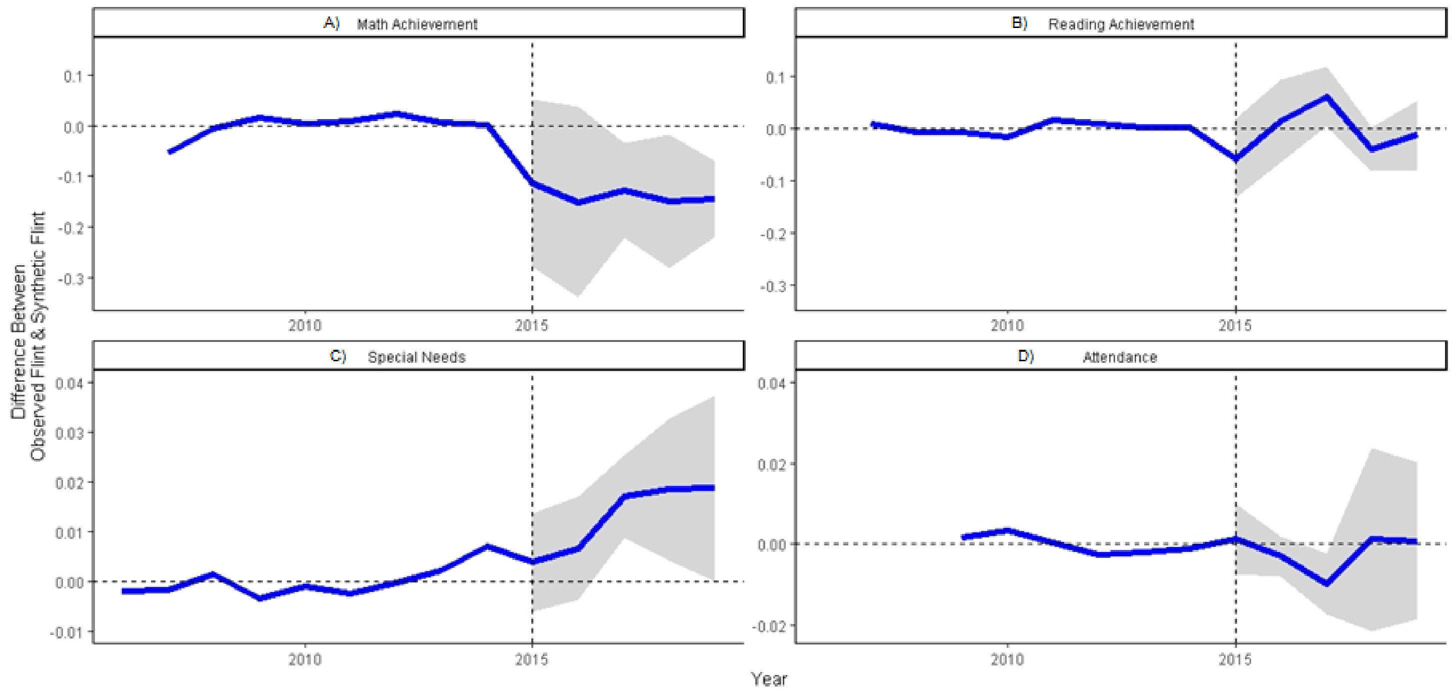
Zahran, Sammy, David Mushinski, Shawn P. McElmurry, and Christopher Keyes. 2020. “Water Lead Exposure Risk in Flint, Michigan after Switchback in Water Source: Implications for Lead Service Line Replacement Policy.” *Environmental Research*. https://doi.org/10.1016/j.envres.2019.108928.

Figure 1. Mean Educational Outcomes Over Time in Flint.



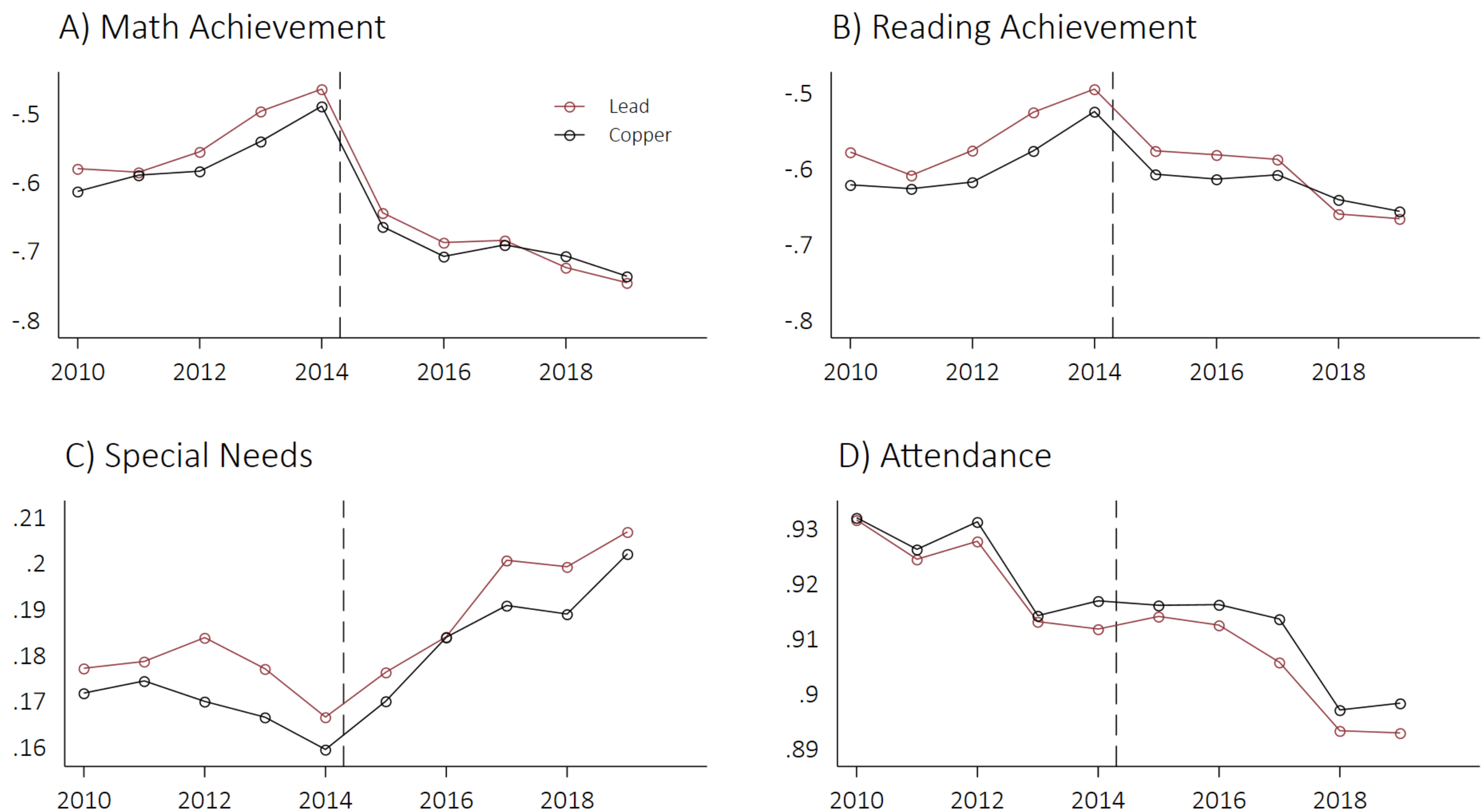
Note. This figure displays descriptive trends in the mean academic outcomes for the Flint geographic district from 2006-2019. Data is taken from the Michigan Department of Education's longitudinal administrative database. Grey dotted line represents time that the Flint Water Crisis begins. Math and reading achievement are observed for only for grades 3-8 and are standardized within test subject, grade, and year to the overall state distribution scores. Math and reading achievement observations begin in 2007. Both special needs and attendance are observed in grades K-12, and special needs status observations begin in 2006 whereas attendance observations begin in 2009.

Figure 2. Synthetic Control Estimates of the Effect of the Flint Water Crisis on Student Outcomes.



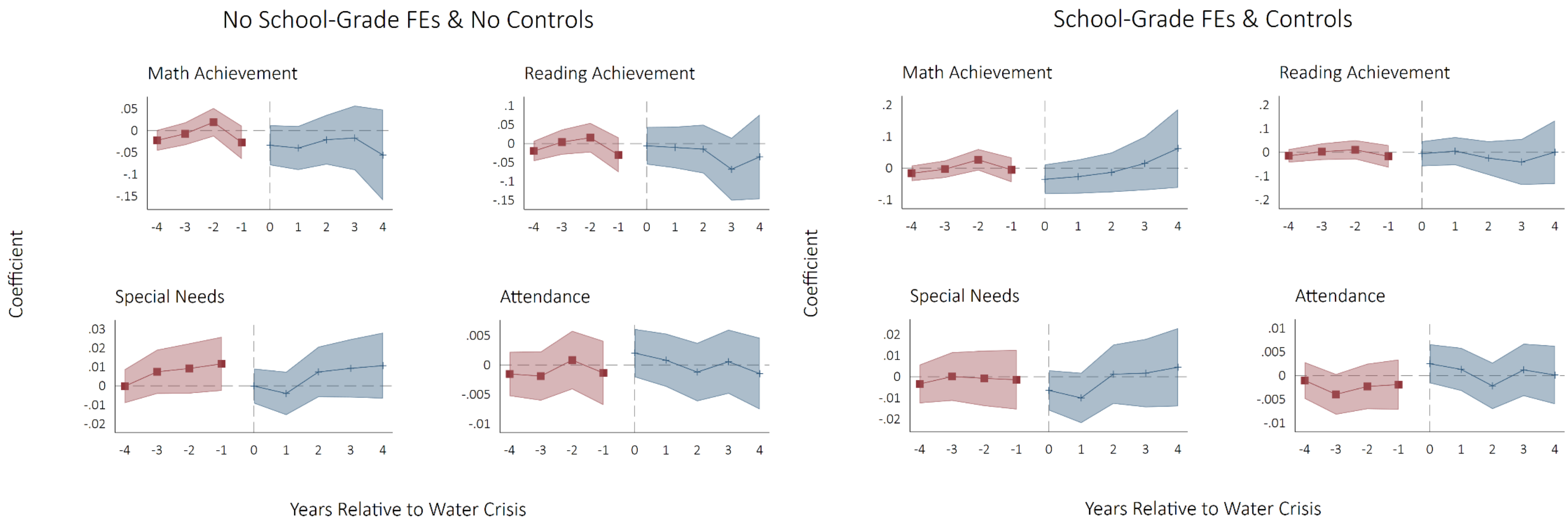
Note. ATT estimates of the causal effects of Flint Water Crisis on standardized math and reading academic achievement, fraction special needs, and fraction of school days attended are plotted over time. 2015 is the first year post-treatment. A treatment effect was estimated for each year in the post-period (2015-2019). The grayed shaded area represents the 95% confidence interval of the treatment effect estimates. Synthetic Flint was constructed by taking a weighted average of the 54 districts listed in table A1 in the appendix.

Figure 3. Mean Educational Outcomes Over Time in Flint by Service Line Material.



Note: This figure displays descriptive trends in the mean academic outcomes for the Flint geographic district from 2010-2019. Education data is taken from the Michigan Department of Education's longitudinal administrative data base. The black lines display students living in homes with copper service lines, while the red lines display students living in homes with lead service lines. Service line material data was collected during the City of Flint's service line inspection and replacement program that was implemented in the aftermath the Crisis. The grey dotted vertical line represents time that the Flint Water Crisis begins. Math and reading achievement are observed in only grades 3-8, whereas special needs and attendance are observed in grades K-12.

Figure 4. Event Study Imputation Estimates of Lead Pipes on Student Outcomes.



Note. This figure displays Dangerous Pipes * Year Dummy event study coefficients. Confidence intervals are shown in shaded red and blue areas. Dangerous Pipes * 2010, the first year in our Flint student-level panel, is the omitted category. The left four figures correspond to Models 1, 5, 9, and 13 and the right four figures correspond to models 4, 8, 12, and 16 from Table 6, respectively. Controls include race-gender-year fixed effects, a vector of interactions between a student's grade in 2013-2014 and year dummies, and a vector of interactions between a student's census block poverty in 2013-2014 and year dummies.

Table 1. Descriptive Statistics on Michigan Geographic Districts, 2013-2014.

	All Districts	All Large Districts	Potential Control Districts	Flint
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Math Achievement	-0.01	0.01	-0.34	-0.51
	(0.30)	(0.30)	(0.20)	(.)
Reading Achievement	0.03	0.04	-0.31	-0.53
	(0.25)	(0.25)	(0.18)	(.)
Fraction Special Needs	0.15	0.14	0.15	0.15
	(0.04)	(0.03)	(0.02)	(.)
Fraction School Days Attended	0.95	0.95	0.94	0.91
	(0.03)	(0.01)	(0.01)	(.)
Fraction Female	0.48	0.49	0.49	0.49
	(0.04)	(0.01)	(0.01)	(.)
Fraction Black	0.08	0.10	0.40	0.76
	(0.16)	(0.17)	(0.26)	(.)
Fraction Hispanic	0.06	0.06	0.12	0.04
	(0.08)	(0.08)	(0.14)	(.)
Fraction Economically Disadvantaged	0.51	0.49	0.75	0.89
	(0.18)	(0.19)	(0.11)	(.)
Fraction Limited English Proficiency	0.03	0.03	0.09	0.03
	(0.06)	(0.06)	(0.11)	(.)
Fraction Attending Charter Schools	0.04	0.04	0.13	0.31
	(0.09)	(0.06)	(0.11)	(.)
Fraction Attending Administrative District	0.79	0.83	0.69	0.45
	(0.18)	(0.11)	(0.14)	(.)
Enrollment	2,883	3947	7496	16210
	(6,083)	(7007)	(15922)	(.)
Number of Districts	548	362	54	1

Note. This table contains geographic school district characteristics from the Michigan Department of Education's longitudinal administrative data. Geographic school districts with greater than 1000 students are considered large. Math and reading achievement are standardized within test subject, grade, and year to the overall state distribution scores. Math and reading achievement are observed for grades 3-8; all other variables are observed for grades K-12. A full list of control districts is reported in Table A1. All variables are from the 2013-2014 academic year.

Table 2. Synthetic Control Weights & Effect Estimates.

Panel A. District Weights		
District Code	District Name	Weight
14020	Dowagiac Union School District	0.14
25240	Beecher Community School District	0.05
35040	Whittemore-Prescott Area Schools	0.07
63250	Oak Park School District	0.20
72020	Houghton Lake Community Schools	0.08
80090	Bloomingdale Public School District	0.01
81070	Lincoln Consolidated School District	0.14
82060	Hamtramck School District	0.13
82120	River Rouge School District	0.11
82430	Van Buren Public Schools	0.06
Sum		1.00

Panel B. Estimates		
	Pre-RMSE	Treatment Effect
Math Achievement	0.0388	-0.1373
		(0.0629)
Reading Achievement	0.0307	-0.0080
		(0.0187)
Special Needs	0.0025	0.0130
		(0.0047)
Attendance	0.0034	-0.0019
		(0.0046)
Overall (Scaled)	0.0689	

Note. This table displays results from a synthetic control model using a sample of geographic school districts taken from the Michigan Department of Education's longitudinal administrative data base. Math and reading achievement are standardized within test subject, grade, and year to the overall state distribution scores. Math and reading achievement are observed in only grades 3-8, whereas all other variables as observed in grades K-12. Only control districts given non-zero weights by the synthetic control model are listed in Panel A; a list of all control districts can be found in Table A2.

Table 3. Student Education Record to Home Service Line Data Match.

		Flint Students Matched to Service Line Data			
	All Flint students (n=17,024)	Matched (n = 10,245)	Not Matched (n = 6,779)	Difference	Difference (w/census block fixed effects)
	(1)	(2)	(3)	(4)	(5)
Student					
Math Achievement (z-score)	-0.51	-0.50	-0.52	0.023	-0.014
Reading Achievement (z-score)	-0.53	-0.51	-0.57	0.052*	0.031
Special Needs (%)	0.16	0.16	0.15	0.014*	0.022*
School Days Attended (%)	0.91	0.91	0.90	0.010**	0.009**
Female (%)	0.49	0.48	0.49	-0.006	0.003
Black (%)	0.76	0.74	0.79	-0.047**	-0.006
Hispanic (%)	0.04	0.05	0.03	0.012**	0.006
Economically Disadvantaged (%)	0.88	0.87	0.90	-0.030**	-0.025**
Limited English Proficiency (%)	0.03	0.03	0.02	0.017**	0.009*
Flint Community Schools (%)	0.45	0.48	0.41	0.061**	0.053**
School					
City (%)	0.57	0.59	0.54	0.055**	0.056**
Suburb (%)	0.38	0.36	0.42	-0.061**	-0.066**
Charter (%)	0.30	0.27	0.35	-0.085**	-0.076**
Magnet (%)	0.54	0.58	0.50	0.080**	0.068**
Enrollment (N)	598.41	600.33	595.51	3.902	-24.590*
Economically Disadvantaged (%)	0.80	0.80	0.81	-0.016**	-0.017**
Black (%)	0.67	0.66	0.68	-0.019*	-0.018**
Hispanic (%)	0.04	0.04	0.04	0.002*	0.003**
White (%)	0.24	0.25	0.23	0.015*	0.014*
First Year Teachers (%)	0.05	0.05	0.05	-0.003	-0.001
District per-pupil expenditures ($)	6410.22	6567.91	6172.36	379.687**	307.118**
Neighborhood					
Black (%)	0.57	0.56	0.60	-0.036**	
Hispanic (%)	0.03	0.03	0.04	-0.003	
Age 65+ (%)	0.12	0.12	0.11	0.010**	
Age 25+: BA (%)	0.10	0.11	0.10	0.009	
Age 25+: <HS Degree (%)	0.18	0.17	0.18	-0.014**	
Below Poverty Line (%)	0.42	0.40	0.45	-0.052**	
Unemployed (%)	0.28	0.27	0.29	-0.012	
Owner-occupied Houses (%)	0.43	0.46	0.39	0.067**	

Note. This table compares the full set of Flint students to the set whose addresses matched the service line data. To estimate the differences between the matched and unmatched sample, we used a series of regressions that adjusted the standard errors for clustering at the school-level and census block. For the student-level characteristics, we clustered the standard errors at the school level for all variables except percent attending Flint, another school district, charter, and living in Flint in 2014-15, for which we adjusted standard errors using census block. For school-level variables, we used adjusted standard errors for census block.

Table 4. The Determinants of Lead Pipes in the Homes of Flint Children.

	(1)	(2)	(3)
Housing characteristics			
Year built	-0.020***	-0.019***	-0.016***
	(0.001)	(0.001)	(0.000)
Housing condition 2014: Poor	0.050*	0.045*	0.002
	(0.026)	(0.026)	(0.015)
Housing condition 2014: Fair	-0.008	-0.015	-0.008
	(0.017)	(0.017)	(0.008)
Residential value ($10,000)	0.006**	0.007**	0.006***
	(0.003)	(0.003)	(0.001)
Land improvement flag	-0.033*	-0.019	-0.017
	(0.020)	(0.019)	(0.011)
Rental	0.004	-0.000	0.005
	(0.012)	(0.012)	(0.007)
Missing: Year built	-38.114***	-36.826***	-30.510***
	(1.477)	(1.559)	(0.475)
Neighborhood characteristics			
% Black		-0.001*	
		(0.000)	
% Hispanic		-0.003	
		(0.002)	
% Persons: 65 and over		0.001	
		(0.002)	
% Persons 25 years and over with bachelors		0.002	
		(0.001)	
% Persons 25 years and over with less than HS		-0.000	
		(0.001)	
% Economically Disadvantaged		0.002**	
		(0.001)	
% Unemployed		0.001	
		(0.001)	
% Owner-occupied		-0.002**	
		(0.001)	
Constant	38.817***	37.495***	31.132***
	(1.476)	(1.563)	(0.477)
Fixed effects for census blocks?	No	No	Yes

Notes: N=10,180
*p < 0.05, **p < 0.01, ***p < 0.001

Note. This table presents results from several regression models that include both household-level and neighborhood-level predictors.

Table 5. Comparing Flint Children with Lead versus Copper Pipes.

	All matched students (n = 10,245)	Copper Pipes (n = 6,183)	Lead Pipes (n = 4,062)	Difference	Difference (w/census block fixed effects)	Difference (w/ census block fixed effects and housing characteristics)
	(1)	(2)	(3)	(4)	(5)	(6)
Student						
Math Achievement (z-score)	-0.5	-0.51	-0.49	0.013	-0.029	-0.025
Reading Achievement (z-score)	-0.51	-0.53	-0.5	0.029	-0.034	-0.035
Special Needs (%)	0.16	0.16	0.17	0.007	0.009	0.01
School Days Attended (%)	0.91	0.92	0.91	-0.005	0.003	0.001
Female (%)	0.48	0.48	0.49	0.002	0.001	-0.033
Black (%)	0.74	0.77	0.69	-0.079**	-0.012	0.022
Hispanic (%)	0.05	0.04	0.06	0.017*	0.001	-0.002
Economically Disadvantaged (%)	0.87	0.86	0.89	0.031**	0.025	0.022
Limited English Proficiency (%)	0.03	0.03	0.05	0.019**	0.01	0.011
Attending school in Flint (%)	0.48	0.45	0.51	0.058**	0.039	0.044
Attending another school district (%)	0.25	0.27	0.23	-0.035**	-0.033	-0.038
School						
City (%)	0.59	0.56	0.63	0.072**	0.044	0.053*
Suburb (%)	0.36	0.38	0.32	-0.055**	-0.043	-0.051*
Charter (%)	0.27	0.28	0.25	-0.024*	-0.006	-0.007
Magnet (%)	0.58	0.56	0.6	0.035**	0.04	0.059*
Enrollment (N)	600.33	606.51	590.94	-15.746	-11.637	-3.789
Economically Disadvantaged (%)	0.8	0.79	0.8	0.008*	0.012*	0.011
Black (%)	0.66	0.66	0.66	0.002	0.017	0.028*
Hispanic (%)	0.04	0.04	0.04	0.002*	-0.001	-0.001
White (%)	0.25	0.25	0.25	-0.003	-0.015	-0.026*
FTE teachers in their first year of teaching (%)	0.05	0.05	0.04	-0.002	-0.002	-0.003
Per-pupil district-level total instructional expenditures ($)	6567.91	6492.08	6683.42	190.476**	75.048	65.572
Neighborhood						
Black (%)	0.56	0.6	0.5	-0.094**		
Hispanic (%)	0.03	0.03	0.04	0.005*		
Age 65 and over (%)	0.12	0.13	0.11	-0.020**		
Age 25 and over: BA (%)	0.11	0.11	0.1	-0.011*		
Age 25 and over: less than high school (%)	0.17	0.16	0.19	0.024**		
Economically Disadvantaged (%)	0.4	0.37	0.44	0.070**		
Unemployed (%)	0.27	0.27	0.29	0.019**		
Owner-occupied Houses (%)	0.46	0.48	0.42	-0.068**		

Note. $*p < 0.05$, $**p < 0.01$

Service line material is considered dangerous based on pipes made out of lead. Water is coded as dangerous if it contains >15 PPB of lead.

[a]The difference column was calculated by regressing dangerous status on each of the variables listed on the table with census blocks fixed effects.

For the service line materials, the overall F-test using all listed variables to predict dangerous status was significant, $F(12, 1663) = 1.49$, $p = 0.12$.

For the water material, the overall F-test using all listed variables to predict dangerous status was significant, $F(10, 1534) = 0.96$, $p = 0.487$.

Table 6. Difference in Difference Imputation Results: Dangerous Service Line.

	Math Achievement				Reading Achievement			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dangerous Pipes * Post	-0.0319	-0.0314	-0.0198	-0.0161	-0.0190	-0.0152	-0.0199	-0.0110
	(0.0218)	(0.0233)	(0.0212)	(0.0225)	(0.0250)	(0.0267)	(0.0246)	(0.0264)
Control Mean	{-.62}	{-.62}	{-.62}	{-.62}	{-.60}	{-.60}	{-.60}	{-.60}
Controls		X		X		X		X
School-Grade Fixed Effects			X	X			X	X
Pre-Trends P-Value	0.0219	0.107	0.0504	0.137	0.122	0.473	0.165	0.526
Students	8034	8018	7959	7944	8024	8008	7953	7938
Student-Years	33159	33084	32851	32781	33057	32982	32755	32685

	Special Needs				Attendance			
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Dangerous Pipes * Post	0.00398	-0.00553	-0.000318	-0.00255	0.000225	-0.00269	-0.000740	0.000700
	(0.00558)	(0.00587)	(0.00559)	(0.00581)	(0.00169)	(0.00172)	(0.00159)	(0.00166)
Control Mean	{.176}	{.176}	{.175}	{.174}	{.916}	{.917}	{.916}	{.916}
Controls		X		X		X		X
School-Grade Fixed Effects			X	X			X	X
Pre-Trends P-Value	0.365	0.753	0.862	0.796	0.640	0.412	0.183	0.433
Students	10245	10225	10238	10218	9921	9898	9866	9847
Student-Years	80818	80654	79642	79477	66797	66666	66208	66092

*$p < 0.05$, **$p < 0.01$

Note. This table displays results from difference-in-differences regressions of the effect of having dangerous pipe materials on a student's academic outcomes during the Flint Water Crisis using a panel spanning 2010-2019. All models include student fixed effects and year fixed effects. Education data is taken from the Michigan Department of Education's longitudinal administrative data base. Service line material data was collected during the City of Flint's service line inspection and replacement program that was implemented in the aftermath the crisis. Lead or galvanized steel service lines are classified as dangerous, whereas copper service lines are classified as not dangerous. Models with controls include race-gender-year fixed effects, a vector of interactions between a student's grade in 2013-2014 and year dummies, and a vector of interactions between a student's census block poverty in 2013-2014 and year dummies. These difference-in-differences models are estimated using imputation (Borusyak, Jaravel, and Spiess 2021

Appendix

The Psychosocial Effects of the Flint Water Crisis on School-Age Children

A1. The Expected Educational Impacts of Lead on Flint Children

How would we expect the amount of increased lead exposure that resulted from the Flint Water Crisis to impact educational outcomes? There are a few studies that provide recent and rigorous evidence studies regarding the effects of blood lead on academic achievement and intelligence (Aizer et al., 2018; Lanphear et al., 2005; Reuben et al., 2017). Aizer et al. (2018) uses an instrumental variables strategy, whereas the other two studies are simple linear regression (controlling for confounding factors such as parental cognition and childhood socioeconomic status). We omit Billings & Schnepel (2018) because they leverage lead reductions that results from a broad public health program that may have had its additional, non-lead driven effects on educational development.

In Aizer et al. (2018), the authors using two different instrumental variables strategies to quantify the causal effects of blood lead, measured in early childhood (i.e. below the age of 6), on third grade math and reading test scores using administrative data from Rhode Island. The first strategy uses one childhood measure of blood lead as an instrument for another from the same child, thereby reducing measurement error (Blood Lead IV). The second strategy uses the timing and geographic roll-out of a lead-free home certificate program as an instrument for a child's blood levels (Certificate IV). The second stage of both approaches estimates the effect of (predicted) blood lead on both math and reading achievement. All point estimates are negative and of the same order of magnitude but estimates from the Blood Lead IV are much more precise than estimates from the Certificate IV.

A second relevant study is Reuben et al. 2017. Here, the authors use rich longitudinal data from New Zealand to estimate the association between blood lead measured at age 11 on IQ in adulthood (age 38). They provide unconditional estimates as well as estimates from regression models that control for sex, maternal IQ, IQ at age 11, and children's socioeconomic status. The association between blood lead and IQ attenuate only slightly when controls are included. A third study, Lanphear et al. 2005, a presents a pooled international analysis that measures the association between blood lead and children's IQ. When we transform their results in order to be comparable to the other studies,[1] their estimates are not substantively different from Rueben et al. 2017.

In summary, the results from these three studies displayed in Table A1 of the appendix, converge surprisingly well; a one microgram per deciliter (ug/dL) increase in blood lead is associated with approximately 0.03 SD reduction in cognitive ability, math achievement, and reading achievement. Given that children experienced roughly a 0.5 ug/dL average increase in blood lead following the Flint Water Crisis, we would expect only a .017 SD decrease in math scores *as a result of lead effects*. In the education literature, a .017 SD effect is quite small

[1] Lanphear et al. 2005 find nonlinear effects of lead (i.e. the marginal blood lead effects are larger for children with less lead in their body than children with more lead in their body). For this reason, we focus on their estimates from the lower end of the lead distribution (<10 ug/dL), where most Flint children fell both before and after crisis. Using their preferred estimate, Lanphear observes a "decline of 6.2 IQ points (95% CI, 3.8-8.6) for an increase in blood lead levels from < 1 to 10 pg/d." Dividing 6.2 by 10 yields the effect of a one ug/dL change in blood lead on unstandardized IQ points per entails an average (0.62). Finally, if we divide .62 by the standard deviation of the IQ measure in their study (19.2), we are left with the effect of a one ug/dL change in lead on standardized IQ (.032)

(Kraft, 2020). Lead is also known to have effects on behavioral outcomes, which we can index through special needs status and daily attendance, though expected the expected effects are unknown.

A2. Synthetic Control Methods

A2.i. Notation

Suppose we have a panel of data with districts i over time t. Flint, our treated district, has $i = 1$. We observe a total of T time periods of some educational outcome Y_{it} for N unique school districts. The final pretreatment year is t^* such that $1 < t^* < T$. Here, $T = 14$ (i.e. we have yearly observations ranging from $2006 - 2019$), $t^* = 9$ (i.e. 2014), and $N = 55$ (i.e. Flint and 54 control districts).

We begin with our data structured as the following $N \times T$ matrix:

$$\begin{array}{c} Flint \\ Detroit \\ \vdots \\ Lansing \end{array} \begin{bmatrix} Y_{1,1} & Y_{1,2} & \cdots & Y_{1,t^*} & Y_{1,t^*+1} & \cdots & Y_{1,T} \\ Y_{2,1} & Y_{2,2} & \cdots & Y_{2,t^*} & Y_{2,t^*+1} & \cdots & Y_{2,T} \\ \vdots & & & & & & \vdots \\ Y_{N,1} & Y_{\mathrm{N},2} & \cdots & Y_{N,t^*} & Y_{N,t^*+1} & \cdots & Y_{N,T} \end{bmatrix}$$

(A2.i.a)

Y_{it}: Educational outcome for district i in time t
T: Scalar for total number of time periods observed
N: Total number of unique districts in our data
t^*: Final year t before treatment begins

To aid in the exposition of our methodology, for the remainder of the paper we partition pretreatment outcome observations (i.e. $t \leq t^*$) into X_{it} and use Y_{it} to refer solely to posttreatment observations (i.e. $t > t^*$).

$$\begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,t^*} & Y_{1,t^*+1} & \cdots & Y_{1T} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,t^*} & Y_{2,t^*+1} & \cdots & Y_{2T} \\ \vdots & & & & & & \vdots \\ X_{\mathrm{N},1} & X_{N,2} & \cdots & X_{N,t^*} & Y_{N,t^*+1} & \cdots & Y_{NT} \end{bmatrix}$$

(A2.i.b)

A2.ii. Potential Outcomes Framework

Using the potential outcomes framework (Rubin, 2005), in order to calculate the causal effect of the Flint Water Crisis on some educational outcome Y_{it}, we must identify the difference between the potential outcome that was observed in Flint, $Y_{1t}(1)$, and the potential outcome that we would have observed if Flint had not experienced its water crisis, $Y_{1t}(0)$.

$$ATT_t = Y_{1t}(1) - Y_{1t}(0)$$

$$ATT = \frac{1}{T - t^*} \sum_{t=1}^{T-t^*} ATT_t$$

(A2.ii.a)
ATT_t: Average treatment on the treated of the Flint Water Crisis in year t
ATT: Overall average treatment on the treated of the Flint Water Crisis
$Y_{1t}(1)$: Potential outcome for Flint (i.e i=1) in year t when treated
$Y_{1t}(0)$: Potential outcome for Flint (i.e i=1) in year t when untreated

Of course, we do not observe both $Y_{it}(1)$ and $Y_{it}(0)$ for any district in our panel. Instead, we observe $Y_{it}(1)$ for Flint and $Y_{it}(0)$ for all our control districts. Thus, we are left needing a reliable way to combine $Y_{it}(0)$ for the 54 control districts to approximate Flint's potential outcome in the absence of treatment, $Y_{1t}(0)$.

In many panel data settings, difference-in-differences methods are a credible way to obtain $\hat{Y}_{1t}(0)$, an estimate of $Y_{1t}(0)$, and thereby estimate the causal effects of an event like the Flint Water Crisis (though some challenges to estimating difference-in-differences with a single treated unit remain, see Ferman and Pinto 2019; Conley and Taber 2011).[2] However, finding a valid counterfactual for Flint is quite difficult. Even setting aside the water crisis, as a waning, mid-size city with residents who are disproportionately black and extremely poor, Flint is unique (see Figure 1). No districts in our panel readily support the common trends assumption required by differences-in-differences models.

A2.iii. Demeaned Synthetic Control Methods

To address the problem of that lack of a clear counterfactual district, we use synthetic control methods (Abadie 2021). The earliest variations iterations of this method were developed by Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmuller (2010). Synthetic control methodologies use pre-treatment data to identify the weighted average of control districts that most closely approximates the treated unit. They have the advantage of reducing bias through improved pretreatment fit and also allow for a more transparent counterfactual selection process, leaving less researcher degrees of freedom.

We implement a recent extension known as demeaned, or intercept-shift, synthetic control (Doudchenko & Imbens, 2017). Such a methodology involves first subtracting off each unit i's pretreatment outcome mean from all pre and posttreatment observations (conceptually similar to a unit fixed effect), and then fitting classic synthetic control on those residuals.

$$\bar{X}_i = \frac{1}{t^*}\sum_{t=1}^{t^*} X_{it}$$

$$X_{it} - \bar{X}_i = X_{it} - \frac{1}{t^*}\sum_{t=1}^{t^*} X_{it}$$

(A2.iii.a)
$\bar{X}_i$: Mean of outcome for unit i in the pretreatment period

[2] The inference strategy presented in Ferman and Pinto (2019) does not generalize to the case where a *weighted* set of control units approximates the treated unit.

We follow the literature (Arkhangelsky et al., 2018; Ben-Michael et al., 2021; Doudchenko & Imbens, 2017) and choose to balance only lagged outcomes, X_{it}, and not covariates. Covariates have been shown to be redundant in synthetic control when used alongside all lagged outcomes (Kaul et al., 2021). Our demeaned synthetic control estimator estimates $\hat{Y}_{1t}(0)$ using a demeaned weighted average of any number of control units. Thus, our estimating equation becomes:

$$ATT_t = Y_{1t}(1) - [\bar{X}_1 + \sum_{i=2}^{N} w_i(Y_{it} - \bar{X}_i))]$$

(A2.iii.b)

w_i: Synthetic control weight for district i

Demeaned synthetic control allows us to match on pretreatment trends rather than on pretreatment means. While Flint's mean academic outcomes are extreme outliers, its trends lie closer to the convex hull of control outcomes; thus, demeaning drastically increases our common support between Flint and the control units. At the same time, demeaning does not stretch our relatively short panel of pretreatment outcomes too thin, as would more complicated extensions, such as incorporating time weights, outcome modeling, or machine learning.

The demeaned synthetic control that we implement is identical to a few control extensions recently suggested in the literature. For example, our approach is equivalent to using augmented synthetic control methods in the specific case where each unit's pretreatment mean is used as covariate with a coefficient constrained to be equal to 1 (Ben-Michael et al., 2018). Our approach is also the identical to fitting synthetic difference-in-differences with uniform time weights and including a unit intercept (Arkhangelsky et al., 2018). As such, the demeaned synthetic control estimator we use can be expressed as a weighted difference-in-differences estimator.

$$ATT_t = [Y_{1t} - \frac{1}{t^*}\sum_{t=1}^{t^*}(\bar{X}_{1t})] - [\sum_{i=2}^{N}(w_i(Y_{it} - \bar{X}_i)$$

(A2.iii.c)

In the case of traditional, unweighted difference-in-differences, unit weights are uniform and sum to one (i.e. each district's w is equal to $\frac{1}{N-1}$). If we were to implement such weights in Equation 5c.iii, we are left with the familiar difference-in-differences estimator:

$$ATT_t = [Y_{1t} - \bar{X}_1] - \frac{1}{N-1}[\sum_{i=2}^{N}[Y_{it} - \bar{X}_i]$$

(A2.iii.d)

As a reminder, X_{it} represents the outcome variable Y_{it} in the pretreatment period.

A2.iv. Estimating Weights

We construct the vector of synthetic control weights, $\boldsymbol{w}$**,** by solving for the values that minimizes the squared distance between demeaned pretreatment outcomes for Flint and demeaned pretreatment outcomes for the other districts.

$$\boldsymbol{w} = \begin{bmatrix} w_2 \\ w_3 \\ \vdots \\ w_N \end{bmatrix}$$

(A2.iv.a)

$\boldsymbol{w}$: $N-1 \times 1$ vector of basic synthetic control weights w_i from $\mathrm{i} = 2, \dots, \mathrm{N}$

Specifically, we solve:

$$\min_{\boldsymbol{w}} \sum_{t=1}^{t*} [(X_{1t} - \bar{X}_1) - \sum_{i=2}^{N} (X_{it} - \bar{X}_i)\, w_i]$$

$$\text{subject to} \quad \sum_{i=2}^{N} w_i = 1$$

$$w_i \geq 0, i = 2, \dots, N$$

(A2.iv.b)

However, our problem in practice is slightly more complicated. We identify a single set of synthetic control weights that simultaneously balances all four of our educational outcomes in pretreatment period. Let us denote each outcome with j, where J is the total outcomes. For every outcome j, we have a unique $N \times T$ matrix described in Equation (5a.ii). Thus, we can solve for the values of $\boldsymbol{w}$ that simultaneously minimize the squared distance between all J demeaned pretreatment outcomes for Flint and all J demeaned pretreatment outcomes for control districts. We first normalize each outcome by dividing by its standard deviation so that differences in scaling do not influence the optimization.

$$\min_{\boldsymbol{w}} \sum_{j=1}^{J} \sum_{t=1}^{t*-1} [(X_{1tj} - \bar{X}_{1j}) - \sum_{i=2}^{N} \left(X_{itj} - \bar{X}_{ij}\right) w_i]$$

(A2.iv.c)

We calculate these weights, estimate treatment effects, and calculate standard errors using the publicly available R package augsynth.[3] The standard error estimation is described in more detail in Ben-Michael, Feller, and Rothstein 2019 and Arkhangelsky et al. 2019. In short, we use heteroskedasticity-consistent standard errors for panel data settings using the R package sandwich (Zeileis, 2004),with variance estimated via the jackknife (Miller, 1974). A desirable feature of these standard errors using the jackknife is that they incorporate uncertainty created by

[3] Available at https://github.com/ebenmichael/augsynth/blob/master/vignettes/augsynth-vignette.md.

the weight selection process in addition to uncertainty due to imperfect fit in the pre-treatment period.

In order to probe whether the treatments effects we observe are heterogeneous by some dichotomous subgroup v, we can decompose our treatment effect by maintaining the same weights as the main effects and estimating the treatment effects within subgroup. Unfortunately, there is no obvious way to yield standard errors for this decomposition. Nonetheless, we believe such an endeavor may still provide useful suggestive information regarding which underlying subgroups may be driving any observed treatment effects.

A3. Synthetic Control Robustness

Here we describe in more detail the ways in which we probe the robustness of our synthetic control results. Our robustness analyses center on We do so by altering our synthetic control models across three distinct dimensions: 1) the number of outcomes estimated simultaneously, 2) the sample of control districts used, and 3) the way in which district membership is defined (i.e. traditional vs. fixed). Tables A6 and A7 present the results of 21 synthetic control models. Across all the specifications, our key results—a large decrease in math achievement and a corresponding increase in special needs—remain unchanged.

First, we fit models that vary by the number of outcomes considered simultaneously. In addition to our preferred model results presented in this section, we fit six news versions of synthetic controls model: each outcome on its own, math achievement and reading achievement jointly, and special needs and attendance jointly. A number set of the models considering only a single outcome attain perfect fit in the pretreatment period. Perhaps counterintuitively, perfect fit is problematic for synthetic control methods; it implies that the models do not converge to a sufficiently unique solution and that we need either a longer pretreatment panel or more outcomes to balance. We therefor omit all results that attain perfect fit.

The second dimension that we vary is the set of control districts used to construct synthetic Flint. One concern might be that are results are sensitive to the specific pool of control districts used. Further, using districts too distant from Flint in covariate space might lead to an apples and orange comparison. We test the sensitivity of our results to this concern by restricting to a narrower set of districts that are more similar to Flint. As discussed earlier, relative to all Michigan districts with at least 1000 students, Flint is at the 99th percentile in terms of fraction black and in terms of fraction economically disadvantaged. Rather than using districts that are in the top decile in terms of fraction black or fraction economically disadvantaged, we use only districts that are in the top 5% in terms of at least one of these characteristics. This includes only 26 of the 54 districts included in our baseline comparison group. Panel B of Figure A3 in the appendix illustrates how these districts were identified. Table A2 provides a list of the 26 districts and Table A3 displays descriptive statistics. Models 8-14 in Tables A6 and A7 present synthetic control results using this more restricted set of control districts.

Finally, we test the robustness of our results to student residential mobility. The key concern is that the treatment might have led to compositional changes in Flint, which could bias our estimates. Because we are measuring things on the basis of geographic school district, a student who transfers to a school in a neighboring district will remain in our Flint sample, unless they change their residential address to outside Flint. To determine if any such movement influences our estimates, we re-estimate our SCM using a panel of data based on only students living in Flint or the comparison districts during the 2013-2014 school year. Using this "fixed sample" approach, a student that leaves Flint following the Water Crisis will nonetheless be contribute to the Flint average in these later years. This fixed sample approach ensures that the results we derive are not simply a function of selective attrition of Flint over time, as very few students attrit from the fixed sample (i.e., only those that switch to private school or move out of Michigan). Table A4 present descriptive statistics of this fixed panel alongside the traditional sample (i.e. our main specification). Models 15-21 in Tables A6 and A7 present synthetic control results using the fixed sample. Though there is some evidence for reading effects when using the fixed sample; see Models 17 and 18, given the mixed evidence we refrain from reading into them too deeply.

A4. Lead Water Test Data

For certain auxiliary analyses, we use data on residential water test results obtained from the state of Michigan (we refer to this data as "water data").[4] The water data contains results from lead tests using a voluntary homeowner-driven sampling program whereby concerned citizens were provided testing kits and conduct sampling on their own. Despite the lack of a statistical sampling strategy, voluntary testing turned out to largely representative of the Flint's housing stock (Goovaerts, 2017). We use data from water tests conducted in 2016, which consists of 17,421 addresses in the city of Flint. Because many addresses had multiple tests within 2016, we create a single binary variable that indicates whether the address had at least one water test return a value of lead concentration greater than one part per billion.

Next, we matched our household water testing data to our analytic sample of 17,024 students to any of the 17,421 unique addresses with valid water test data in 2016. To do so, University of Michigan research staff first parsed addresses to facilitate the algorithm responsible for the match using the ArcGIS Geocoding Parser.[5] From there, they attempted to match students to water test results from the water data using student home address. We were able to match to home water test results to 5,355 students, or 31.46% of our analytic sample 17,024. Both pipes and water data were matched to just 3,430 students. The comparatively fewer Flint addresses with water data largely explains why we were able to match a smaller proportion of students to water test results.

[4] Obtained at the following URL in January 2021: https://www.michigan.gov/flintwater/0,6092,7-345-76292_76294_76297---,00.html.

[5] Parsing involved splitting an address into various fields (e.g., address number, street name pre-directional such as north or south, street name, zip code). Roughly 80% of the original water material addresses that were fully parsed (n = 17,024). This step was not necessary for the service line materials addresses because they had been previously parsed using the Google Maps API by the researchers who collected the data.

References

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, *59*(2), 391-425. Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, *105*(490), 493–505. https://doi.org/10.1198/jasa.2009.ap08746

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque country. *American Economic Review*, *93*(1), 113–132. https://doi.org/10.1257/000282803321455188

Aizer, A., Currie, J., Simon, P., & Vivier, P. (2018). Do low levels of blood lead reduce children's future test scores? *American Economic Journal: Applied Economics*, *10*(1), 307–341. https://doi.org/10.1257/app.20160404

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2018). *Synthetic Difference in Differences*. http://arxiv.org/abs/1812.09970

Ben-Michael, E., Feller, A., & Rothstein, J. (2021). The augmented synthetic control method. *Journal of the American Statistical Association*, (just-accepted), 1-34.Billings, S. B., & Schnepel, K. T. (2018). Life after lead: Effects of early interventions for children exposed to lead. *American Economic Journal: Applied Economics*. https://doi.org/10.1257/app.20160056

Conley, T. G., & Taber, C. R. (2011). Inference with "difference in differences" with a small number of policy changes. *Review of Economics and Statistics*. https://doi.org/10.1162/REST_a_00049

Doudchenko, N., & Imbens, G. W. (2017). Difference-In-Differences and Synthetic Control Methods : A Synthesis ∗. *NBER Working Paper*.

Ferman, B., & Pinto, C. (2019). Inference in differences-in-differences with few treated groups and heteroskedasticity. *Review of Economics and Statistics*. https://doi.org/10.1162/rest_a_00759

Goovaerts, P. (2017). The drinking water contamination crisis in Flint: Modeling temporal trends of lead level since returning to Detroit water system. *Science of the Total Environment*. https://doi.org/10.1016/j.scitotenv.2016.09.207

Kaul, A., Klößner, S., Pfeifer, G., & Schieler, M. (2021). Standard Synthetic Control Methods: The Case Of Using All Pre-Intervention Outcomes Together With Covariates. *Journal of Business & Economic Statistics*, (just-accepted), 1-34.Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*. https://doi.org/10.3102/0013189x20912798

Lanphear, B. P., Hornung, R., Khoury, J., Yolton, K., Baghurst, P., Bellinger, D. C., Canfield, R. L., Dietrich, K. N., Bornschein, R., Greene, T., Rothenberg, S. J., Needleman, H. L., Schnaas, L., Wasserman, G., Graziano, J., & Roberts, R. (2005). Low-level environmental lead exposure and children's intellectual function: An international pooled analysis. *Environmental Health Perspectives*. https://doi.org/10.1289/ehp.7688

Miller, R. G. (1974). The Jackknife--A Review. *Biometrika*. https://doi.org/10.2307/2334280

Reuben, A., Caspi, A., Belsky, D. W., Broadbent, J., Harrington, H., Sugden, K., Houts, R. M., Ramrakha, S., Poulton, R., & Moffitt, T. E. (2017). Association of childhood blood lead levels with cognitive function and socioeconomic status at age 38 years and with IQ change and socioeconomic mobility between childhood and adulthood. *JAMA - Journal of the American Medical Association*. https://doi.org/10.1001/jama.2017.1712

Roth, J. (2021). Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends. *Working Paper*, 1–54.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, *100*(469), 322–331. https://doi.org/10.1198/016214504000001880

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v011.i10

Figure A1. Timeline of Flint Water Crisis.

December 2011: With Flint's municipal government bankrupt, Michigan governor Rick Snyder appoints the first of a long string of emergency city managers to sort out the city's budget troubles.

April 2014: To reduce costs, Flint stops using water from Lake Huron and instead begins obtaining its drinking water from the Flint River.

May 2014: Flint residents begin complaining about the odor, color, and taste of their tap water.

October 2014: The General Motors plant in Flint discontinues its use of the municipal water supply because its corrosivity is damaging car parts.

February 2015: Despite the continued presence of sediment and discoloration, a city of Flint consultant tells the public that the tap water is safe to drink. However, employees at the federal EPA inform the city of Flint and the Michigan Department of Environmental Quality that lead and other contaminants are leaching into the water system.

September 2015: Local Flint pediatrician and researcher Dr. Mona Hanna-Attisha announces that the number of children with elevated blood lead levels had doubled since the water switch. Virginia Tech engineering professor Dr. Mark Edwards finds that lead readings in the tap water of 5,000 Flint homes exceed safe levels.

October 2015: The state of Michigan detects an increase in child blood lead levels in Flint since the city's switch in routine testing required by the US CDC. Flint switches its municipal water back to the original Detroit source, though until a protective mineral film to redevelops inside the leaching pipes the city's water remains unsafe to drink.

November 2016: A class action lawsuit is filed against the city of Flint and the state of Michigan.

January 2016: Governor Rick Snyder declares a state of emergency. The top Flint health official publicly advises residents against using tap water. The National Guard is called into Flint to distribute bottled water.

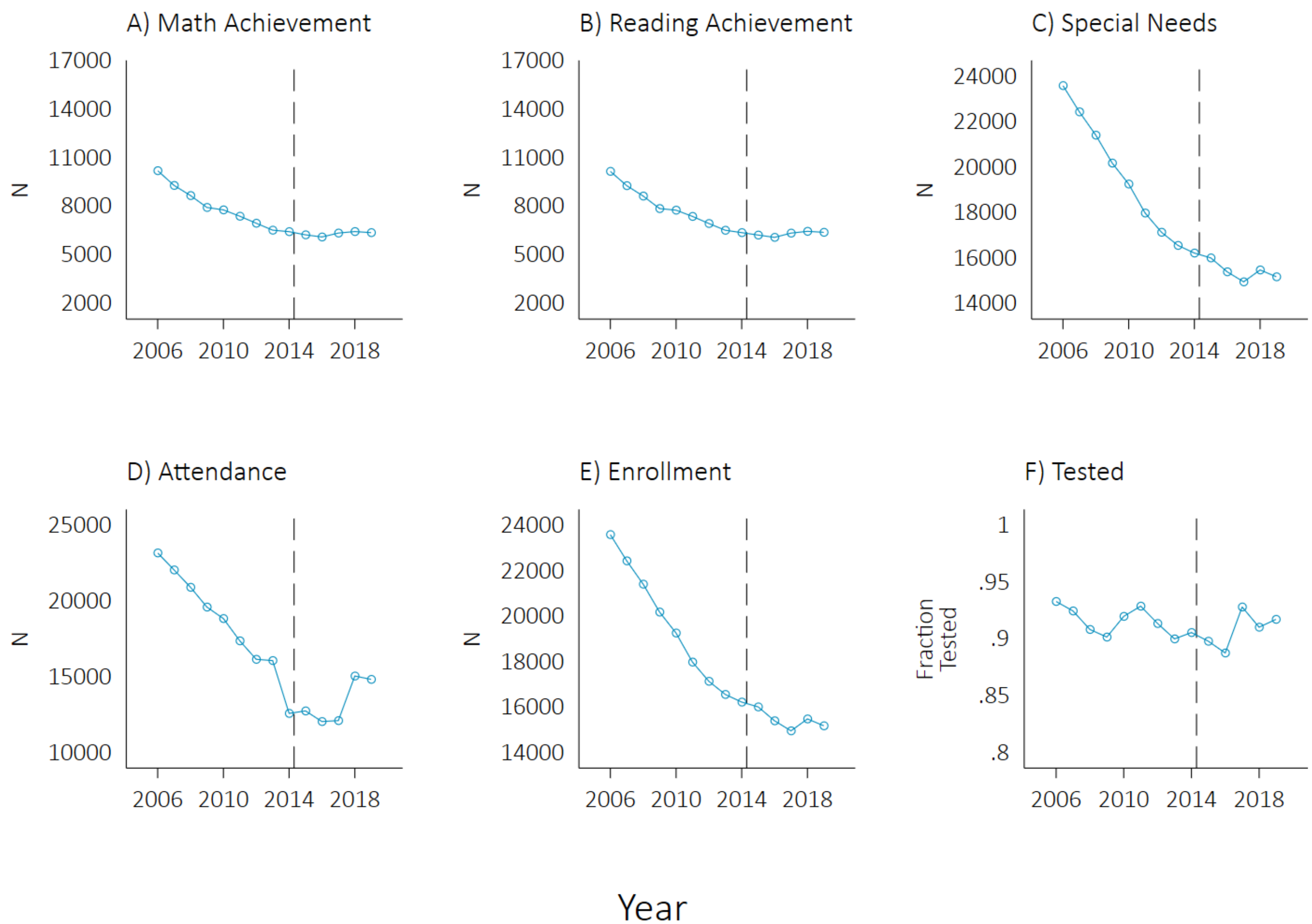
March 2016: The first lead service line is inspected and replaced as a part of the City's FAST Start program.

January 2019: Roughly 8,000 out of the 20,000 homes inspected as part of the FAST Start program have been found to have lead or galvanized service lines and were subsequently replace.

June 2019: EPA declares that Flint's water currently meets all health-based standards and is safe to drink, but then-Flint mayor Karen Weaver dismisses this conclusion as premature.

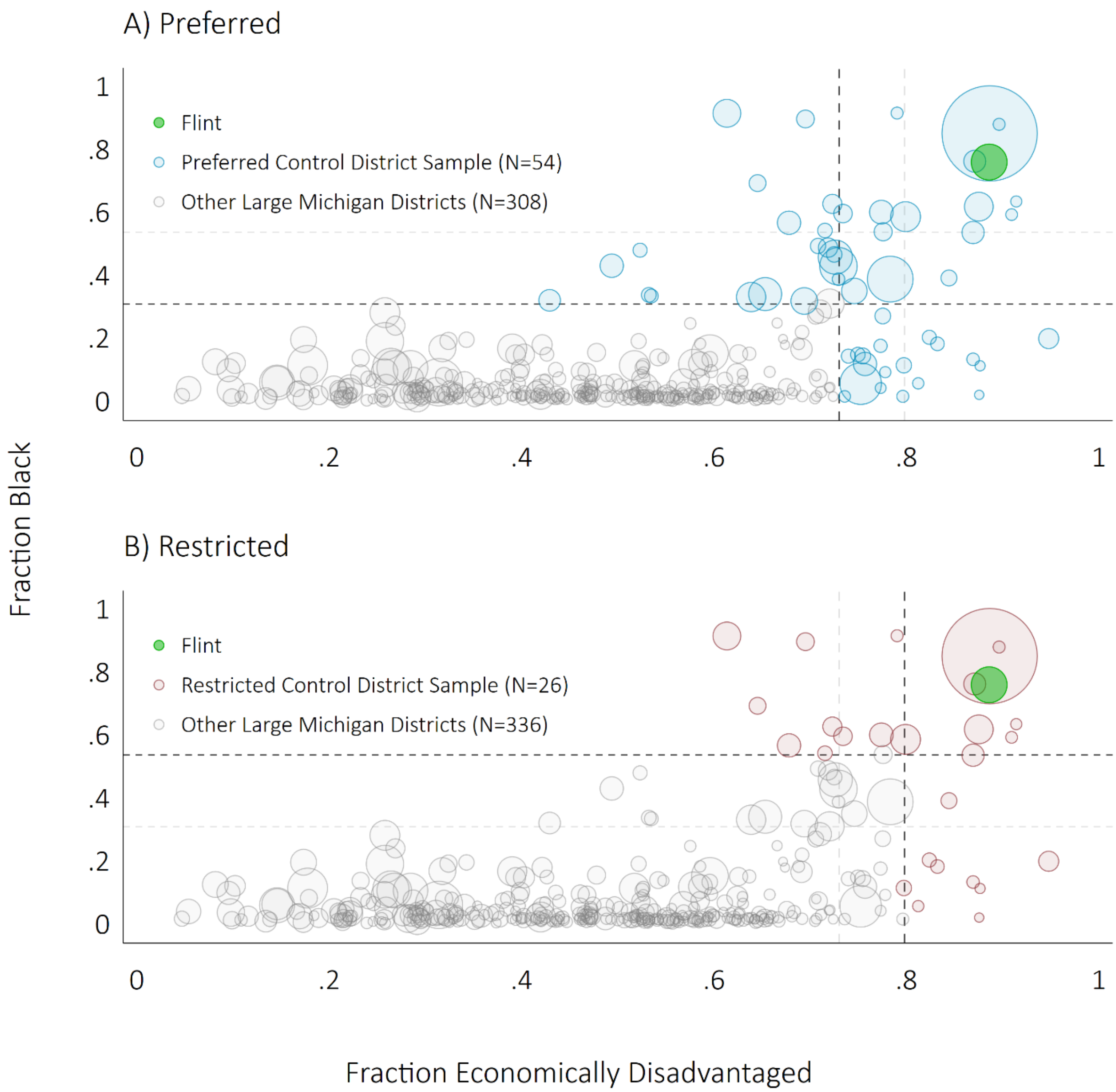
August 2020: The state of Michigan reaches a $600 million settlement with the victims of the Flint Water Crisis, with the bulk of this money going to those who were children at the time of the crisis.

Figure A2. Observations of Educational Outcomes Over Time in Flint.



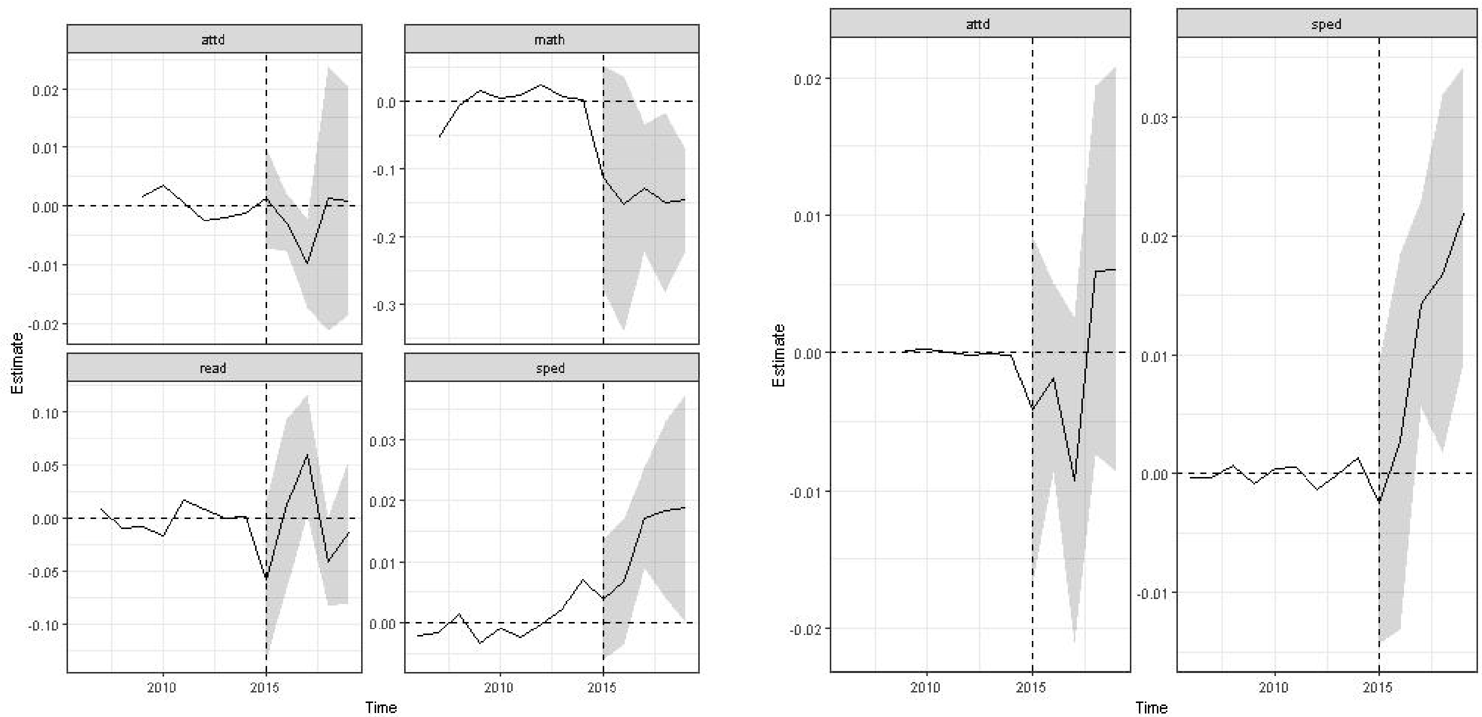
Note: This figure displays descriptive trends in the number of observations of academic outcomes for the Flint geographic district from 2006-2019. Data is taken from the Michigan Department of Education's longitudinal administrative data base. The grey dotted vertical line represents time that the Flint Water Crisis begins. Math and reading achievement are observed in only grades 3-8, whereas special needs and attendance are observed in grades K-12. Traditional sample is used for our main synthetic control analysis, while the fixed sample is use for a robustness check described in Section 5C.

Figure A3. Synthetic Control Sample Selection.



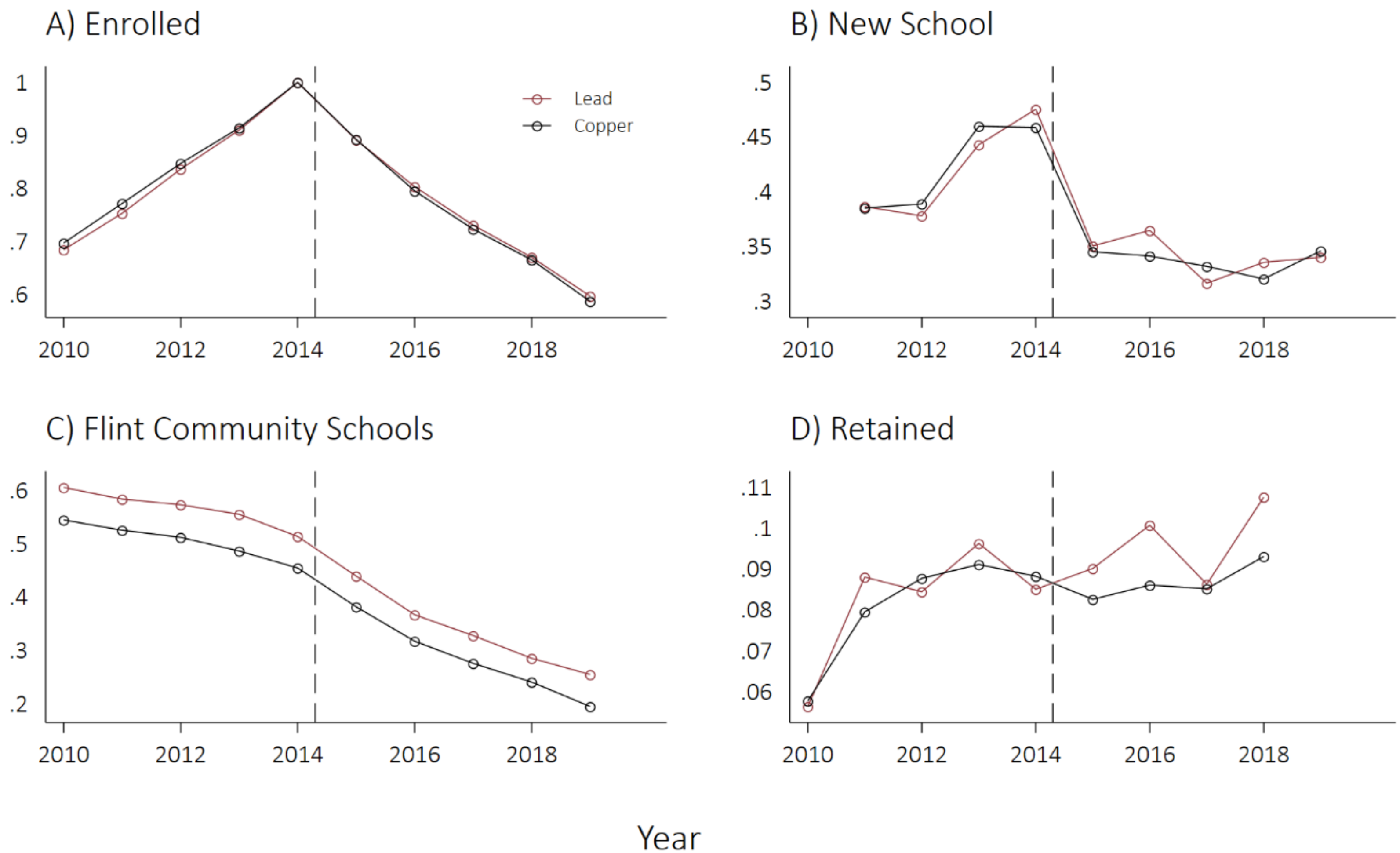
Note. Each circle represents a Michigan geographic school district. Data is taken from the Michigan Department of Education's longitudinal administrative data base. All districts with enrollment greater than 1,000 students are displayed, and the size of each school's circle is proportional to its student enrollment. Dashed lines in Panel A are at X=.73 and Y=.31, the 90th percentile of fraction economically disadvantaged and fraction black, respectively. Dashed lines in Panel B are at X=.8 and Y=.54, the 95th percentile of fraction economically disadvantaged and fraction black, respectively. All variables were measured in the 2013-2014 academic year

Figure A4. Synthetic Control Robustness: Special Needs Plot.



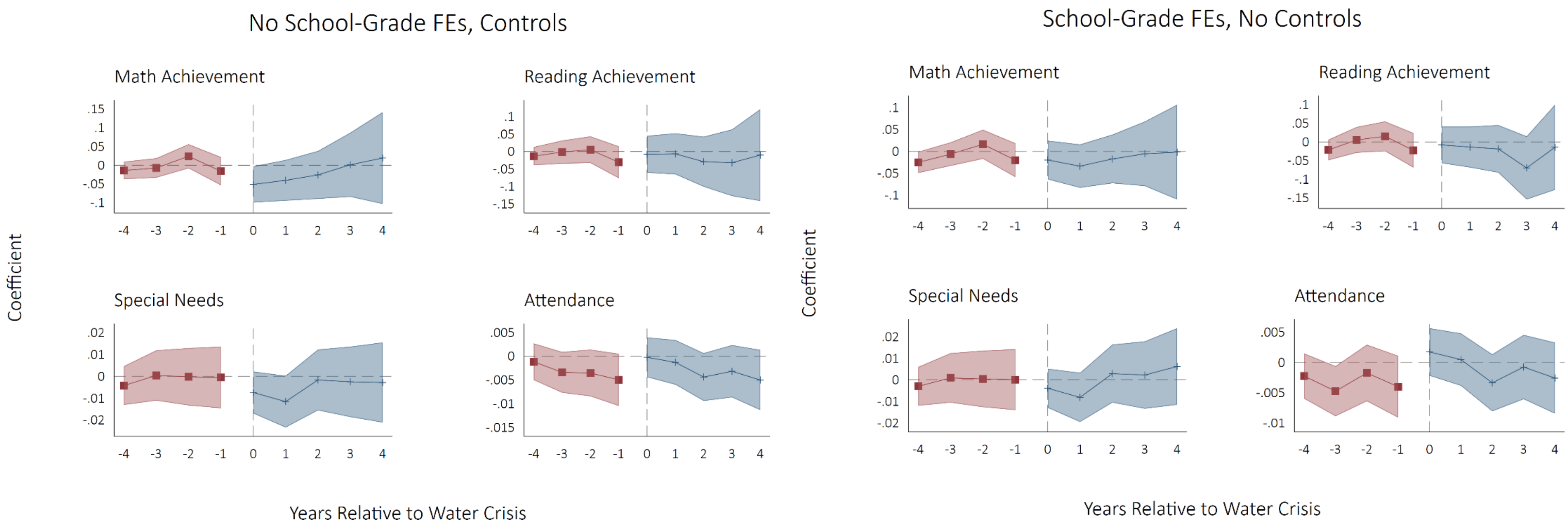
Note. This figure presents estimates of an alternative version of the synthetic control that consider each outcome individually (left-side of panel) as well as models that include two of the four outcomes together. The right-side of the figure presents results that includes attendance and special need services together, corresponding to models from Column 7 of Tables A6 and A7.

Figure A5. Mean Mobility Outcomes Over Time by Service Line Material.



Note. This figure displays descriptive trends in the means mobility outcomes for the Flint geographic district from 2006-2019 by service line material. Data is taken from the Michigan Department of Education's longitudinal administrative data base. Grey dotted line represents time that the Flint Water Crisis begins.

Figure A6. Event Study Imputation Estimates of Lead Pipes on Student Outcomes.



Note. This figure displays Dangerous Pipes * Year Dummy event study coefficients. Dangerous Pipes * 2010, the first year in our Flint student-level panel, is the omitted category. The left four figures correspond to Models 2, 6, 10, and 12 and the right four figures correspond to models 3, 7, 11, and 15 from Table 6, respectively. Controls include race-gender-year fixed effects, a vector of interactions between a student's grade in 2013-2014 and year dummies, and a vector of interactions between a student's census block poverty in 2014-2014 and year dummies.

Table A1. Previous Estimates of Lead on Achievement.

Aizer et al. 2018	Blood Lead IV		Certificate IV	
	Test Score	Test Score/SD	Test Score	Test Score/SD
Reading (Age 8)	-0.396	-0.030	-0.931	-0.072
	[0.0372]	[0.003]	[0.516]	[0.040]
Math (Age 8)	-0.266	-0.020	-0.431	-0.033
	[0.0347]	[0.003]	[0.471]	[0.036]

Rueben et al. 2017	Unadjusted		Covariate Adjusted*	
	Test Score	Test Score/SD	Test Score	Test Score/SD
Cig (Age 38)	-0.394	-0.026	-0.322	-0.021
	[0.140]	[0.009]	[0.089]	[0.006]

*Covariates include sex, maternal IQ, childhood IQ (age 11), and childhood socioeconomic status.

Lanphear et al. 2005			Covariate Adjusted**	
			Test Score	Test Score/SD
IQ (Age 7)			-0.620	-0.032
			[0.133]	[0.007]

**Covariates include birth weight, maternal IQ, and maternal education.

Note: This table displays previous estimates of the effect lead on achievement taken from three studies (Aizer et al. 2018; Reuben et al. 2017; Lanphear et al. 2005).

Table A2. List of Control Districts.

District Code	District Name	Full Control Group (n = 54)	Restricted Control Group (n = 26)
13020	Battle Creek Public Schools	X	
82160	Wayne-Westland Community School District	X	
80020	Bangor Public Schools (Van Buren)	X	
25240	Beecher Community School District	X	X
25060	Bendle Public Schools	X	X
11010	Benton Harbor Area Schools	X	X
80090	Bloomingdale Public School District	X	X
73180	Bridgeport-Spaulding Community School District	X	
25080	Carman-Ainsworth Community Schools	X	
50070	Clintondale Community Schools	X	
82030	Dearborn City School District	X	
82040	Dearborn Heights School District #7	X	
82240	Westwood Community School District	X	X
82010	Detroit City School District	X	X
14020	Dowagiac Union School District	X	
50020	East Detroit Public Schools	X	X
82250	Ecorse Public Schools	X	X
63020	Ferndale Public Schools	X	
50090	Fitzgerald Public Schools	X	X
41120	Godfrey-Lee Public Schools	X	X
41020	Godwin Heights Public Schools	X	X
41010	Grand Rapids Public Schools	X	
82060	Hamtramck School District	X	X
82320	Harper Woods The School District	X	X
18060	Harrison Community Schools	X	
80120	Hartford Public Schools	X	
63130	Hazel Park School District	X	
72020	Houghton Lake Community Schools	X	
38170	Jackson Public Schools	X	
39010	Kalamazoo Public Schools	X	
41140	Kelloggsville Public Schools	X	
41160	Kentwood Public Schools	X	
33020	Lansing Public School District	X	
81070	Lincoln Consolidated School District	X	
82090	Lincoln Park School District	X	
82045	Melvindale-North Allen Park Schools	X	X
50160	Mount Clemens Community School District	X	
25040	Mt. Morris Consolidated Schools	X	X
61010	Muskegon Public Schools	X	X
63250	Oak Park School District	X	X
61190	Orchard View Schools	X	
63030	Pontiac City School District	X	X
82110	Redford Union Schools District No. 1	X	X
82120	River Rouge School District	X	X
82130	Romulus Community Schools	X	X
73010	Saginaw School District	X	X
50200	South Lake Schools	X	
82140	South Redford School District	X	X
63060	Southfield Public School District	X	X
82430	Van Buren Public Schools	X	
50220	Van Dyke Public Schools	X	
33215	Waverly Community Schools	X	
35040	Whittemore-Prescott Area Schools	X	X
81020	Ypsilanti Community Schools	X	X

Note. The control samples were selected from the distribution of all Michigan geographic districts with enrollment greater than 1,000 students during the 2013-2014 academic year. Data is taken from the Michigan Department of Education's longitudinal administrative data base. The full control sample includes districts in the top 10% of either fraction black or fraction economically disadvantaged. The restricted sample includes districts in the top 5% of either fraction black or fraction economically disadvantaged.

Table A3. Descriptive Statistics on Restricted Control Sample.

	Full Control Group Mean (SD)	Restricted Control Group Mean (SD)	Flint Mean (SD)
Math Achievement (SD)	-0.34	-0.46	-0.51
	(0.20)	(0.16)	(.)
Reading Achievement (SD)	-0.31	-0.41	-0.53
	(0.18)	(0.16)	(.)
Fraction Special Needs	0.15	0.16	0.15
	(0.02)	(0.02)	(.)
Fraction School Days Attended	0.94	0.93	0.91
	(0.01)	(0.02)	(.)
Fraction Female	0.49	0.48	0.49
	(0.01)	(0.01)	(.)
Fraction Black	0.40	0.51	0.76
	(0.26)	(0.29)	(.)
Fraction Hispanic	0.12	0.12	0.04
	(0.14)	(0.16)	(.)
Fraction Economically Disadvantaged	0.75	0.81	0.89
	(0.11)	(0.09)	(.)
Fraction Limited English Proficiency	0.09	0.10	0.03
	(0.11)	(0.12)	(.)
Fraction Attending Charter Schools	0.13	0.16	0.31
	(0.11)	(0.12)	(.)
Fraction Attending Administrative District	0.69	0.65	0.45
	(0.14)	(0.14)	(.)
Enrollment	7496	8457	16210
	(15922)	(22096)	(.)
Number of Districts	**54**	**26**	**1**

Note. This table contains geographic school district characteristics from the Michigan Department of Education's longitudinal administrative data base. Math and reading achievement are standardized within test subject, grade, and year to the overall state distribution scores. Math and reading achievement are observed in only grades 3-8, whereas all other variables as observed in grades K-12. A list of control districts in the full and restricted samples can be found in Table A1. All variables are from the 2013-2014 academic year.

Table A4. Descriptive Statistics on Fixed Michigan District Panel.

	2010		2014	2018	
Traditional Sample	X		X	X	
Fixed Sample		X	X		X
Math Achievement	-0.58	-0.59	-0.51	-0.73	-0.73
Reading Achievement	-0.62	-0.64	-0.53	-0.68	-0.68
Fraction Special Needs	0.14	0.16	0.15	0.16	0.18
Fraction School Days Attended	0.92	0.93	0.91	0.89	0.89
Fraction Female	0.49	0.49	0.49	0.48	0.48
Fraction Black	0.77	0.78	0.76	0.69	0.69
Fraction Hispanic	0.04	0.04	0.04	0.05	0.05
Fraction Economically Disadvantaged	0.77	0.87	0.89	0.92	0.90
Fraction Limited English Proficiency	0.02	0.02	0.03	0.04	0.04
Fraction Attending Charter Schools	0.25	0.27	0.31	0.37	0.27
Fraction Attending Administrative District	0.66	0.54	0.45	0.31	0.25
Number of Observations	19,254	10,141	16,210	15,466	11,213

Note. This table contains geographic school district characteristics from the Michigan Department of Education's longitudinal administrative data. Math and reading achievement are standardized within test subject, grade, and year to the overall state distribution scores. Math and reading achievement are observed in only grades 3-8, whereas all other variables as observed in grades K-12. The Traditional Sample defines a student's geographic district using a student's home address *in that year*, whereas the Fixed Sample defines a student's geographic district using a student's home address during the 2013-2014 school year (final pre-treatment period). Thus, the Traditional and Fixed samples for 2013-2014 are mechanically identical.

Table A5. Synthetic Control Heterogeneity Estimates.

	Gender (1)			Grade (2)		Administrative District (3)	
	All	Boys	Girls	Grades 3-5	Grades 6-8	Flint Community Schools	Other District
Math Achievement	-0.14	-0.15	-0.13	-0.18	-0.10	-0.30	-0.20
Reading Achievement	-0.01	-0.01	0.00	-0.03	0.02	-0.13	-0.10
Special Needs	0.01	0.02	0.01	0.01	0.01	0.03	0.01
Attendance	0.00	0.00	0.00	0.00	0.00	-0.01	0.00

This table decomposes our main synthetic control effect estimates by dichotomous subgroup (using the same set of identified weights).

Table A6. Synthetic. Control Robustness: Estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
Fixed Sample															X	X	X	X	X	X	X
Traditional Sample	X	X	X	X	X	X	X	X	X	X	X	X	X	X							
Full Control Group	X	X	X	X	X	X	X								X	X	X	X	X	X	X
Restricted Control Group								X	X	X	X	X	X	X							
Math Achievement		-0.1618	-0.1247	-0.1373					-0.1376	-0.1195	-0.0909					—	-0.1221	-0.1262			
		(0.0511)	(0.0216)	(0.0629)					(0.0611)	(0.0143)	(0.0115)					—	(0.0577)	(0.0672)			
Reading Achievement			-0.0287	-0.0080	—					-0.0021	-0.0151	0.0228					-0.0919	-0.0620	—		
			(0.0703)	(0.0187)	—					(0.0429)	(0.0316)	(0.0455)					(0.0535)	(0.0388)	—		
Special Needs				0.0130		—	0.0106				0.0217		—	0.0113				0.0184		—	0.0155
				(0.0047)		—	(0.0057)				(0.0053)		—	(0.0053)				(0.0054)		—	(0.0055)
Attendance	—			-0.0019			-0.0006	—			0.0008			0.0009	0.0019			-0.0016			0.0020
	—			(0.0046)			(0.0039)	—			(0.0035)			(0.0055)	(0.0026)			(0.0013)			(0.0023)
Pre-RMSE	0.000	0.003	0.031	0.069	0.000	0.000	0.002	0.000	0.031	0.040	0.097	0.010	0.000	0.005	0.001	0.000	0.026	0.058	0.000	0.000	0.008

Note. This table displays effect estimates from a synthetic control models using a sample of geographic school districts taken from the Michigan Department of Education's longitudinal administrative data base. Estimates from models that achieve perfect fit are not displayed.

Table A7. Synthetic Control Robustness: Weights.

District Code	(2)	(3)	(4)	(7)	(9)	(10)	(11)	(12)	(14)	(15)	(17)	(18)	(21)
11010	0	0	0	0	0	0	0	0	0	0.2	0	0.12	0.09
14020	0.14	0	0.14	0						0	0.01	0	0
25040	0	0	0	0.02	0	0	0.05	0.09	0.15	0	0	0	0
25240	0	0.19	0.05	0	0.29	0.28	0.2	0.2	0	0	0.27	0.28	0
33215	0	0	0	0						0	0.01	0.17	0.19
35040	0	0	0.07	0.04	0.09	0	0	0	0.02	0	0	0	0.05
41160	0	0	0	0.17						0.28	0	0	0.07
50020	0	0	0	0.15	0	0	0	0	0	0	0	0	0
50090	0	0	0	0.04	0	0	0	0	0.22	0	0	0	0
50220	0	0	0	0						0	0.26	0	0
61010	0	0.28	0	0	0.43	0.5	0.09	0.34	0	0.23	0	0.16	0.1
63020	0.12	0.1	0	0						0	0	0	0
63130	0	0	0	0						0	0.1	0	0
63250	0	0	0.2	0.27	0	0	0.21	0.13	0.19	0.14	0	0.07	0.16
81070	0	0	0.14	0						0	0	0.08	0
82060	0.18	0.14	0.13	0	0.2	0.15	0.17	0.05	0	0	0.22	0.12	0
82120	0	0.1	0.11	0	0	0.07	0.16	0.18	0.14	0.15	0	0	0.24
82140	0	0	0	0.13	0	0	0	0	0.16	0	0	0	0
82430	0.38	0.11	0.06	0						0	0.02	0	0

Note. This table displays weights from a synthetic control models using a sample of geographic school districts taken from the Michigan Department of Education's longitudinal administrative data base. Only control districts with a weight in any model greater than .1 are displayed; a list of all control districts can be found in Table A1. Weights from models that achieve perfect fit are not displayed.

Table A8. Traditional Difference in Difference Results: Dangerous Service Line.

	Math Achievement		Reading Achievement		Special Needs		Attendance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dangerous Pipes * Post	-0.0387	-0.0262	-0.00875	-0.00168	-0.00537	-0.00382	-0.00369*	-0.000536
	(0.0199)	(0.0191)	(0.0225)	(0.0220)	(0.00540)	(0.00526)	(0.00174)	(0.00160)
Control Mean	{-.62}	{-.62}	{-.60}	{-.60}	{.176}	{.175}	{.917}	{.917}
School-Grade Fixed Effect		X		X		X		X
Students	8219	8094	8208	8073	10180	10165	10001	9931
Student-Years	35933	35104	35837	34992	80753	79158	67438	66118

*$p < 0.05$, **$p < 0.01$

Note. This table displays results from OLS difference-in-differences regressions of the effect of having dangerous pipe materials on a student's academic outcomes during the Flint Water Crisis using a panel spanning 2010-2019. All models include student fixed effects and year fixed effects. Education data is taken from the Michigan Department of Education's longitudinal administrative data base. Service line material data was collected during the City of Flint's service line inspection and replacement program that was implemented in the aftermath the crisis. Lead or galvanized steel service lines are classified as dangerous, whereas copper service lines are classified as not dangerous. These difference-in-differences models are estimated using OLS regression.

Table A9. Difference in Difference Results: Alternative Lead/Copper Definitions.

	(1)	(2)	(3)	(4)
	Math Achievement	Reading Achievement	Special Needs	Attendance
A.				
Lead Pipes * Post	-0.0262	-0.00168	-0.00382	-0.000536
	(0.0191)	(0.0220)	(0.00526)	(0.00160)
Control Mean	{-.62}	{-.60}	{.175}	{.917}
Students	8094	8073	10165	9931
Student-Years	35104	34992	79158	66118
B.				
Lead2 Pipes * Post	-0.0253	0.00351	-0.00479	-0.00145
	(0.0194)	(0.0223)	(0.00533)	(0.00164)
Control Mean	{-.62}	{-.60}	{.175}	{.917}
Students	7907	7886	9913	9683
Student-Years	34290	34183	77217	64475
C.				
Lead3 Pipes * Post	-0.0292	-0.00288	-0.00427	-0.00167
	(0.0197)	(0.0225)	(0.00540)	(0.00166)
Control Mean	{-.62}	{-.60}	{.176}	{.917}
Students	7723	7702	9685	9460
Student-Years	33504	33399	75436	63008

* $p<.05$; ** $p<.01$

Note. This table displays results from difference-in-differences regressions of the effect of having lead + copper (as opposed to only lead) pipe materials on a student's academic outcomes during the Flint Water Crisis using a panel spanning 2010-2019. All models include student fixed effects and year fixed effects and control for having galvanized steel pipes. In the difference-in-differences models presented in the main text, the "Lead" variable is equal to l for all service line materials considered dangerous (and scheduled for replacement by Flint's FAST Start team) and is equal to 0 for all other service line materials. In the Panel A above, the "Lead2" variable is equal to 1 only for lead or galvanized steel service, equal to 0 for only copper or safe non-copper service lines, and is set to missing otherwise. In Panel B above, the "Lead3" variable is equal to 1 for only lead service lines, is equal to 0 only for copper service lines, and is set to missing otherwise. These difference-in-differences models are estimated using OLS regression.

Table A10. Difference in Differences Results by Grade Cohort, Gender, and Administrative District.

	(1) Math Achievement	(2) Reading Achievement	(3) Special Needs	(4) Attendance
By Grade Cohort				
A. Grades PK-2 in 2014				
Lead Pipes * Post	—	—	-0.00916	0.00363
	—	—	(0.0115)	(0.00284)
Control Mean	{-.72}	{-.65}	{.168}	{.914}
B: Grades 3-7 in 2014				
Lead Pipes * Post	-0.0191	-0.00221	-0.0113	-0.000857
	(0.0191)	(0.0223)	(0.00810)	(0.00225)
Control Mean	{-.61}	{-.58}	{.177}	{.918}
C: Grades 8-12 in 2014				
Lead Pipes * Post	—	—	0.00898	-0.00105
	—	—	(0.00772)	(0.00335)
Control Mean	{-.56}	{-.60}	{.151}	{.917}
By Gender				
D: Male				
Lead Pipes * Post	-0.0125	-0.00363	-0.00355	0.000268
	(0.0269)	(0.0310)	(0.00816)	(0.00231)
Control Mean	{-.63}	{-.72}	{.231}	{.913}
E: Female				
Lead Pipes * Post	-0.0383	-0.00342	-0.00389	-0.000468
	(0.0270)	(0.0305)	(0.00647)	(0.00226)
Control Mean	{-.61}	{-.48}	{.115}	{.921}
By Administrative District				
F: Flint Community Schools				
Lead Pipes * Post	-0.0287	0.0119	0.00117	-0.00218
	(0.0283)	(0.0322)	(0.00770)	(0.00242)
Control Mean	{-.76}	{-.76}	{.179}	{.91}
Lead Pipes * Post	-0.0123	-0.00390	-0.00812	0.000790
	(0.0257)	(0.0305)	(0.00721)	(0.00212)
Control Mean	{-.51}	{-.48}	{.171}	{.923}

* $p<.05$; ** $p<.01$

This table displays results from various subgroup analysis using difference-in-differences regressions of the effect of having lead pipe materials on a student's academic outcomes during the Flint Water Crisis using a panel spanning 2010-2019. All models include student fixed effects and year fixed effects and control for having galvanized steel pipes. These difference-in-differences models are estimated using OLS regression.