Evaluating the Impact of Arsenic-Contaminated Drinking Water on Educational Outcomes

Madalyn Romberger

October 21, 2025

Abstract

Research shows that high arsenic exposure harms children's development, but its effects in high-income countries or at lower concentrations remain unclear. This study examines how arsenic-contaminated water affects children's attendance and test scores in Maryland. I use a unique dataset linking school-level water treatment and local arsenic concentrations to estimate students' exposure. My findings suggest reducing in-school arsenic exposure through water treatment improves attendance but has little effect on test scores. Effects vary across grade spans, underscoring the need for further research to understand the consequences of arsenic exposure.

1 Introduction

In the United States, concerns about access to clean drinking water are growing, particularly in the wake of water crises involving toxins such as lead. In schools, water purity is often unknown and threatened by outdated infrastructure and a lack of water treatment centers. Many states have taken action to address these concerns and improve students' water quality, but testing and policies are limited to pollutants with well-established negative impacts, namely lead. Contaminants like arsenic are often overlooked, and their effects on adolescents are not well-established. Economic research on arsenic is limited to countries such as Bangladesh and China, where childhood exposure to high concentrations of arsenic has been linked to decreased intellectual function, test scores, earnings, and productivity (Asadullah & Chaudhury, 2008; Pitt, Rosenzweig, & Hassan, 2021; Wang et al., 2007). This paper seeks to be the first to establish the impact of arsenic on children's educational outcomes in the United States and the overarching context of high-income, developed countries.

This paper focuses on elementary and middle school students in Maryland, where exposure to arsenic occurs through contaminated drinking water resulting from a high natural occurrence of arsenic in underground aquifers and the high prevalence of well water use within schools and households. Students' water treatment status is based on access to public water utilities or treated well water at the school level. This is combined with inferred arsenic exposure based on contamination in the well geographically closest to the school and acts as a proxy for potential exposure, allowing me to study this under-researched contaminant.

To understand how arsenic exposure impacts students' attendance and test scores, I leverage geographical variation in arsenic contamination and the timing of school-level water treatment in a difference-in-difference estimation strategy, employing stacking estimation as used in Cengiz, Dube, Lindner, and Zipperer (2019). Unlike other staggered timing difference-in-differences approaches such as Callaway and Sant'Anna (2021) and Sun and Abraham (2021), stacking supports both binary and continuous treatment, allowing for a consistent methodology across school and cohort specifications while avoiding concerns of bad comparisons biasing results.

My findings suggest that arsenic negatively impacts attendance. I observe a significant decrease in the percentage of students missing more than twenty days of schools, and an increase in the percent of students missing fewer than 5 days of school. Limited effects are observed for math or reading testing proficiency. Results are heterogeneous between elementary and middle schools, suggesting the potential for differential impacts across students' ages. Additionally, results are robust to different specifications of arsenic exposure.

In addition to its policy relevance, my research supports three different facets of the economics literature. First, this paper adds to the literature on the impact of environmental pollution on children's educational outcomes. Existing literature has established the negative impacts of air and soil pollution on children's test scores, attendance rates, and behavior, especially as a result of lead pollution (Aizer & Currie, 2019; Aizer, Currie, Simon, & Vivier, 2018; Gazze, Persico, & Spirovska, 2022; Hollingsworth, Huang, Rudik, & Sanders, 2025). This paper expands the limited subset of such research relating to water pollution to provide a more comprehensive understanding of the different elements of children's environments and their potential impacts. My findings suggest that arsenic may have similar impacts.

Using hand-collected school water treatment data, I enhance the literature on the effects of arsenic, which is limited primarily to high concentration levels in low-income countries. Existing studies have access to survey data containing individual-level arsenic exposure from wells, urine measurements, and toenail clippings and find adverse impacts on boys' test scores, intellectual function, and cognition that manifests into lower schooling attainment and earnings (Asadullah & Chaudhury, 2008; Pitt et al., 2021; Wasserman et al., 2004). However, conclusions from these studies may not be externally valid to the context of the United States due to differences in level and circumstance of exposure. I find that arsenic may negatively impact students' health and development at low levels of exposure in high-income countries. Reducing exposure results in attendance improvements, and positive but insignificant improvements in testing proficiency.

My results also suggest arsenic exposure below the EPA-regulated limit may negatively impact students' development, which is currently only established in the medical literature. Similar to results for lead or high exposure to arsenic, chronic low-dose exposure to arsenic in my sample also make students physically ill, resulting in decreased attendance. To further enhance the understanding of acute versus chronic exposure, my study derives the cumulative effects of arsenic exposure over a student's time in school to understand the importance of length and intensity of exposure. This differs from existing pollution studies, which traditionally exploit students' one-time exposure to contaminants.

Overall, my findings suggest that arsenic may negatively impact the attendance and testing proficiency of elementary and middle school students, and this effect may vary across different schools and age groups. The small sample size and insignificant results for test scores underscore the need for continued research to understand the nuanced effects of arsenic exposure in high-income countries. Lastly, this study highlights the importance of re-evaluating existing regulations and policies to ensure students have access to clean drinking water in schools and tests the longstanding EPA regulation for arsenic, which currently permits contamination of up to 10 micrograms per liter (ug/L) in drinking water.

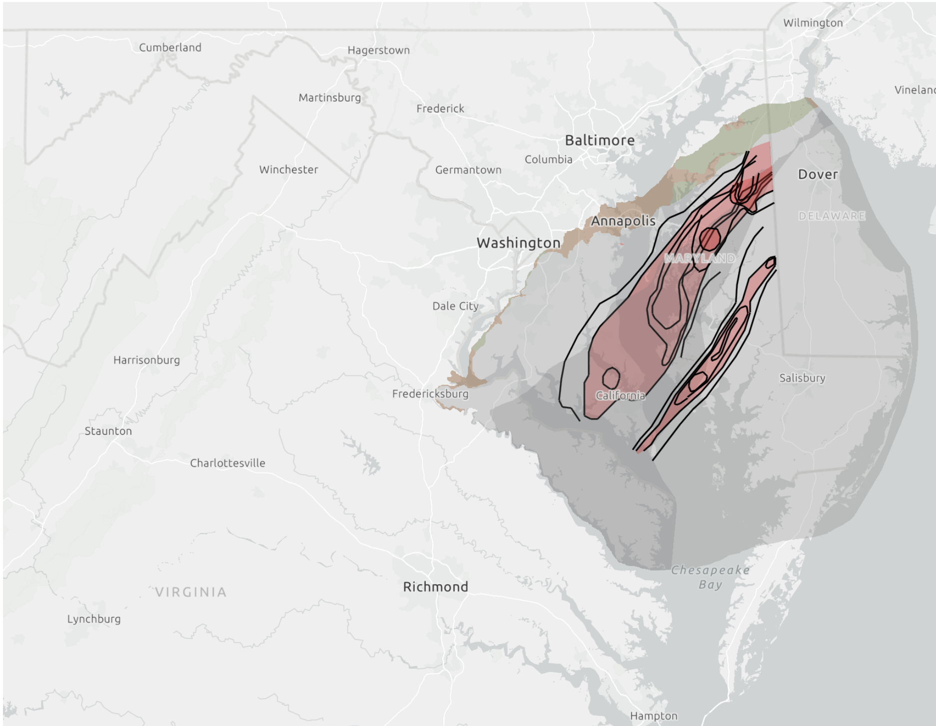
2 Background

Arsenic is a carcinogenic and neurotoxic metalloid that exists in the natural environment and is a byproduct of industrial, mining, and farming activities. Its widespread presence puts millions of individuals worldwide, including in the United States, at risk. While most exposure occurs through drinking water, it can also occur by consuming contaminated foods or exposure to contaminated soil, dust, or treated wood, posing a significant and severe health risk (ATSDR, 2007). Often only presenting symptoms after prolonged exposure, the most common side effects of arsenic exposure in adults are skin lesions and internal cancers (Council, 1999). In children, arsenic negatively impacts brain function and development by stunting the brain's growth and degrading existing neurotransmitter systems (Htway, Sein, Nohara, & Win-Shwe, 2019). Early life and in-utero exposure to arsenic is incredibly harmful and can manifest in memory, intelligence, and attention problems and increased mortality as a result of a higher likelihood of cancers (Smith et al., 2006; Tolins, Ruchirawat, & Landrigan, 2014; Tsai, Chou, The, Chen, & Chen, 2003). Therefore, removing arsenic from students' drinking water should theoretically improve test scores and behaviors.

Concerns about arsenic exposure are particularly relevant given that 43 million people in the United States consume water from groundwater wells, and an estimated 2.1 million people drink from wells with arsenic concentrations above the EPA standard of 10 ug/L (Ayotte, Medalie, Qi, Backer, & Nolan, 2017; DeSimone, Haminton, & Gilliom, 2009). In Maryland, where arsenic is naturally occurring and not expected to be related to anthropogenic contamination, it is estimated that 13% of residents rely on wells (MDE, 2010). Surveys of these wells found arsenic contamination ranging from below detection limits (generally 2 ug/L) to 131 ug/L, with approximately 37% of sampled wells having detectable arsenic and 10% of sampled wells having concentrations above 10 ug/L (Drummond & Bolton, 2010). Figure 1 highlights the geographical variation in contamination across Maryland, with the contours indicating exposed areas and darker contours indicating areas of higher exposure. Contamination is most prevalent in the Aquia and Piney Point Aquifers, which largely affect the southeastern portion of the state (Figure 1).

Arsenic exposure is difficult to identify because arsenic is colorless, odorless, and tasteless in drinking water, therefore people are unlikely to know they're being exposed to arsenic without a water test. Further, testing humans for arsenic exposure is more challenging than testing for lead, which can be done easily with a blood test. Urine tests for arsenic indicate recent exposure but cannot accurately predict what level of arsenic a person has been exposed to or absorbed. Fingernail clippings provide more accurate results for long-term exposure, but only for high levels, and are difficult to obtain. Therefore, measuring the treatment status of wells likely to contain arsenic is useful for gauging exposure to this

Figure 1: Piney Point and Aquia Aquifer range in Maryland



Source: Drummond & Bolton (2010)

elusive contaminant.

Arsenic cannot be naturally abated by the environment; it can only change form and be removed by being attached or separated from other particles through treatment and filtration. EPA-acceptable methods for arsenic removal include pre-oxidation, pH adjustment, ion exchange, activated alumina (AA), reverse osmosis (RO), enhanced lime softening, and enhanced coagulation/filtration (EPA, 1999). Water treatment facilities in the United States must filter and treat their water to ensure that it meets quality standards enforced by the EPA before distributing it to customers such as schools. Both the WHO and EPA suggest a 10 ug/L standard for arsenic in drinking water. The EPA also reports adverse health effects at exposure levels of 3, 5, 10, and 20 ug/L (EPA, 2000). However, the current legal limit of 10 ug/L is the result of a cost-benefit analysis weighing adverse effects on health with high remediation costs (EPA, 2000).

3 Empirical Strategy

This study examines the effects of arsenic exposure on students' attendance and test scores by analyzing changes in arsenic exposure at the school level. Specifically, I evaluate how the implementation of water treatment affects these educational outcomes. Pre-existing arsenic groundwater concentrations approximate students' potential exposure, and changes in school water treatment are used to exploit plausibly exogenous changes in this exposure. Results reflect the effects of reducing potential exposure to arsenic. Specifications are provided first at the school level, which is the level of treatment, and then at the cohort level to exploit further variation in duration and intensity of exposure (Appendix B).

At the school level, I investigate if removing arsenic from drinking water through water treatment affects attendance and test score outcomes. To do so, I create treatment-year-specific datasets corresponding to the year schools begin treating their water, ranging from 2003 to 2023. Each h-specific dataset, referred to as a "stack", includes the schools treated in year h and control schools. Control schools are other schools in Maryland that do not receive treatment in the 11-year window ($t = -4, .., 6$) around year h. This window represents the years a student may attend the school before their outcomes are observed. All other schools are dropped from the stack to avoid negative weighting by bad comparisons between early and late-treated schools. All stacks are then appended into one dataset for analysis to maximize power and estimate the average effect across event times, β_1. The specification is as follows:

$$y_{sdth} = \sum_{\substack{\tau=-4 \\ \tau\neq-1}}^{6} \gamma_\tau D_{sh,t-\tau} \times \text{Arsenic}_s + \sum_{\substack{\tau=-4 \\ \tau\neq-1}}^{6} D_{sh,t-\tau} + X_{st_{2002}} \cdot t + \delta_{dh} \cdot t + \delta_{sh} + \delta_{th} + \epsilon_{sdth} \quad (1)$$

with corresponding differences in differences specification:

$$y_{sdth} = \alpha + \beta_1 \text{Post}_{sdth} \times \text{Arsenic}_s + X_{st_{2002}} \cdot t + \delta_{dh} \cdot t + \delta_{sh} + \delta_{th} + \epsilon_{sdth} \quad (2)$$

y_{sdth} represents attendance or test score outcomes for school s in district d observed in year t. Attendance outcomes include the percentage of students who miss more than 20 days of school (high absenteeism) and the percentage of students who miss fewer than five days of school (low absenteeism). Test score outcomes are presented as the percentage of students who score proficient or better on reading or math exams at school s in year t, weighted by the number of test takers in each grade. The results will discuss these simply as the percentage of proficient students.

$D_{sh,t-\tau}$ is a binary variable equal to one if school s is treated in year t. This occurs if a school supplies clean, treated water that is free from contaminants, i.e., arsenic, to students. This can occur in two ways: (1) schools connect to a public water utility that monitors water quality and treats the water for contaminants, or (2) schools begin treating well water on-site. Exposure to lead is not expected to change with this form of treatment because it is usually present in the school's piping and not in the source water. Schools are considered always treated if their water has always come from a public utility or a treated well, and never treated if the opposite is true. To proxy for the severity of exposure at each school, $Arsenic_s$ is equal to one if the concentration of arsenic in the closest measured groundwater well is greater than three ug/L. This measure is chosen because it is the lowest measure found by the EPA to induce negative side effects (EPA, 2000). While 10 ug/L is the EPA maximum allowable concentration and would be a preferable comparison, my small sample size hinders precise estimation at this level.

The coefficients of interest from Equation 1, γ_τ, represent the impact of water treatment on student outcomes over time. Specifically, they capture the change in test scores or attendance for schools with arsenic exposure greater than three ug/L relative to their pre-treatment baseline and relative to control schools. I can use these coefficients to assess how the effects of treatment evolve over time and if the effects are immediate and sustained. Positive coefficients for test scores indicate improvements post-treatment and highlight the negative impacts of arsenic exposure. For attendance results, a similar conclusion can be drawn for negative coefficients in the high absenteeism results and positive coefficients in the low absenteeism results. The event study plots in Figure 3 display these patterns, while Table 3 presents β_1 coefficients from Equation (2). These coefficients can be interpreted as the aggregate post-treatment effect of removing students' potential arsenic exposure via water treatment.

This specification relies on the exogeneity of treatment timing, which cannot be confirmed theoretically due to a lack of information regarding states, districts, or school-level decision-making to implement water treatment. Existing information regarding connections to public water utilities or the decision to treat well water does not indicate that this decision is driven by arsenic exposure. One potential explanation could be that economies of scale drive this decision; as communities grow larger it is more economical to provide public utilities. District-by-year fixed effects, $(\delta_{dh} \cdot t)$, control for time-varying factors that could influence the decision to implement water treatment, such as changes in district composition, policies, or funding opportunities. Including these fixed effects reduced potential biases related to endogeneity in the timing of treatment adoption.

Additional elements in this specification include controls observed for all schools in 2002, $X_{st_{2002}} \cdot t$. These controls include pupil-to-teacher ratio, the percentage of students receiving subsidized lunch, the proportion of non-White students, and the percentage of female students. These controls capture differences in school-level resources and demographics, controlling for correlations between resources, race, and income with test scores and attendance outcomes. To maintain consistency across schools, all controls were observed in 2002, the year before I observed outcomes. Due to many schools treating their water before I observe characteristic information, including characteristics gathered from different points in time and treatment statuses' would not produce accurate comparisons. Further, school-by-stack (δ_{sh}) and year-by-stack (δ_{th}) fixed effects account for unobserved time-invariant differences across schools within stacks and time-varying shocks impacting all schools within stacks, respectively. Standard errors are clustered at the school level because it is the level of treatment.

The previously outlined school-level specification provides an overview of arsenic's impact but could be improved by following the same students over time. Without individual-level data, I extrapolate cohort-level math and reading outcomes from grade-level proficiency. Attendance data is not availible at the grade level, so I do not include cohort-level attendance estimates. I use this information to build "cohorts" that are unique at the school and start year level, which allows me to track the same students over time and estimate the length of their exposure, gaining power and precision. This method assumes a consistent cohort and cannot account for attrition from students who are held back or switch to another school. The cohort specification and results are included in Appendix B and support the results of the school-level analysis.

For both methodologies, the two-way fixed effects (TWFE) methodology may be biased in this specification due to the staggered timing of water treatment, particularly if treatment effects are heterogeneous across periods. More specifically, because treatment occurs at different times, TWFE may incorrectly calculate treatment effects by using comparisons between early- and late-treated groups, which no longer follow parallel trends. To address this concern, I use the stacking approach outlined by Cengiz et al. Cengiz et al. (2019) to manually remove comparisons between early and late-treated groups from the specification, which could otherwise bias results. Employing the most up-to-date stacking methodologies, I also estimate the stacked model using weights introduced by Wing, Freedman, and Hollingsworth Wing, Freedman, and Hollingsworth (2024), who suggest that if the number of units treated in each period differs, traditional stacking methods could be biased and driven by the treatment effects of few schools. Instead, this specification weights results in the average treatment effect by the proportion of treated units in each stack, which resolves this concern. These results are presented in Appendix A.

This study assumes that, without treatment, the potential outcomes for treated and untreated schools would follow parallel trends. The adherence to parallel trends can be observed in the event study plots of Figure 3. Additionally, I do not expect anticipation effects to bias estimates. While there is no uniform or formal notification of water treatment implementation, secondary notification through news channels or rumors might influence outcomes. However, I did not find substantial evidence of these forms of notification and do not believe they will significantly affect student or parent decisions or school procedures before treatment. Estimates could still be biased if time-varying school-specific factors are covariant with students' attendance or test scores.

4 Data

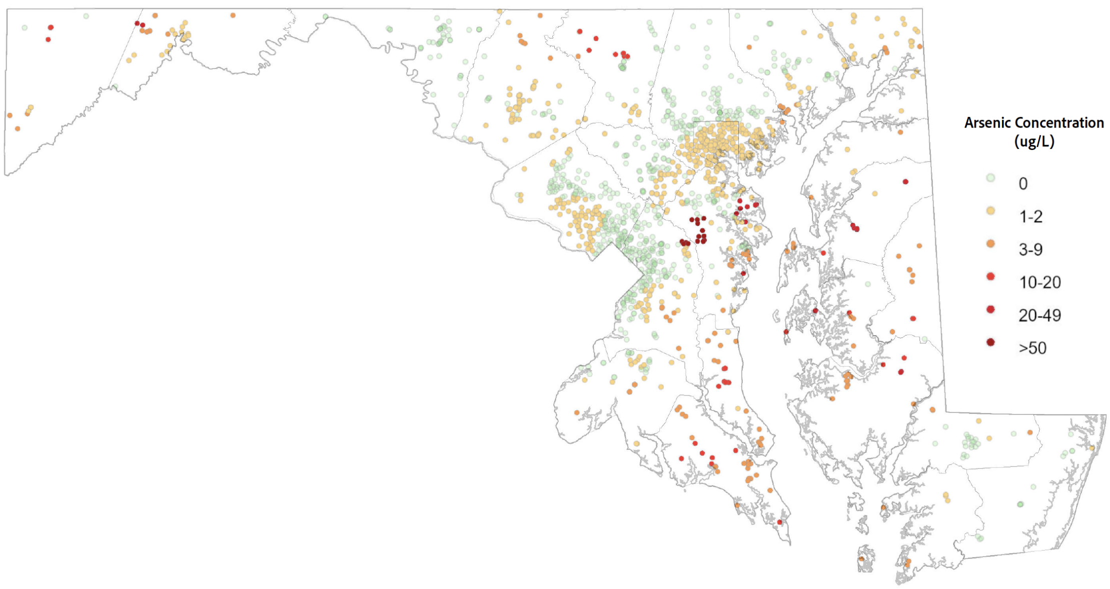
4.1 School Data

To assess the impact of arsenic contamination on children's educational outcomes, I compile data from local, state, and national sources to understand pollution levels, children's educational indicators, and treatment status. I limit my sample to public elementary adn middle schools. To understand which schools are most impacted by arsenic-contaminated drinking water, I collect arsenic well testing results from the United States Geological Survey (USGS), Maryland Geological Survey (MGS), and the National Water Quality Monitoring Council (NWQMC). Combining these sources provides the most robust geographic coverage to provide a more reliable understanding of students' exposure, with over 1,509 unique arsenic results after removing duplicates from the same well, retaining the measure with the highest concentration.

Arsenic cannot be detected in very small amounts using current testing procedures. The USGS has a minimum detection limit of 1 ug/L, MGS of 2 ug/L and the NWQMC includes a variable for "not detected". In these situations, I record the arsenic measure as a zero. Concerns regarding the number of zeros or undetected arsenic measures are expected to bias results toward zero and could result in an underestimation of the true extent and impact of contamination. To decrease dependence on zeros, I provide alternative specifications in

Table 5 that use the average and maximum across the two closest wells and observe relatively consistent results across specifications.

Figure 2: Map of school locations shaded by arsenic concentration



Source: Figured generated by author using data from USGS, MGS, and MDE

From this comprehensive data, I extracted the coordinates and measured arsenic concentration from each well. Using ArcGIS, I mapped each school to the geographically closest well using coordinate matching. The measured concentrations at the school level are displayed in Figure 2, highlighting the geographical variation and intensity of exposure. The geographical variation in this figure mirrors that of Figure 1, which indicates that arsenic is largely concentrated in the Piney Point and Aquia aquifers. Other notable areas of high concentration include western Maryland, which also has pockets of arsenic-contaminated groundwater. In this figure, each point represents an individual school. Dense clusters of schools indicate urbanicity, which visually appears to be negatively correlated with arsenic exposure.

To assess the impact of arsenic exposure on these outcomes, I leverage school-level water treatment status. This information was hand-collected from each school district in Maryland through a series of Public Information Access requests of schools, public facilities, and other

Table 1: School-Level Treatment Decomposition

Concentration	Never-Treated	Treated	Always-Treated	Total
As = 0	19	75	476	570
0 < As < 3	8	20	75	103
3 ≤ As < 5	6	4	7	17
5 ≤ As < 10	8	7	16	31
As ≥ 10	3	5	7	15
Total	44	111	581	736

organizations and cross-referenced with school master plans and news articles when available. The collected information contains the school's water source (groundwater well or public) and the year the school begins receiving water from that source. If schools are on well water and begin filtering water within the school, the date of treatment activation is recorded.

Treatment status for Garrett, Howard, and Kent County school districts is unavailable, so they are excluded from the dataset. Baltimore City and Baltimore County districts are also excluded because they started providing bottled water in all district schools during the sample period. This change aimed to eliminate lead exposure from contaminated piping, making these schools unsuitable as controls. Within the remaining districts, if information regarding the start date of treatment is unavailable, the school is assumed to be always treated, and the school's operating date is used as a proxy.

Table 1 combines the concentration information from Figure 2 with treatment status during the study period to provide an overview of the number of schools that fit into each level of arsenic exposure and treatment status. The largest group of schools has low exposure and is treated before the study period ("always treated"), followed by low exposure and treated during the sample period, providing ample control observations for this study. For schools that have arsenic exposure, most have exposure between five and ten ug/L.

Table 3 provides summary statistics for school and cohort characteristics and outcomes, separated by treatment status, broken into four categories: never treated, treated during the sample period, always treated, and no arsenic exposure.

Public school characteristics are collected from the Common Core of Data (CCD). The CCD is a panel of school characteristics from 1986-2022 compiled from surveys of each school in the United States. From this source, I retain a school identifier, the type of school (elementary or middle), the percentage of students receiving free or reduced lunch (used as a proxy for income), the percentage of non-White students, and the percentage of female students. I generate a variable for the pupil-to-teacher ratio, calculated as the ratio of school enrollment to full-time equivalent teachers. The values of these variables from 2002, before I observed any outcomes, are used as controls and remain consistent across time and do not vary with the introduction of water treatment.

A few key differences in school characteristics across treatment status are observed in Table 2. To start, treated schools have lower enrollment and subsequent lower pupil-to-teacher ratios. This is likely a result of their more rural location, corresponding to the lower portion and rightmost portion of Figure 2. Demographic characteristics are similar across schools exposed to arsenic but differ from those in unexposed schools. Schools with no arsenic exposure are more racially diverse and of lower socioeconomic status, which could be correlated to their urbanicity.

The school outcomes included in Table 2 are sourced from the Maryland School Report Card and include attendance and test score measures for all Maryland public schools, which are published annually. Measures of attendance include the percentage of students who missed more than 20 days (high absenteeism) and the percentage of students who missed fewer than five days (low absenteeism). This information is collected from 2003-2023 and is truncated at 5% and 95% by FERPA; this truncation is not expected to impact results. Low absenteeism is relatively consistent across treatment categories, while treated schools have higher rates of high absenteeism. A potential source for this variation could be lower access to healthcare or other resources influencing children's ability to attend school.

I compile test score information into the percentage of students scoring proficient or better in math and the percentage of students scoring proficient or better in reading. These

Table 2: School Level Summary Statistics

	Never-Treated	Treated	Always Treated	No Arsenic
Panel A: Arsenic Exposure				
As Concentration	4.920	7.143	4.571	0.000
	(3.463)	(19.039)	(12.322)	(0.000)
As Avg of 2 Closest	4.380	5.614	4.505	0.182
	(2.909)	(9.948)	(9.473)	(2.356)
As Max of 2 Closest	5.040	8.400	6.438	0.365
	(3.494)	(19.133)	(16.342)	(4.711)
Panel B: Pre-Period School Characteristics				
Enrollment	516.680	373.867	603.971	592.197
	(212.166)	(178.146)	(234.019)	(231.776)
Pupil to Teacher Ratio	16.739	14.821	16.980	16.047
	(2.200)	(2.249)	(2.375)	(2.769)
% Female	0.497	0.481	0.486	0.485
	(0.026)	(0.025)	(0.023)	(0.027)
% Non-White	0.150	0.157	0.270	0.491
	(0.114)	(0.125)	(0.232)	(0.341)
% Receiving Free/Reduced Lunch	0.215	0.279	0.255	0.330
	(0.118)	(0.151)	(0.188)	(0.230)
Panel C: Outcomes				
% Missed >20 Days	0.083	0.091	0.082	0.083
	(0.027)	(0.044)	(0.034)	(0.033)
% Missed <5 Days	0.345	0.337	0.364	0.383
	(0.033)	(0.060)	(0.059)	(0.060)
% Prof or Better Math	0.807	0.806	0.801	0.766
	(0.098)	(0.114)	(0.098)	(0.133)
% Prof or Better Reading	0.841	0.833	0.832	0.803
	(0.062)	(0.093)	(0.081)	(0.106)
N	25	36	105	570

This table includes school level characteristics. Measures of arsenic are presented as micrograms per liter (ug/L). All outcome variables are average over the observed period. All school characteristics are observed in 2002, the year prior to treatment. Treated and never-treated are both limited to schools with arsenic exposure greater than 0.

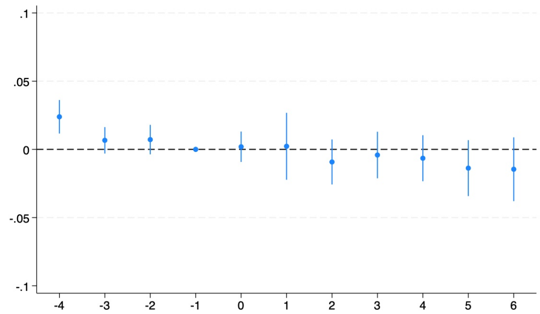
results are from the Maryland School Assessment (MSA), which began testing 3rd through 8th grade students against a standard curriculum in 2003. I observe these test scores through 2016 when the standardized testing method was changed, rendering results in-comparable. Further, concerns regarding a change in proficiency standards beginning in 2013 could result in bias in an indeterminable direction, so proficiency measures from these years are excluded from the estimation. The percentage of students in each grade that score basic (below expectation), proficient (at expectation), or advanced (above expectation) are reported. I aggregate these outcomes to create the percentage of students scoring proficient or better within each school and year and weight the estimate by the number of test takers in each grade. Test scores and attendance outcomes are linked to CCD characteristics using a common school identifier. Across treatment status, proficiency is relatively consistent, with approximately 80% of students scoring proficient or better in math and 83% scoring proficient or better in reading. This is likely correlated with the lower socioeconomic status, higher racial diversity, and overall urbanicity of schools not exposed to arsenic (Gagnon & Mattingly, 2018; Reardon, Kalogrides, & Shores, 2019)).

The difference between treated and control observations in Table 2 suggests that there may be underlying concerns about the endogeneity of treatment status. Differences between treatment and exposure schools highlight the necessity of including controls to absorb factors such as race and income, which would be correlated with test scores and attendance outcomes. Further, my research did not uncover any information suggesting that the decision to treat schools was driven by arsenic exposure in these areas. One potential motive could be economies of scale, with the provision of community water treatment becoming more cost-effective as populations increase, which I control for using district-by-year fixed effects. Concerns of endogeneity influenced my decision to utilize a difference-in-difference approach for estimation, which only requires that treated and control schools trend similarly in the absence of treatment and allows for level differences.

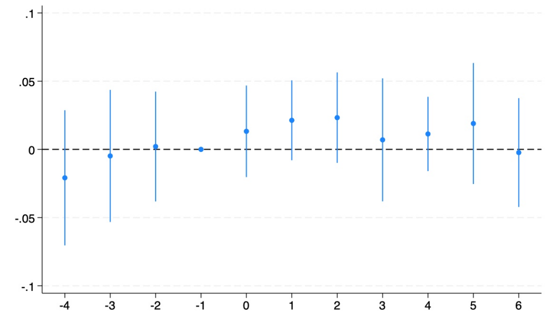
5 Results

Event study estimates for school-level outcomes from Equation (1) are presented in Figure 3. For comparability across specifications, four pre-periods and six post-periods are presented, representing the years a student may attend school before being tested. Aside from reading proficiency (Panel D), these event studies suggest that treated and untreated schools were trending similarly prior to treatment.

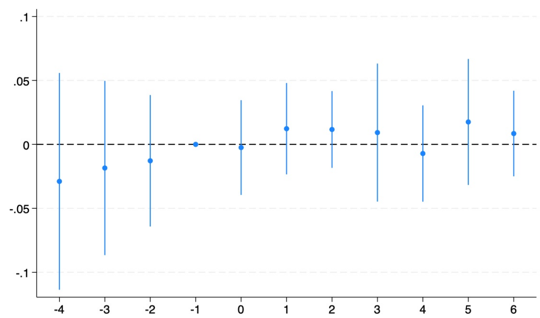
Figure 3: School Event Study Plots



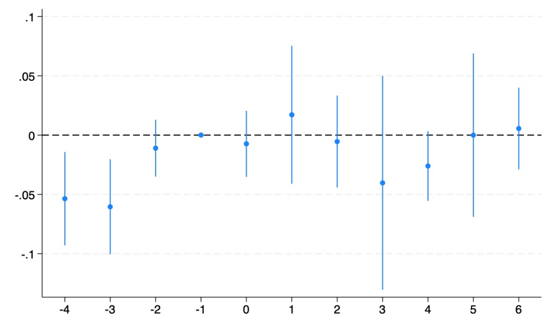
Panel A: Missed > 20 Days



Panel B: Missed < 5 Days



Panel C: Math Proficient or Better



Panel D: Reading Proficient or Better

Note: These plots are a result of unweighted stacking regression from Equation (1) with 2002-level controls for pupil-to-teacher ratio, percentage of students receiving free lunch, percentage of non-White students, and percentage of female students with district-by-year, year-by-stack, and school-by-stack fixed effects. The horizontal axis measures years since treatment, and the vertical axis is the percentage point change in the outcome of interest. Standard errors are clustered at the school level.

Table 3: Difference-in-Difference School Results

	Miss > 20 Days	Miss < 5 Days	Math Prof	Reading Prof
Arsenic$\geq 3\times$Post	-0.015*	0.019	0.021	0.021
	(0.008)	(0.013)	(0.023)	(0.023)
N	87,412	87,412	59,280	59,306

Notes: Standard errors clustered at the school level are shown in parentheses. Estimates are from the unweighted stacking regression of Equation (2) with 2002-level controls for pupil-to-teacher ratio, percentage of students receiving free lunch, percentage of non-White students, and percentage of female students. All regressions include district-by-year, year-by-stack, and school-by-stack fixed effects. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Panel A suggests a small downward trend in the percentage of students missing more than 20 days of school. This result suggests a decrease of 1.5 percentage points on a mean of 9.1% after treatment for schools with arsenic concentrations above three ug/L (Table 3). This reduces high absenteeism by approximately 16.5% and reduces the gap in absenteeism between treated and always treated or no arsenic schools. This is supported by the upward trend in students' with very low absenteeism, missing fewer than five days of school (Panel B). Medically, low-level chronic exposure to arsenic is known to be carcinogenic, but gastrointestinal symptoms are often absent. My results suggest there may be some side effects of arsenic exposure that affect students' attendance beyond gastrointestinal symptoms and diversify the understanding of the impact of arsenic on young student bodies. Comparatively, existing literature from Bangladesh finds that high exposure levels result in attendance changes (Murray & Sharmin, 2015).

Comparatively, the percentage of students scoring proficient or better on reading exams only increases by 2.1 percentage points, and this result is not statistically significant (Panels C and D). Other pollution research that finds that pollution exposure impacts math test scores more than reading (Hollingsworth et al., 2025; Jacqz, 2022). However, this result is not significant nor sustained across time, suggesting that arsenic exposure did not affect proficiency. Wide confidence intervals in my estimates suggest substantial variation in treatment effects. This is likely a result of the limited number of schools in my sample exposed to arsenic and treated during the sample time, which could reflect unobserved heterogeneity within schools or varying degrees of true arsenic exposure. Future research with larger sample sizes and more accurate exposure measures could help narrow these estimates. Still, because no other research exists on lower-level exposure or exposure within the United States for arsenic pollution, my results provide a reference point for future research.

5.1 Heterogeneity Analysis

To understand which students are most impacted by exposure to arsenic, I split the sample into elementary and middle school sub-samples and present these results in Table 4. These results suggest that improvements to students' attendance following water treatment are larger for middle school students. 2.7 percentag points fewer students miss more than 20 days of school, and 7.7 percentage points more students miss fewer than 5 days of schools. These gains are smaller for elementary school students. For testing proficiency, when stratified by grade span, results suggest that reading proficiency significantly increased by 3.1 percentage points for middle school students, with insignificant effects for elementary reading proficiency, or math proficiency in either school type.

One potential mechanism that may be driving this difference is the accumulation of exposure over the student's lifetime. As arsenic passes through the body, it is absorbed by soft tissues such as the brain. The accumulation of arsenic in these soft tissues may only show measurable effects after prolonged periods. This could also be compounded by the higher neuroplasticity of younger children's brains, allowing them to overcome damage by arsenic better than middle school students. Alternatively, it could be that differences in the type and level of learning within the middle school may be more impacted by the degradation of cognition caused by the arsenic. Further, results suggest that there may be more significant health improvements across middle school students' lifetimes that allow them to attend school more consistently once arsenic is removed. Overall, both groups are physically impacted by the presence of arsenic in their drinking water and see their absenteeism decline after its removal.

For comparison, Hollingsworth et al. (2025) find negative impacts of pollution on younger and older children. For lead, which is similar to arsenic, exposure in young children compounds into lower attainment and higher delinquency (Aizer & Currie, 2019; Aizer et al., 2018; Chandramouli, Steer, Ellis, & Emond, 2009). However, it is also suggested that exposure at age six may be more harmful than infantile exposure (Hornung, Lanphear, &

Table 4: Heterogeneity by School Type

	Elementary	Middle
Panel A: Missed > 20 Days		
Arsenic$\geq$ 3$\times$Post	-0.017***	-0.027**
	(0.006)	(0.012)
N	71,083	37,566
Panel B: Missed < 5 Days		
Arsenic$\geq$ 3$\times$Post	0.019	0.077**
	(0.020)	(0.034)
N	71,083	37,566
Panel C: Math Proficient		
Arsenic$\geq$ 3$\times$Post	-0.008	-0.005
	(0.011)	(0.034)
N	53,576	25,712
Panel D: Reading Proficient		
Arsenic$\geq$ 3$\times$Post	0.006	0.031***
	(0.032)	(0.011)
N	53,576	25,505

Notes: Standard errors clustered at the school level are shown in parentheses. Estimates are from the unweighted stacking regression of Equation (2) with 2002-level controls for pupil-to-teacher ratio, percentage of students receiving free lunch, percentage of non-White students, and percentage of female students. All regressions include district-by-year, year-by-stack, and school-by-stack fixed effects. The sample was split to retain only elementary schools or middle schools and run separately. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Dietrich, 2009). This provides evidence to support my results that older children may be more impacted by exposure. The heterogeneity in results highlights the inability of school-level data to follow students' exposure and outcomes across long periods, which requires me to assume that middle school students' treatment before entering the school is the same as the middle school treatment status up to that point. This assumption could affect results if middle school students' exposure is much different than their elementary school exposure. Overall, the results from this section suggest opportunities for further research into who is most impacted by exposure to arsenic.

5.2 Robustness

In all previous estimates, exposure to arsenic was calculated as the level observed in the closest geographical well to the school. These measurements occur at different times and by different sources, and many are coded as zeros. As discussed in the data section, this may lead to an underestimation of the number of schools exposed to arsenic, biasing the results toward zero and underestimating the true effects of exposure. To test how reliant my results are on the composition of arsenic measures, I present two alternative measures of arsenic exposure in Table 5.

Table 5: School Estimates by Arsenic Measure

	Missed >20	Missed <5	Math Prof	Reading Prof
Panel A: Closest Well				
As $\geq 3\times$ Post	-0.015*	0.019	0.021	0.021
	(0.008)	(0.013)	(0.023)	(0.023)
N	87,412	87,412	59,280	59,306
Panel B: Avg of 2 Closest Wells				
As $\geq 3\times$ Post	-0.009	0.016	0.027	0.040*
	(0.008)	(0.015)	(0.034)	(0.021)
N	108,044	108,044	63,551	64,888
Panel C: Max of 2 Closest Wells				
As $\geq 3\times$ Post	-0.008	0.011	0.025	0.028
	(0.006)	(0.013)	(0.028)	(0.019)
N	108,044	108,044	63,551	64,888

Notes: Standard errors clustered at the school level are shown in parentheses. Estimates are from the unweighted stacking regression of Equation (2) with 2002-level controls for pupil-to-teacher ratio, percentage of students receiving free lunch, percentage of non-White students, and percentage of female students. All regressions include district-by-year, year-by-stack, and school-by-stack fixed effects. Three separate methods are used to determine arsenic concentration and associated treatment. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

These alternative measures include an average of the concentration from the two closest wells to the school and the maximum measure between the two closest wells to the schools. By incorporating more nearby data points, these measures are expected to provide a more comprehensive understanding of potential exposure as they may capture arsenic contamination that the single closest well measure misses. Therefore, both measurements are expected to reduce the impact of "zero" concentration measures in the data. All models retain arsenic greater than three ug/L as the treatment variable and only change the method for producing that dummy based on the above characteristics.

Results across models and outcomes remain relatively consistent. Significance of some results varies slightly between models, but conclusions remain consistent (Table 5). Across all models, the average of the two closest wells exhibits coefficients with the highest magnitude, yet the significance of results across these models does not change. Overall, these results highlight that my specification is robust to using different measures for arsenic contamination, and noise from this measure is unlikely to influence the precision or interpretation of the results. Further, the results of this paper reflect the most conservative estimation method of only utilizing the closest single well measurement.

6 Discussions and Conclusion

This study aimed to investigate the impact of arsenic exposure on children's educational outcomes in the United States, specifically in Maryland. By expanding the pollution literature to include the impacts of arsenic, this research sought to understand the severity of this public health concern. The findings of this paper suggest potential negative impacts of arsenic exposure in Maryland on attendance and test scores, but further research is necessary to validate these effect.

Across school and cohort models, the results indicated potential negative impacts of arsenic exposure on high absenteeism, which should be interpreted cautiously for some outcomes due to weak parallel trends in the pre-period. This result aligns with literature from countries with much higher contamination levels that report negative attendance associated with exposure. Further, while not significant, my results suggest arsenic exposure may negatively impact math proficiency, fitting into a larger literature on environmental pollutants and cognitive outcomes. Unlike math, similar suggestions cannot be made for reading, whose results vary across specifications and do not meet the required parallel trends assumption.

Heterogenous effects across schools and cohort analysis, as well as between elementary and middle school specifications, suggest that arsenic's impact might vary based on the duration and length of exposure and the age of exposed students. Effects realized from cumulative exposure to arsenic underscore the long-term effect suggested by significant impacts of arsenic on the math and reading proficiency of middle school students, as well as improvements in attendance after exposure was removed. This result expands the time horizon of exposure in existing literature, which primarily focuses on early life exposure to pollution and finds long-term effects. Potentially, these effects are realized as early as middle school and may culminate into worsened life outcomes, an area for potential continued research.

The findings of this paper are limited by the structure of the data and the inability to follow the same student over long periods of time and control for student-level characteristics. While I attempt to exploit as much variation as possible using the cohort identification presented in Appendix A, this proves insufficient to find statistical significance or balance the noise in the arsenic measure, ultimately resulting in noisy results on how arsenic impacts children's educational performance.

The absence of student-level exposure data meant that exposure could only be used as a proxy, reflecting students' exposure in a school setting rather than at home. Although these exposures were expected to be similar, there currently exists no way to measure this without a tailored study. Additionally, arsenic measures from the closest well measured as far back as the 1950's, and with arsenic expected to increase in groundwater concentration over time, this suggests my measures for arsenic may be an underestimation of the accurate levels in students' drinking water. Further, the arsenic data is littered with zeros, reflecting the inability to accurately measure small concentrations of arsenic, which could bias results downward further if children are affected by these small concentrations.

Overall, this paper attempts to uncover the effect of a sparsely researched contaminant and suggests there may be harmful effects on the health and development of children. I provide some preliminary evidence that at levels below current regulatory limits, children's exposure may lead to physical and cognitive detriment and prompts the need for deeper analysis within the health and economics literature. Further, this research provides the basis for future research on this topic. This research should focus on obtaining more granular data on individual exposure levels and resulting health and education outcomes to establish more definitive causal relationships, and those findings should be considered when policymakers revise water quality standards. Lastly, the lack of available exposure information prompts the need for more rigorous testing procedures and more clear treatment protocols to ensure the safety and well-being of students. In conclusion, this study highlights a need for ongoing research and policy efforts to address pollutants and their impacts on human capital accumulation, particularly in vulnerable populations such as children.

References

Aizer, A., & Currie, J. (2019, October). Lead and Juvenile Delinquency: New Evidence from Linked Birth, School, and Juvenile Detention Records. *The Review of Economics and Statistics*, *101*(4), 575–587. Retrieved 2024-05-14, from https://doi.org/10.1162/rest_a_00814 doi: 10.1162/rest_a_00814

Aizer, A., Currie, J., Simon, P., & Vivier, P. (2018, January). Do Low Levels of Blood Lead Reduce Children's Future Test Scores? *American Economic Journal: Applied Economics*, *10*(1), 307–341. Retrieved 2024-05-14, from https://www.aeaweb.org/articles?id=10.1257/app.20160404 doi: 10.1257/app.20160404

Asadullah, M. N., & Chaudhury, N. (2008, February). *Poisoning the Mind: Arsenic Contamination and Cognitive Achievement of Children* [SSRN Scholarly Paper]. Rochester, NY. Retrieved 2024-06-05, from https://papers.ssrn.com/abstract=1091620

ATSDR, C. (2007). *Arsenic ublic health statement.* Retrieved 2024-07-16, from https://www.atsdr.cdc.gov/ToxProfiles/tp2-c1-b.pdf

Ayotte, J. D., Medalie, L., Qi, S. L., Backer, L. C., & Nolan, B. T. (2017, November). Estimating the High-Arsenic Domestic-Well Population in the Conterminous United States. *Environmental Science & Technology*, *51*(21), 12443–12454. doi: 10.1021/acs.est.7b02881

Callaway, B., & Sant'Anna, P. H. C. (2021, December). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. Retrieved 2024-06-22, from https://www.sciencedirect.com/science/article/pii/S0304407620303948 doi: 10.1016/j.jeconom.2020.12.001

Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019, August). The Effect of Minimum Wages on Low-Wage Jobs*. *The Quarterly Journal of Economics*, *134*(3), 1405–1454. Retrieved 2024-06-23, from https://doi.org/10.1093/qje/qjz014 doi: 10.1093/qje/qjz014

Chandramouli, K., Steer, C. D., Ellis, M., & Emond, A. M. (2009, November). Effects of early childhood lead exposure on academic performance and behaviour of school age children. *Archives of Disease in Childhood*, *94*(11), 844–848. doi: 10.1136/adc.2008.149955

Council, N. R. (1999). *Arsenic in Drinking Water*. The National Academies Press. Retrieved 2024-04-29, from https://nap.nationalacademies.org/catalog/6444/arsenic-in-drinking-water doi: 10.17226/6444

DeSimone, L., Haminton, P., & Gilliom, R. (2009). *Quality of water from domestic wells in principal aquifers of the United States, 1991–2004—Overview of major findings: U.S. Geological Survey Circular 1332* (Tech. Rep. No. 1332). U.S. Geological Survey. Retrieved 2024-07-16, from https://pubs.usgs.gov/circ/circ1332/

Drummond, D. D., & Bolton, D. W. (2010). Arsenic in Ground Water in the Costal Plain Aquifers of Maryland RI 78. Retrieved from http://www.mgs.md.gov/reports/RI_78.pdf

EPA, E. P. A. (1999). *Technologies and Costs for Removal of Arsenic from Drinking Water* (Tech. Rep.). EPA. Retrieved from https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P1004WDI.txt

EPA, E. P. A. (2000). *Arsenic in Drinking Water Rule Economic Analysis* (Tech. Rep. No. EPA 815-R-00-026). EPA. Retrieved 2024-07-16, from https://nepis.epa.gov/Exe/ZyPdf.cgi?Dockey=20001YQT.txt

Gagnon, D. J., & Mattingly, M. J. (2018). Racial/Ethnic Test Score Gaps and the Urban Continuum. Retrieved 2024-09-08, from https://jrre.psu.edu/sites/default/files/2019-08/33-2.pdf (Publisher: The Pennsylvania State University Libraries) doi: 10.18113/P8JRRE3302

Gazze, L., Persico, C. L., & Spirovska, S. (2022, October). The Long-Run Spillover Effects of Pollution: How Exposure to Lead Affects Everyone in the Classroom. *Journal of Labor Economics*. Retrieved 2022-11-08, from https://www.journals.uchicago.edu/doi/10.1086/723175 doi: 10.1086/723175

Hollingsworth, A., Huang, J. M., Rudik, I., & Sanders, N. J. (2025, May). A Thousand

Cuts: Cumulative Lead Exposure Reduces Academic Achievement. *Journal of Human Resources*, *60*(3), 950–976. Retrieved 2025-10-09, from https://jhr.uwpress.org/content/60/3/950 (Publisher: University of Wisconsin Press Section: Articles) doi: 10.3368/jhr.0222-12169R2

Hornung, R. W., Lanphear, B. P., & Dietrich, K. N. (2009, August). Age of Greatest Susceptibility to Childhood Lead Exposure: A New Statistical Approach. *Environmental Health Perspectives*, *117*(8), 1309–1312. Retrieved 2024-09-11, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2721877/ doi: 10.1289/ehp.0800426

Htway, S.-M., Sein, M.-T., Nohara, K., & Win-Shwe, T.-T. (2019, January). Effects of Developmental Arsenic Exposure on the Social Behavior and Related Gene Expression in C3H Adult Male Mice. *International Journal of Environmental Research and Public Health*, *16*(2). Retrieved 2023-10-07, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6352149/ doi: 10.3390/ijerph16020174

Jacqz, I. (2022, September). Toxic test scores: The impact of chemical releases on standardized test performance within U.S. schools. *Journal of Environmental Economics and Management*, *115*. Retrieved 2024-07-16, from https://www.sciencedirect.com/science/article/pii/S0095069622000146 doi: 10.1016/j.jeem.2022.102628

MDE, M. (2010). *Residential Wells.* Retrieved 2024-07-16, from https://mde.maryland.gov/programs/water/water_supply/Pages/default.aspx

Murray, M. P., & Sharmin, R. (2015, December). Groundwater arsenic and education attainment in Bangladesh. *Journal of Health, Population and Nutrition*, *33*(1). Retrieved 2024-07-03, from http://www.jhpn.org/content/33/1/20 doi: 10.1186/s41043-015-0029-6

Pitt, M. M., Rosenzweig, M. R., & Hassan, M. N. (2021, September). Identifying the Costs of a Public Health Success: Arsenic Well Water Contamination and Productivity in Bangladesh. *The Review of Economic Studies*, *88*(5), 2479–2526. Retrieved 2023-11-19, from https://academic.oup.com/restud/article/88/5/2479/6046946 doi: 10.1093/

restud/rdaa087

Reardon, S. F., Kalogrides, D., & Shores, K. (2019, January). The Geography of Racial/Ethnic Test Score Gaps. *The American Journal of Sociology*, *124(4)*(124(4)). Retrieved 2024-09-11, from https://cepa.stanford.edu/content/geography-racialethnic-test-score-gaps

Smith, A. H., Marshall, G., Yuan, Y., Ferreccio, C., Liaw, J., von Ehrenstein, O., ... Selvin, S. (2006, August). Increased Mortality from Lung Cancer and Bronchiectasis in Young Adults after Exposure to Arsenic in Utero and in Early Childhood. *Environmental Health Perspectives*, *114*(8), 1293–1296. Retrieved 2024-04-29, from https://ehp.niehs.nih.gov/doi/10.1289/ehp.8832 (Publisher: Environmental Health Perspectives) doi: 10.1289/ehp.8832

Sun, L., & Abraham, S. (2021, December). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199. Retrieved 2024-07-03, from https://www.sciencedirect.com/science/article/pii/S030440762030378X doi: 10.1016/j.jeconom.2020.09.006

Tolins, M., Ruchirawat, M., & Landrigan, P. (2014). The developmental neurotoxicity of arsenic: cognitive and behavioral consequences of early life exposure. *Annals of Global Health*, *80*(4), 303–314. doi: 10.1016/j.aogh.2014.09.005

Tsai, S.-Y., Chou, H.-Y., The, H.-W., Chen, C.-M., & Chen, C.-J. (2003, August). The effects of chronic arsenic exposure from drinking water on the neurobehavioral development in adolescence. *Neurotoxicology*, *24*(4-5), 747–753. doi: 10.1016/S0161-813X(03)00029-9

Wang, S.-X., Wang, Z.-H., Cheng, X.-T., Li, J., Sang, Z.-P., Zhang, X.-D., ... Wang, Z.-Q. (2007, April). Arsenic and fluoride exposure in drinking water: children's IQ and growth in Shanyin county, Shanxi province, China. *Environmental Health Perspectives*, *115*(4), 643–647. doi: 10.1289/ehp.9270

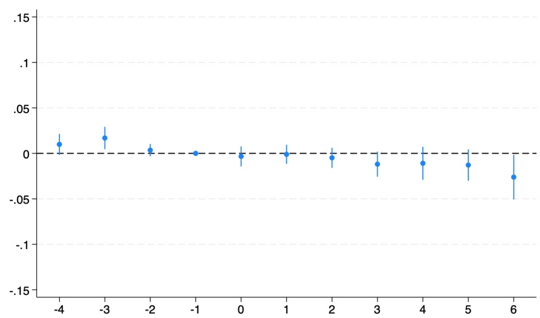
Wasserman, G. A., Liu, X., Parvez, F., Ahsan, H., Factor-Litvak, P., van Geen, A., ... Graziano, J. H. (2004, September). Water Arsenic Exposure and Children's Intellectual

Function in Araihazar, Bangladesh. *Environmental Health Perspectives*, *112*(13), 1329–1333. Retrieved 2024-01-11, from https://ehp.niehs.nih.gov/doi/full/10.1289/ehp.6964 (Publisher: Environmental Health Perspectives) doi: 10.1289/ehp.6964

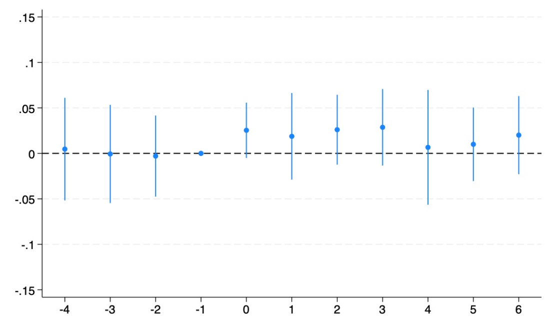
Wing, C., Freedman, S. M., & Hollingsworth, A. (2024). Stacked Difference-in-Differences. *NBER Working Papers 32054*. Retrieved 2024-06-23, from https://www.nber.org/system/files/working_papers/w32054/w32054.pdf

A Additional Tables and Figures

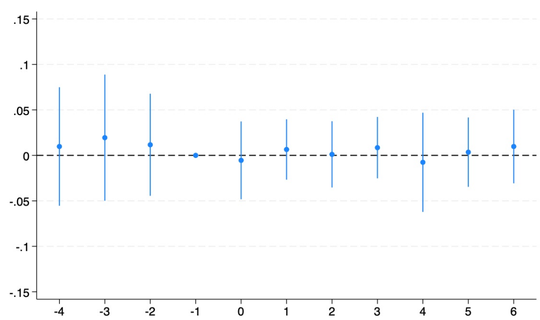
Figure A.1: Elementary Event Study Plots



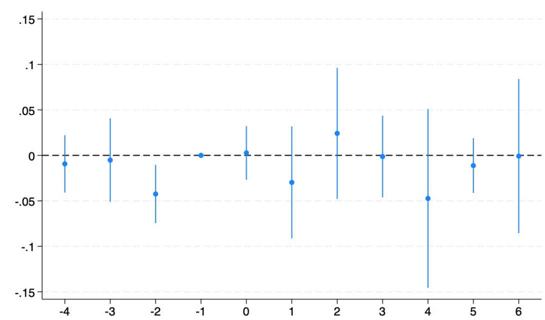
Panel A: Missed > 20 Days



Panel B: Missed < 5 Days



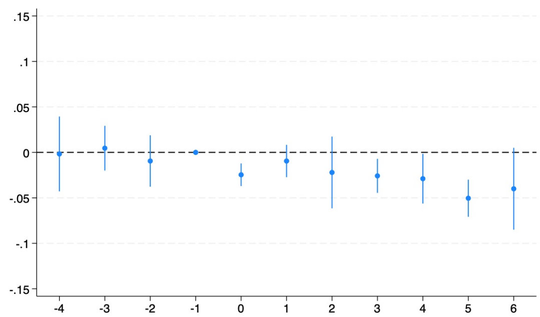
Panel C: Math Proficient or Better



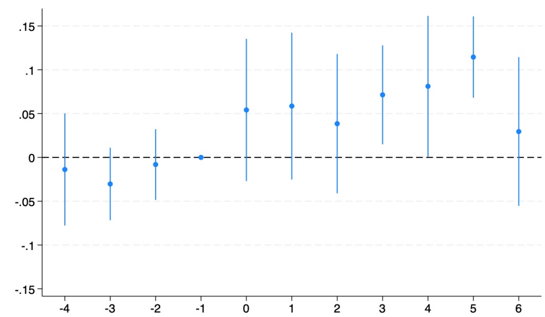
Panel D: Reading Proficient or Better

Note: These plots are based on unweighted stacking regressions from Equation (1), limited to elementary schools, with 2002-level controls for pupil-to-teacher ratio, percentage of students receiving free lunch, percentage of non-White students, and percentage of female students. All regressions include district-by-year, year-by-stack, and school-by-stack fixed effects. The horizontal axis measures years since treatment, and the vertical axis reports the percentage point change in the outcome of interest. Standard errors are clustered at the school level.

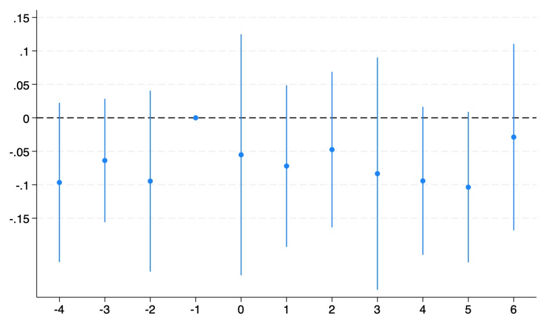
Figure A.2: Middle Event Study Plots



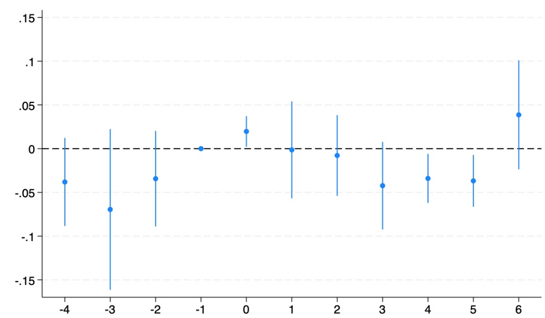
Panel A: Missed > 20 Days



Panel B: Missed < 5 Days



Panel C: Math Proficient or Better



Panel D: Reading Proficient or Better

Note: These plots are based on unweighted stacking regressions from Equation (1), limited to middle schools, with 2002-level controls for pupil-to-teacher ratio, percentage of students receiving free lunch, percentage of non-White students, and percentage of female students. All regressions include district-by-year, year-by-stack, and school-by-stack fixed effects. The horizontal axis measures years since treatment, and the vertical axis reports the percentage point change in the outcome of interest. Standard errors are clustered at the school level.

B Appendix B: Cohort Results

This section includes a detailed analysis of the secondary specification using cohorts. A cohort refers to a group of students within each school and start year. One main benefit of this specification is that it enables me to estimate students' length of exposure more accurately. By estimating how long students have been exposed to arsenic exceeding three ug/L, I can better understand cumulative impacts on math and reading proficiency. I explain how these cohorts are constructed, followed by specifications that mirror those presented within the paper at the school level.

B.1 Cohort Specification

Table B.1 provides a theoretical example of how these cohorts are built. For example, if a third-grade student's test scores at School A are observed in 2005, I then observe the same cohort of students' fourth-grade proficiency in 2006 and fifth-grade proficiency in 2007. These students would be considered cohort "3", which is a unique indicator of the cohort within the school.

One main benefit of this specification is that it enables me to estimate students' length of exposure more accurately. I assume the same students whose test scores I observe have attended school since kindergarten or sixth grade, depending on the school type. I extrapolate cohorts' kindergarten start year from the observed grade and test year and compare this to the schools' first year of treatment, with the difference providing the number of years exposed (if any).

Table B.1: Cohort Design

		Cohort ID									
School	Year	1	2	3	4	5	6	7	8	9	10
School A											
	2005	5	4	3	2	1					
	2006		5	4	3	2					
	2007			5	4	3					
School B											
	2005						5	4	3	2	1
	2006							5	4	3	2
	2007								5	4	3

To investigate the effects of exposure, I generate stacks by year of treatment h, mirroring the stacking methodology of the school-level specification. Control cohorts remain those who do not begin treatment in the 11-year window ($t = -4, .., 6$) around treatment year h and include cohorts from all observed schools. The following specification is used to explore heterogeneous effects by length of exposure across math and reading test scores, averaged across treatment years:

$$\begin{aligned}\text{Proficient}_{csdth} = \beta_1 \text{Arsenic}_s \times \text{Years}_{csth} + \beta_2 \text{Arsenic}_s + \beta_3 \text{Years}_{csth} \\ + X_{st_{2002}} \cdot t + \delta_{dh} \cdot t + \delta_{ch} + \delta_{th} + \epsilon_{csdth}\end{aligned} \tag{3}$$

Here, $\text{Proficient}_{csdth}$ identifies the percentage of students in cohort c, within school s in district d in year t who are proficient or better in math or reading. Arsenic_s is equal to one if arsenic exposure at school s is greater or equal to three ug/L and is determined by the concentration of arsenic measured in the closest geographical groundwater well. Years_{csth} is a continuous measure of the total years a student in cohort c is exposed to untreated water by year t. This measure is the difference between a student's start year and the year water treatment begins in school s, or the current year, whichever occurs first.

The coefficient on the interaction between these Arsenic_s and Years_{csth}, β_1, measures the impact of each additional year of exposure to three ug/L or more of arsenic over the student's tenure and provides an understanding of the impact of arsenic over time. The controls in this specification, $X_{st_{2002}} \cdot t$, are the same as those in the school-level specification. District-by-year fixed effects, $\delta_{dh} \cdot t$, are included to mitigate concerns of endogeneity of treatment timing. Lastly, Cohort-by-stack (δ_{ch}) and year-by-stack (δ_{th}) fixed effects are included, and standard errors are clustered at the school level.

The continuity of treatment in this specification inhibits the use of event studies to test parallel trends. However, I assume that because cohorts are nested inside of the school level if parallel trends are satisfied for the school-level models, it is reasonable to assume they are satisfied in the cohort regression.

B.2 Cohort Results

While the school results suggest the potential for arsenic exposure to negatively impact students' academic performance, the following section will expand these results at the cohort level to gain more precision with cohort results from equation (2).

The findings for math proficiency outcomes are consistent across the school and cohort specifications, suggesting that arsenic exposure may negatively impact math and reading proficiency. The coefficients in Table B.3 can be interpreted as the cumulative effect of each additional year of exposure to arsenic above three ug/L on the percent of students scoring proficient or better in math and reading. Compared to the results from the school-level

Table B.2: Cohort Level Summary Statistics

	Never-Treated	Treated	Always Treated	No Arsenic
As Concentration	4.899	6.100	4.572	0.000
	(3.380)	(15.568)	(12.264)	(0.000)
As Avg of 2 Closest	4.368	5.173	4.504	0.180
	(2.839)	(8.494)	(9.427)	(2.381)
As Max of 2 Closest	5.020	7.507	6.437	0.359
	(3.411)	(15.796)	(16.264)	(4.761)
% Prof or Better Math	0.798	0.798	0.793	0.760
	(0.160)	(0.155)	(0.152)	(0.170)
% Prof or Better Reading	0.836	0.835	0.828	0.800
	(0.103)	(0.114)	(0.116)	(0.137)
Enrollment	513.935	388.831	599.191	592.570
	(209.065)	(168.409)	(238.066)	(229.485)
Pupil to Teacher Ratio	16.711	14.747	16.977	16.085
	(2.181)	(2.186)	(2.354)	(2.968)
% Female	0.497	0.482	0.486	0.485
	(0.026)	(0.026)	(0.023)	(0.026)
% Non-White	0.149	0.161	0.268	0.492
	(0.112)	(0.125)	(0.232)	(0.341)
% Receiving Free/Reduced Lunch	0.215	0.279	0.256	0.331
	(0.116)	(0.149)	(0.186)	(0.230)
Years Exposed	5.847	4.548	0.096	0.755
	(1.299)	(2.152)	(0.576)	(1.927)
N	355	450	1,484	7,784

Treated and never-treated are both limited to schools with arsenic exposure greater than 0.

analysis, cohort results suggest the effect on test scores could be larger in magnitude. For math proficiency, while not significant, the results suggest a 0.9 percentage point decrease in math proficiency for each year exposed to arsenic contamination above three ug/L. For a third-grade student, this translates to a 3.6 percentage point decrease, on a mean proficiency of 80.4%. On average, students in treated schools are exposed for 4.5 years, accumulating to as much as a 4.05 percentage point difference between treated and control schools.

Another similarity to the school-level results is the smaller effect of arsenic contamination on reading proficiency. The same exposure would result in a 0.2 percentage point decrease in test scores per year exposed, aggregating to a 0.9 percentage point difference on a mean of 83.6% scoring proficient or better in reading. These results follow similar trends to Hollingsworth et al. (2025) and Jacqz (2022), as mentioned in the school-level results.

Overall, my results are difficult to compare to other pollution literature on students' test score outcomes because they use individual-level data and provide results as a standard deviation change in test scores. The most similar and comparable result I find is from Aizer et al. (2018). This paper found that a ”one-unit decrease in average blood lead levels reduces the probability of being substantially below proficiency in reading (math) by 0.96 (0.79) percentage points on a baseline of 12 (16) percent”. While lead and arsenic may not be directly comparable, my results are much smaller in magnitude, with each marginal ug/L per year of arsenic exposure leading to only a 0.09 percentage point decrease in the percent of students scoring proficient or better in math on a mean of 81.7%.

Table B.3: Cohort Results

	Math Proficiency	Reading Proficiency
Years Exposed	-0.004	0.000
	(0.003)	(0.003)
Arsenic $\geq$3 $\times$ Years	-0.004	-0.008
	(0.014)	(0.013)
N	69,808	69,788

Notes: Standard errors clustered at the school level are shown in parentheses. Estimates are from the unweighted stacking regression from Equation (3) with 2002-level controls for pupil-to-teacher ratio, percentage of students receiving free lunch, percentage of non-White students, and percentage of female students. All regressions include district-by-year, year-by-stack, and cohort-by-stack fixed effects. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

B.3 Heterogeneity by School Type

When estimating equation (3) separately by school type, the cohort level regressions produce similar outputs between elementary schools and the “all school” category, suggesting that elementary school results drive the aggregate result within the cohort regression. For both math and reading proficiency, the middle school-only regression produces significant results at the 1% level that differ dramatically from the theoretical result and elementary and aggregate results. The results suggest that each additional year of arsenic exposure increases proficiency by 3.7 percentage points or 5.0 percentage points for math and reading, respectively (Table B.4). The likely reason for the dramatic difference in results is that they are driven by few schools that may be experiencing some other shock simultaneously that is not controlled for, as suggested by the drastically fewer observations for middle school regressions. These results may be significant but should be cautiously interpreted due to the limited number of schools.

Table B.4: Cohort Results by School Type

	Math Proficiency			Reading Proficiency		
	All	Elem	Middle	All	Elem	Middle
Years	-0.004	0.001	-0.014	0.000	0.002	-0.003
	(0.003)	(0.003)	(0.010)	(0.003)	(0.002)	(0.006)
As$\geq$3$\times$Years	-0.004	-0.021	0.005	-0.008	-0.024	0.012***
	(0.014)	(0.020)	(0.007)	(0.013)	(0.016)	(0.004)
N	69,808	51,093	15,020	69,788	51,073	15,020

Notes: Standard errors clustered at the school level are shown in parentheses. Estimates are from the *unweighted* stacking regression from Equation (3) with 2002-level controls for pupil-to-teacher ratio, percentage of students receiving free lunch, percentage of non-White students, and percentage of female students. All regressions include district-by-year, year-by-stack, and cohort-by-stack fixed effects. The sample was split to retain only elementary or middle schools and run separately. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

B.4 Cohort Results by Arsenic Measure

To test the robustness of my estimates on the decision of which wells to use to generate the variable for potential exposure, I provide estimates using well concentrations from the closest well, the average of the two closest wells, and the maximum between the two closest wells (Table B.5). For cohort math and reading proficiency, results are similar between the closest well and max, the two closest wells. However, when using the average of two wells to produce the treatment dummy for arsenic exposure greater or equal to three ug/L, results become positive. These results would suggest that test scores improve each additional year that they are exposed to arsenic contamination. However, it is also notable that all results are small in magnitude and significantly insignificant, which could be a result of noise in the arsenic measure. Lastly, arsenic exposure is observed at the school level and not the cohort level directly, limiting the variation in these estimates further. Comparatively, school-level results were similar, but the sign was consistent across all specifications.

Table B.5: Cohort Results by Arsenic Measure

	Math Proficiency	Reading Proficiency
Panel A: Closest Well		
Years Exposed	-0.004	0.000
	(0.003)	(0.003)
Arsenic $\geq$3$\times$Years	-0.004	-0.008
	(0.014)	(0.013)
Panel B: Avg of 2 Closest Wells		
Years Exposed	-0.004	0.000
	(0.003)	(0.003)
Arsenic $\geq$3$\times$Years	0.010	0.006
	(0.010)	(0.007)
Panel C: Max of 2 Closest Wells		
Years Exposed	-0.004	0.000
	(0.003)	(0.003)
Arsenic $\geq$3$\times$Years	-0.004	-0.008
	(0.014)	(0.013)
N	69,808	69,788

Notes: Standard errors clustered at the school level are shown in parentheses. Estimates are from the unweighted stacking regression of Equation (3) with 2002-level controls for pupil-to-teacher ratio, percentage of students receiving free lunch, percentage of non-White students, and percentage of female students. All regressions include district-by-year, year-by-stack, and cohort-by-stack fixed effects. Three separate methods are used to determine arsenic concentration and associated treatment. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

C Appendix C: Weighted Stacking Specification

Wing et al. (2024) suggest that traditional stacking methods, as originally by Cengiz et al. (2019) and used as the baseline for this paper, may be biased if the number of control observations varies between treatment timings. To control for this bias, Wing et al. (2024) suggest weighting stacked regressions by the number of treated observations in each substack. My results are consistent across methods. Cohort results are the same for the three decimal places displayed, but they differ when the results are extended to five decimal places.

Table C.1: Cohort Regression Results, Unweighted and Weighted Specifications

	Math Proficiency		Reading Proficiency	
	(1)	(2)	(3)	(4)
Years Exposed	-0.004	-0.003	0.000	0.000
	(0.003)	(0.003)	(0.003)	(0.002)
As$\geq$3 $\times$ Years Exposed	-0.004	-0.003	-0.008	-0.008
	(0.014)	(0.015)	(0.013)	(0.013)
N	69,808	69,808	69,788	69,788
Weights	0	1	0	1

Notes: Standard errors clustered at the school level are in parentheses. Each column corresponds to a separate stacked cohort regression of Equation (3). Models include 2002-level controls for pupil-to-teacher ratio, percentage of students receiving free/reduced lunch, percentage of non-White students, and percentage of female students. All regressions include district-by-year, year-by-stack, and cohort-by-stack fixed effects. Columns (1) and (3) are unweighted; Columns (2) and (4) use observation weights following Wing et al. (2024). Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.2: Unweighted and Weighted School Results

	Math Prof		Math Adv		Reading Prof		Reading Adv		Miss > 20 Days		Miss < 5 Days	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Arsenic$\geq 3\times$Post	0.021	0.021	0.016	0.015	0.021	0.021	0.005	0.005	-0.015*	-0.015*	0.019	0.018
	(0.023)	(0.023)	(0.027)	(0.027)	(0.023)	(0.023)	(0.015)	(0.015)	(0.008)	(0.008)	(0.013)	(0.013)
N	59,280	47,056	59,280	47,056	59,306	47,076	59,306	47,076	87,412	67,687	87,412	67,687
Weights	0	1	0	1	0	1	0	1	0	1	0	1

Notes: Standard errors clustered at the school level are shown in parentheses. Estimates in odd columns are from the unweighted specification by Cengiz et al. (2019), and estimates in even columns are from the weighted specification suggested in Wing et al. (2024), based on Equation (2) with 2002-level controls for pupil-to-teacher ratio, percentage of students receiving free lunch, percentage of non-White students, and percentage of female students. All regressions include district-by-year, year-by-stack, and school-by-stack fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.