The Effects of Exposure to Waterborne Lead on AFQT Scores

Joseph Ferrie
Northwestern University

Bhashkar Mazumder
UC Irvine and NBER

Karl Schulze
Yale University *

November 3, 2026

Abstract

Lead pipes are much more likely to contaminate drinking water at both very high (low acidity) and very low (high acidity) levels of pH. We exploit differences in water chemistry across cities with historical lead piping to identify the causal effects of waterborne lead exposure on cognitive ability. Combining historical piping and pH measurement data with the AFQT scores of the universe of applicants to the US military from 1976 to 1991, we find that childhood exposure to lead pipes in cities with extreme pH levels lowers AFQT scores by at least 1.3 percentiles. By contrast, differences in water's pH levels have no effect where pipes are not made of lead. Lead exposure corresponds to reductions in the present value of lifetime earnings of about $14,000. Under conservative assumptions, a policy that removes lead service lines in cities with extreme pH would be highly cost effective, with a marginal value of public funds (MVPF) of 2.8.

*Mazumder received funding from the National Institute on Aging of the National Institutes of Health award R01AG079914 .

1 Introduction

Exposure to lead through drinking water is a salient public policy issue in the United States. Lead is known to cause a multitude of health problems, and infants and children are especially vulnerable. The Environmental Protection Agency (EPA) notes that childhood exposure to lead is associated with behavior and learning problems, lower IQ and hyperactivity, slowed growth, hearing problems, and anemia [U.S. Environmental Protection Agency, 2026b]. Among adults, the EPA concludes that lead is associated with cardiovascular disease, hypertension, decreased kidney function, and reproductive problems. Public policy has largely addressed two of the main vectors for lead transmission: lead in residential paint was banned in 1978 and lead in gasoline was fully banned by 1996. The remaining vector for lead, waterborne lead, remains a national problem. In 2024, the EPA estimated that 9 million U.S. households still had lead service lines (LSLs) [U.S. Environmental Protection Agency, 2024a].[1] The immense social costs associated with lead prompted the Biden administration to publish a rule in 2024 requiring that virtually all drinking water pipes containing lead be eliminated in a decade [Federal Register, 2024], a project expected to cost tens of billions of dollars.

A key question is whether the economic benefits of large-scale water infrastructure remediation projects would be commensurate with its costs. We contribute to this discussion by studying the benefits of reducing waterborne lead exposure on one critical outcome, cognitive ability. Impaired cognitive ability can impact many other critical downstream outcomes such as health, income, and incarceration. Identifying the *causal* effects of lead exposure is challenging and there are only a few longitudinal studies tracing the impact of childhood lead exposure on later-life cognition. There are two key challenges in such studies. First, it requires having a sample of sufficient size that connects measures of cognition to early-life lead exposure.[2] Second, it can be difficult to identify causal relationships between exposure and outcomes in settings where either other influences experienced later in life can interact with lead exposure to impact cognitive ability, or, where there are multiple exposure vectors (e.g. leaded gasoline, lead-based paint, lead in water) of varying relative intensities over time or across space.

Our analysis uses a dataset extremely well suited to these challenges. We use the Armed Forces Qualification Test (AFQT) scores of all applicants to the US military from 1976 to 1991. Our estimation sample includes between 500,000 and 1 million individuals, the vast majority of whom are still alive today. Using AFQT scores is valuable for several reasons. First, it is a highly validated measure of cognitive ability that captures elements of crystallized and fluid intelligence as well as executive functioning [Hambrick et al., 2023]. Second, there is a large economics literature highlighting AFQT's effects on individual's earnings (e.g. Neal and Johnson [1996], Heckman et al. [2006]). This allows us to quantify the economic

[1] The EPA dramatically reduced the estimate of LSLs in 2025 to 4 million, but the National Resources Defense Council is skeptical that this is accurate, they estimate that there are between 9 and 12 million LSLs. [NRDC, 2025].

[2] The Dunedin Study [Poulton et al., 2015] is exceptional in that it has followed individuals born 1972-73 for more than 50 years. The ongoing study contains data on lead exposure, BLLs, and cognitive capacity over the lives of 95% of the original enrollees. However the sample size is only 1037 individuals

impact of exposure to lead on individuals through one important channel.

Our research design is also well suited for isolating the effects of exposure to waterborne lead by exploiting plausibly idiosyncratic variation in lead exposure during childhood that arises from differences in water chemistry across areas that use lead piping [Ferrie et al., 2012]. Specifically, the likelihood that lead contaminates drinking water is a non-monotonic function of the pH level of the water. When the pH level is very low (high acidity), lead begins to dissolve and when the pH level is very high (low acidity), lead begins to flake. Thus, our main comparisons use variation across cities that historically had lead pipes but had different pH levels. Importantly, for identification, we show that acidity differences are not associated with test score differences when pipes *do not* contain lead. Differences in pH are also plausibly unrelated to other vectors of lead exposure such as leaded gas or lead paint. To address any other potential confounding factors, we use inverse propensity score weighting and additional observable controls.

Since we are unable to observe the actual lead exposure experienced by our sample of test-takers, our approach is necessarily an "intent to treat" (ITT) design that captures the reduced form relationship between test scores and lead piping interacted with extreme pH. However, as we discuss below, we bolster the credibility of our design by showing a striking "first-stage" relationship between the interaction of lead pipes and extreme pH on modern day rates of extreme blood lead levels (BLLs) of children under the age of three obtained from the Centers for Disease Control and Prevention (CDC) for 27 states.

We also use a second design that compares the effect of extreme pH on individuals who grew up in cities with lead pipes compared to those who grew up in cities with other piping materials. Here we also use propensity score adjustment and include controls to deal with possible differences in the composition of cities with lead piping versus those without. Using either approach, we find statistically large and economically meaningful effects of lead exposure. Growing up in a city that used lead pipes and had "extreme pH" levels (<= 7 or >= 8), results in AFQT scores that are reduced by between 1.3 and 1.8 percentiles. This compares to a mean AFQT percentile score of about 41 in cities that had lead piping. This also represents an effect of between roughly 0.05 to 0.08 standard deviations.[3] These effects are substantially larger (between 2.4 and 5 percentiles or 0.1 to 0.2 standard deviations) if we restrict our attention to a smaller set of cities where we are most confident in our pH measures. These effects may be further attenuated due to the fact that we do not observe the location of test-takers when they are children inducing measurement error into our assignment of childhood lead exposure.[4]

We use the National Longitudinal Survey of Youth (NLSY79) to translate these test score gains into economic benefits. Our effects imply that eliminating an individual's exposure to waterborne lead would lead to an increase in lifetime earnings of about $59,000 in 2025 dollars, or about $14,000 in present value terms when discounted to birth. By way of comparison, this present value impact is more than twice as high as the effects of reducing air pollution exposure early in life on lifetime earnings due to

[3]The standard deviation of AFQT percentiles is about 24 for cities with lead piping.
[4]We assess the degree of migration and its potential downward bias on our results in section 5.4.

the 1970 Clean Air Act [Isen et al., 2017].[5] Based on prior work, we estimate that the cost of removing waterborne lead per child, including current and future cohorts affected, is in the range of $2,000 to $7,000. Our estimate of the marginal value of public funds (MVPF) [Hendren and Sprung-Keyser, 2020] using conservative assumptions is 2.8. Importantly, this estimate focuses only on private impacts and is likely an underestimate, as it does not incorporate the further individual and societal benefits from improved health and reduced crime [Feigenbaum and Muller, 2016, Aizer and Currie, 2019]. Thus, a policy aimed at eliminating LSLs in cities with extreme pH appears to be highly cost effective.

We make several contributions to the literature. First, we focus on a critical outcome that is a strong marker for lead exposure, namely cognitive skills, and its relationship with waterborne lead exposure, a contemporary source of much policy discussion . This complements prior work focusing on test scores in school [Aizer et al., 2018] or differential exposure from the removal of leaded gasoline [Grönqvist et al., 2020]. Second, our outcome is measured at a relatively early point in the life-cycle, when there are fewer compounding effects of later influences that might interact with lead exposure. This complements results from Fletcher and Noghanibehambari [2025], who focus on the mortality effects of lead exposure, which can pick up a vast array of channels that come into play later in the life course. Third, unlike much of the literature using relatively small samples to link lead poisoning to IQ, we use samples of between 500,000 and 1 million applicants to the military. Fourth, by using modern day AFQT scores, we are able to quantify a large portion of the economic costs of early life lead exposure and demonstrate the cost effectiveness of a remediation program to remove lead pipes from a subset of cities with extreme pH levels. This is highly relevant given the recent attention focused on policies to remove lead piping in the US.

The work most similar to ours is by Ferrie et al. [2012] who showed a nonlinear relationship between water acidity and AGCT scores of about 5,000 WWII soldiers who enlisted in cities with lead pipes.[6] In contrast to that study, we use vastly larger samples on the universe of test takers over a 16 year period, most of whom are still alive today. With the larger samples we are able to implement a wide array of robustness checks including using a smaller sample of cities where we are most confident in pH measurements. We also use a modern and highly validated test score that has been directly linked to earnings. We also extend the analysis to comparisons to cities without lead pipes. Notably, we have also significantly increased the sample of cities with piping data since Ferrie et al. [2012].

The rest of the paper proceeds as follows. In section 2 we present background material on lead exposure and its effects on cognitive ability, how pH affects lead solubility and flaking in water, and the historical development of lead piping across American cities. Section 3 describes the data we use in our analysis. In section 4 we describe our empirical strategy. Section 5 presents our results, section 6

[5] Isen et al. [2017] estimate that the Clean Air Act would increase lifetime earnings by $4,300 in 2008 dollars which would be roughly $6500 in 2025 dollars. Their estimate is a present value at birth computed with a 3 percent real discount rate, so our numbers are comparable.

[6] Lee et al. [2022] also use the nonlinearity of water chemistry and lead exposure during childhood to identify effects on cognitive ability for about 1100 individuals in the Health and Retirement Survey linked to their location in adolescence in the 1940 Census.

shows our cost-benefit calculations, and section 7 concludes.

2 Background

2.1 Lead and its Effects on Cognitive Ability

By the beginning of the twentieth century, lead was ubiquitous in US daily life. Lead was used in gasoline, paint, tin cans, public drinking fountains, toys, car batteries, cosmetics, toothpaste tubes, lead smelters and household plumbing fixtures [Warren, 2001]. Lead has now been removed from virtually all of these applications except the last, household plumbing where exposure arises from lead service lines (LSLs) carrying water from street mains into dwellings, lead solder joining copper, iron, or steel interior piping, and brass fixtures. The EPA did not ban lead service lines until 1986, and even then it only did so for new construction. The EPA's ban applied only to post-1986 additions to the housing stock and did not require the removal of existing lead service lines. Even after the EPA's ban was promulgated in 1986, some states did not pass legislation to effectuate it until 1991.[7] Only two states (Maine in 1926 and Oregon in 1984) had banned lead service lines before 1986. At the local level, model plumbing codes, which were the basis for municipal plumbing codes, continued to allow lead through 1988.[8] As of 2025, 9.2 million households, or about 7 percent of all US households still had LSLs [U.S. Environmental Protection Agency, 2026a].

The toxicity of lead to the human body at high doses has been known for more than 2,000 years.[9] The deleterious effect of lead exposure on the cognitive capacity of children who exhibited clinical signs of encephalopathy after acute exposure (i.e. resulting in blood lead levels or BLLs > 60 μg/dl) was first recognized more than eighty years ago [Byers and Lord, 1943]. As a result, through the 1960s, the definition of a "clinical" level of exposure was presentation with a BLL above this threshold [Needleman, 2004, p. 213]. Subsequent research identified cognitive deficits even in children with substantially lower BLLs, leading the Centers for Disease Control and Prevention (CDC) to begin to lower its "acceptable" childhood BLL in stages, until it reached 10 μg/dl by the early 2000s [Needleman, 2004, p. 214]. By 2010, it was widely recognized that even this level was too high and that in fact there is no measurable level of exposure at which cognitive deficits do not result [Grandjean, 2010].

2.2 Plumbosolvency as a Research design

One approach that has been adopted to deal with these challenges is using data that connects individuals' post-childhood outcomes to their place of residence early in life, determining the material use in LSLs at those locations, and leveraging the well-established finding in water engineering that the "plum-

[7] Appendix D in U.S. EPA, "Guidance for Developing and Maintaining a Service Line Inventory," Office of Water (4606M), EPA 816-B-22-001, August 2022.

[8] Devine, Christina, Luis Garcia-Bakarich, William Platten, Krista Jajeh, and Simoni Triantafyllidou. "Review of historical plumbing codes for lead service line inventories in the United States." *Water Practice & Technology* 19, no. 9 (2024): 3867–3881.

[9] The first known description of the effects of lead exposure was provided in the second century BCE by the Greek physician and poet Nicander, who described "a frothing mouth, asperity of the tongue, and dry throat, together with dry retching, chills, delusions, and overwhelming fatigue." [Montgomery et al., 2010, p. 199] [Grout, 2008, p. 11-14]

bosolvency" of LSLs, the quantity of lead that leaches into the water it carries per LSL unit of interior surface area, varies non-linearly with the pH of the water being transported through them.[10] The LSL-pH interaction as a proxy for waterborne lead exposure is essentially orthogonal to the two other major exposure vectors, leaded gasoline and lead-based paint. Furthermore, because this interaction is unknown to households and therefore likely unrelated to their residential locations, the relationship between the LSL-pH interaction experienced early in life and subsequent outcomes can plausibly be interpreted as causal. In our analysis below we present evidence in support of this assumption.

In Figure 1 we use data from 2008 to 2012 on elevated blood lead levels (BLLs) for children under the age of 3 from the CDC for 27 states to illustrate the nonlinear relationship between historical pH levels in cities with LSLs and lead exposure in young children. Elevated BLLs are defined as those in excess of 10μg/dl. The relationship between pH and elevated BLLs is estimated using kernel smoothing. The dashed line in the figure, representing cities without any lead in their piping, shows that these cities had incidence rates of elevated blood levels of just 2 per 10,000 and that these rates are fairly stable at all levels of pH. In contrast cities with lead piping (solid line) have rates of elevated BLLs of about 3.5 to 4 when pH is less than 7 and above 3 when pH is between 8 and 8.5.[11] However, when pH is between 7 and 8 the rates of elevated BLLs are more similar between cities with and without lead. This non-linear relationship, which is only observed in cities with LSLs, is precisely what we exploit in our identification strategy. Specifically, we will define extreme pH as levels <=7 or >=8.[12] This is the first direct confirmation of a first-stage relationship between LSLs-pH and elevated BLLs in the literature of which we are aware.[13]

While Figure 1 suggests that there are no meaningful differences in elevated BLLs by extreme pH in cities without lead pipes, we also directly analyze which pre-existing characteristics of cities in 1900 may be associated with having an extreme pH. We do this by estimating propensity score models on our sample of cities with lead piping.[14] We then adjust for these differences using inverse propensity score weights (IPW) in some specifications. Specifically, we find that among cities with lead pipes, extreme pH tends to be more common in urban areas with a lower share of Black individuals, lower fraction with mortgages, and higher fertility rates (see column 3 of Appendix Table A2). Adjusting for these observable differences results in complete balance on these observed covariates, with a joint test for balance resulting in an F-statistic of 0.7 (see Appendix Table A3). Appendix Figure A4 also shows a considerable degree of overlap of the distribution of p-scores between locations with and without extreme pH.

[10] The nonlinear relationship between pH and plumbosolvency—which reaches a minimum near pH 7.5 and increases as pH moves away from this point—is documented in Moore [1973], Schock and Gardels [1983] and Schock [1990]. Also see papers cited in Ferrie et al. [2012] on page 101.

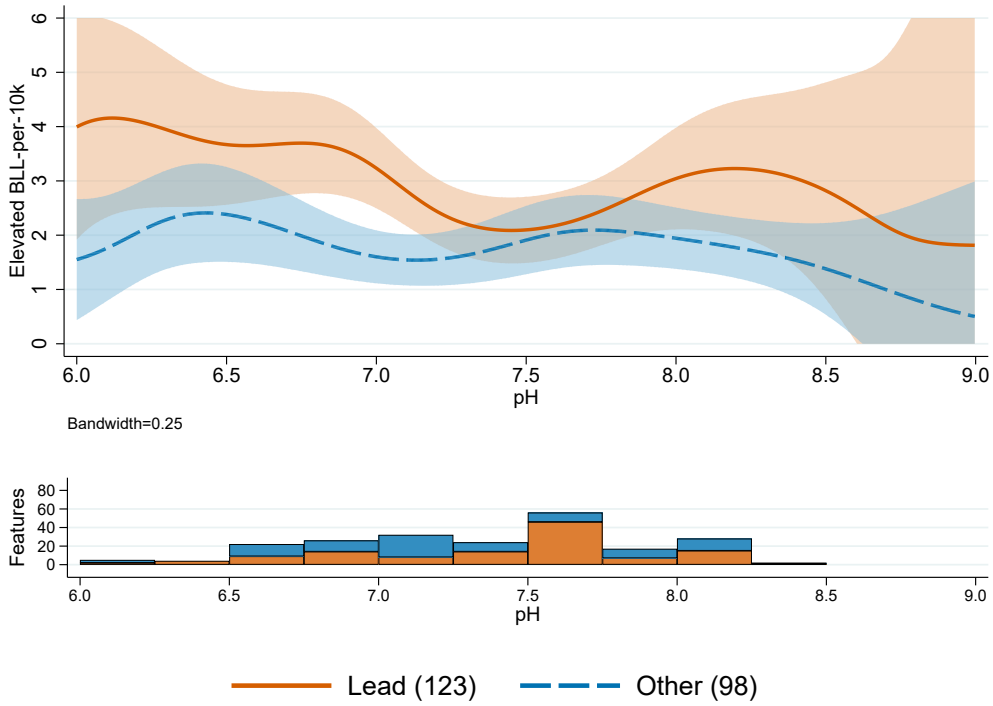
[11] There are virtually no cities in our data that have pH above 8.5.

[12] We also assess the sensitivity of our results by using alternate extreme-pH definitions, predictions from polynomial controls of pH to varying degrees, as well as non-parametric smoothing of test score residuals over pH. These approaches provide results consistent with this range. Results using more restrictive definitions of extreme pH result in large reductions in the sample. For example, there are only 19 cities with lead piping that have pH levels of ≤ 6.75 or ≥ 8.25. Cutting the sample further using these definitions makes estimates noisy and unreliable.

[13] Earlier work by Sherlock et al. [1982] show a linear dose-response relationship between lead contained in water and blood lead levels in individuals who ingest the water.

[14] See section A.3 for a more detailed discussion of our procedure.

Figure 1: Relationship between number of elevated blood lead levels (2008-2012) and pH



Notes: This figure shows the non-parametric relationship of the normalized number of confirmed cases of blood lead levels over $10\mu g/dL$ for children under 3 in a county against a 1990 census place's pH-level, as used in our analysis. Number of census places used are in the legend in parentheses. pH readings of under 6 and over 9 are used in estimation, but excluded from the display. Data on elevated blood lead levels comes from the CDC (https://www.cdc.gov/nceh/lead/data/state.htm) for all states reporting data for 2008 to 2012. AK, AR, HI, ID, MT, ND, NE, NV, SC, SD, UT, and WY do not report data. CO, DC, LA, ME, MD, NM, TN, TX, and VA are not included due to missing data. We digitize state reports from CA for ages less than 6 and blood lead levels of over $9.5\mu g/dL$, the inclusion of which has little effect on the results. This leaves 30 states, 27 of which have lead or other piping data. Included states may differ in reporting procedures. Counts of confirmed cases are averaged over 2008 and 2012 and normalized by the ACS 2012 5-year estimates of the population under 5 years old. Smoothing uses a Gaussian kernel and a bandwidth of 0.25 points of pH.

Ferrie et al. [2012] was the first study to use identification based on the nonlinear relationship between water pH and lead exposure. They link a small subset of World War II U.S. Army enlistees (n = 5503) to their place of birth, and examine their results on the Army General Classification Test (AGCT) – the forerunner of the Armed Forces Qualification Test (AFQT) examined in the present study – as a function of service line material, water pH, and household and individual-level controls.[15] The results show the anticipated hump-shape in the relationship between pH levels and the AGCT score in places with LSLs but not otherwise. In subsequent work, Lee et al. [2022] use 1,089 respondents in the 1998-2016 Health and Retirement Study (HRS) linked to their 1940 location in the U.S. Census of Population. Using similar data on service line material and water pH, they show that cognitive deficiencies are already present among those above the age of 60 and residing in 1940 in places with both LSLs and extreme water pH.

Most recently, Fletcher and Noghanibehambari [2025] use an instrumental variables (IV) strategy also exploiting the interaction of LSL and extreme pH. They focus on World War II U.S. Army enlistees who were under the age of 26 when they took the AGCT and also study mortality. However, their approach yields high standard errors that preclude reliable inference.

In contrast to these prior studies, we utilize a vastly larger sample of about a million individuals comprising the universe of all applicants to the US military between 1976 and 1991. This enables us to construct a large sample of individuals, many of whom are alive today, allowing us to precisely estimate the effect of LSLs and extreme pH on cognitive ability. We also have a much larger sample of cities for which we know the water chemistry.[16] Finally, by using AFQT scores, we can further translate these effects into lifetime earnings losses.

2.3 Choice of piping type in American cities

Municipal waterworks grew extremely rapidly in the US during the second half of the 19th century with a ten fold increase from 1870 to 1900 [Feigenbaum and Muller, 2016]. The main choice faced by cities at this time was between wrought iron, galvanized iron, or lead. Based on the 1897 Manual of American Waterworks a majority of the 545 cities surveyed used at least some lead in their water systems [Feigenbaum and Muller, 2016]. Using this data, Feigenbaum and Muller [2016] found that larger, denser and more highly educated cities were more likely to use some lead. This makes sense given that lead pipes were costlier but higher quality, being far more durable and malleable.

In our own data, which we describe in greater detail in section 3, we are able to delineate whether cities used *only* lead ("lead"), had a combination of lead and other materials ("mixed") or used entirely other materials such as iron ("other"). The geographic breakdown of the cities by piping type is shown below in Figure 2. The figure suggests that the use of lead was not especially geographically concentrated

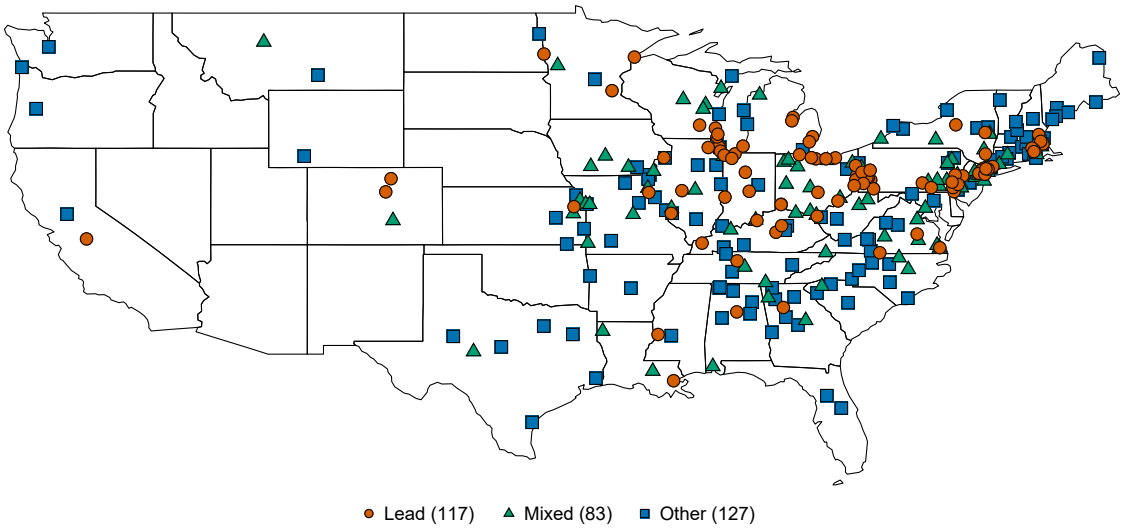
[15] Even though individual level AGCT scores were thought to be lost to history, Ferrie et al. [2012] exploit the fact that for a narrow period of time from March to June of 1943, data entry personnel were instructed to enter AGCT scores in the field for weight in punch card records at enlistment sites. Hence, the relatively small sample available for analysis.

[16] For example, Ferrie et al. [2012] only had data on the water chemistry for 100 cities, whereas we have pH on 1000 cities and towns as we discuss below

compared to other piping types, though there is much less use of lead in Southern cities.

For some of our analysis where we compare the effects of extreme pH in cities with lead piping to cities with other kinds of piping materials, we need to address potential selection into the choice of using lead. We do this by using inverse propensity score weighting (IPW), where we first predict a city's use of lead piping based on 1900 characteristics and then adjust for these probabilities with weights.[17] We find the likelihood of having lead pipes is higher in cities with a higher share of Black individuals, higher literacy rates, higher non-farm occupational incomes, greater rates of having a mortgage, higher schooling rates and lower manufacturing wages and establishments (see Appendix Table A2). With the exception of school enrollment, we see no significant differences between cities with and without lead pipes after IPW adjustment, and a joint test of balance results in an F-statistic of 1.1 (see Appendix Table A4). In Appendix Figure A3 we show the distribution of each piping type by pH level (and race) and do not see any concerning patterns. Appendix Figure A4 also shows a considerable degree of overlap of the distribution of p-scores between locations with and without lead piping.

Figure 2: Map of pH readings by piping type



Notes: Includes pH reliability groups 1 to 2.

3 Data

3.1 Data Sources on Piping

The raw water sources for U.S. cities and towns and the materials used for their water service lines (the portion of the system leading from the water main in the street into each dwelling) were identified primarily through *The Manual of American Water-Works*, edited by Moses Baker and published by *The*

[17] See section A.3 for details.

Engineering News four times (1888, 1889–1890, 1893, 1897).[18][19] This was intended as an annual and comprehensive directory of every public and private water system in the country and was the result of direct correspondence with each of more than 1,400 utilities.

The *Manual* records the finances and operating characteristics of each system. It lists the water sources (specific rivers, streams, lakes, reservoirs, and wells) for each system as well as any methods of water treatment employed. In each volume except the third (1893), it also contains an entry for "Services" recording the material used for each system's service lines (e.g., "lead," "wrought iron").

The *Manual* ceased publication after the 1897 volume, so its reports on sources, treatment, and service line materials are up-to-date only through that year. For years after 1897, we supplemented the information in the *Manual* in three ways:

1. As new systems came online, they were listed in *The Engineering News*, which has continued publication under various names to the present;
2. The journal *Public Works* undertook its own survey of water service materials and raw water sources in more than 800 cities and towns in 1922;
3. Massachusetts conducted surveys of water systems throughout that state (1898–1906) and reported the service line materials for 74 cities and towns, while New Hampshire periodically reported the measured lead content of tap water (1915–1932) for all cities and towns.

See Appendix section A.1 for further details on water piping data and how we prioritized information for the different sources.

3.2 Data Sources on Water Chemistry

We obtained data on the pH of the water source of each city and town from periodic reports from the U.S. Geological Survey (USGS) assessments of the suitability of public water supplies in the U.S. for industrial uses. See Appendix section A.2 for further detail. The first of these appeared in 1923 and 1932 and covered the largest cities. The most comprehensive was published in two volumes for 1952, covering every water system in the U.S.. Another volume, for the 100 largest cities, appeared in 1962.

Most of our pH data come from the 1952 volumes. Measurements of pH were reported in two ways: as annual averages of daily measurements made before and after treatment by local water systems themselves, and as measurements from samples drawn on a single day and analyzed by the USGS. Our universe includes data on the water supplies of around 1,500 U.S. cities and towns from 1923 to 1984. For each place, we typically observe: the water source (specific surface-water or ground-water location from which water was drawn), method(s) of treatment, and the chemical analysis of the raw and treated water. In

[18] Baker's volumes were themselves based on weekly reports beginning in 1881 in *The Engineering News* in its "History and Statistics of American Water Works" feature. These reports through 1883 had been previously compiled for 793 cities and towns by J.R.R. Croes in *Statistical Tables From the History and Statistics of American Water Works* (New York, 1883).

[19] We have not used the lead service line inventories that the U.S. Environmental Protection Agency (EPA) required be submitted by all water systems in October 2024. For our purposes, these inventories are not relevant as they indicate the presence or absence of lead service lines only in 2024 and not at the time the individuals in our data were potentially exposed to lead in their water.

cases where separate measurements were provided for both raw and treated water, we used the raw water number. The treated pH measurements were accepted only in the absence of raw measurements and only if any specified treatment (e.g., filtration, chlorination) had only a negligible effect on pH.[20]

In addition, we have constructed a categorical measure of the reliability of our pH measures. We use this to reduce measurement error in our analysis. The highest rating is 1 and these are assigned to cities where the USGS identified the pH associated directly with the local surface water. We classify the reliability as 2 if the information is from wells and 3 if the information comes from sources of water further upstream (e.g. watershed feeding an aqueduct). Values greater than 3 are given to readings from a diffuse set of secondary sources where the readings are often intended for a different purpose. Our baseline sample uses reliability groups 1 and 2. For some analyses we restrict the sample to locations with a reliability of only 1 ("most reliable") or 1, 2 and 3 ("less reliable"). Appendix Table A1 shows examples of piping types and city pH measurements.

3.3 Data on AFQT Scores

Our outcome is based on a battery of tests called the Armed Services Vocational Aptitude Battery (ASVAB) given to all applicants to the US military. Several components of the ASVAB (e.g. math and verbal) are combined to produce the Armed Forces Qualification Test score (AFQT) which is used to admit applicants. Scores on the ASVAB are also used for occupational assignment. We acquired the AFQT scores on the universe of applicants to the US military between 1976 and 1991 and we restrict our sample to males between the ages of 17 and 20. The AFQT score is a percentile relative to a nationally representative sample of 18 to 23 year olds from the Profile of American Youth 1980. In addition to test scores we have individual level information on race, year of birth, age, years of schooling and importantly, zipcode of residence at the time of application.

3.4 Merging Piping Data to Test Scores

We link our data on piping type (lead, mixed, other) and pH levels to locations in the United States Geological Survey (USGS)'s Geographic Names Information System (GNIS) by city name and state. We then use GNIS latitude and longitude coordinates to overlay points to identify ZIP codes linked to each GNIS location.[21] The piping information linked to each zip code can then be merged into the military

[20] Our ability to observe pH at up to four points in time for some places (1923, 1932, 1952, and 1962) allows us to address concerns regarding the relevance of these point-in-time measurements to the plumbosolvency actually experienced by the individuals in our data. We are satisfied that pH is sufficiently stable to be a reliable measure of source water pH over the long run. Of greater concern is the increase over the twentieth century in the fraction of U.S. water systems that actively manipulated pH to reduce corrosion. The relevant pH for our identification strategy is the pH of the water passing through the service line.

[21] We link ZIP codes to city pH readings by first overlaying each piping type onto 1990 census place boundaries. An exception is New York City (which is its own 1990 place) which we merge at the county-level. Then, using GEOCORR relational files, we map these 1990 place boundaries to every ZIP code area it touches, defined as having at least 1 percent of a ZIP code area overlapping a 1990 place areas. One exception is if multiple GNIS features are linked to the same 1990 place (e.g. Bergen, NJ and Jersey City, NJ). In the case that these features share the same reading (e.g. Evanston and South Evanston), we combine them into a single feature. In the case of disagreement in the piping data, we link each ZIP code to the closest feature using 1990 population-weighted ZIP code centroids as long as the ZIP code centroid is within 1 mile of the GNIS location. A second edge case occurs if there is no 1990 place boundary to which a GNIS location with piping data is linked. In this case, we link all ZIP codes with centroids within 3 miles of the GNIS location. Table A10 shows the robustness of our baseline results to choices of different rules on the geographic

enlistee data based on zipcode of residence at enlistment.

3.5 Summary Statistics

In Table 1 we show key summary statistics for our various estimation samples. In our full sample (first column) we have over 6 million test takers with a mean AFQT percentile of 47. The mean age at entry is 18 at which point most applicants have 11 or more years of schooling. About 22 percent of the full sample are Black applicants.

The remaining three columns show sample characteristics when we divide the sample by the city's piping type and only include those cities with a reliability score of 1 or 2 for their pH value. We see that in cities with lead piping, AFQT scores are about 6 percentiles lower, nearly half of the sample are Black applicants, and fewer test takers are from the South. Similar, but somewhat less pronounced patterns are apparent for the samples that have mixed piping or other piping types.

4 Empirical Strategy

We use two regression models to identify the effects of lead exposure on test scores. Our first model shown in equation 4.1 only uses a sample of individuals who grew up in locations that used lead service lines. We regress the AFQT score of individual i who grew up in location j on an indicator variable for having an extreme pH, $D(\text{Extreme } pH)$. We also include an indicator for being a Black applicant, an indicator for being neither a Black nor a White applicant. In our preferred specifications, we add fixed effects for birth year, state, and survey year. Using these covariates absorbs additional variation and normalizes comparisons within these groups. To account for potential differential selection into the AFQT sample that may be related to test score outcomes, we control for the log-1990 population of the zip codes that we link to in each city. Taken together, these covariate adjustments mean that we are comparing the AFQT scores of same-year applicants of the same race and age but who are in different cities in the same state and have similar 1900 census covariates and 1990 populations. In addition, we present results both unweighted and IPW for the likelihood of one's location in childhood having extreme pH.[22] In all specifications, we cluster standard errors at the feature level. The coefficient on $D(\text{Extreme } pH)$ provides an estimate of the effect of lead exposure during childhood on test scores in late adolescence. The regression equation for this specification is as follows:

$$Y_{ijbt} = \alpha + \beta D(\text{Extreme } pH)_j + \theta X_{ij} + \delta_b + \delta_t + \delta_{s(j)} + u_{ijbt} \tag{4.1}$$

Our second model shown in equation 4.2 uses a much larger sample of locations and includes those areas without lead pipes. Here we add an indicator for having lead pipes, $D(\text{Lead pipes})$, and include the interaction between having lead pipes and extreme pH, which is now our coefficient of interest. One

overlap of ZIP code boundaries

[22] See section 2.2 and Appendix section A.3 for more discussion of the propensity score adjustment

Table 1: Descriptive statistics

	Full AFQT sample	*Lead piping*	*Mixed piping*	*Other piping*
AFQT Perc.	47.0	40.8	42.1	43.9
	(24.4)	(23.9)	(24.0)	(24.5)
South (%)	35.0	19.9	24.1	59.9
	(47.7)	(39.9)	(42.8)	(49.0)
Black (%)	21.4	46.6	40.3	36.1
	(41.0)	(49.9)	(49.0)	(48.0)
Other Non-White (%)	4.7	4.9	7.4	2.4
	(21.1)	(21.7)	(26.2)	(15.4)
Birth year	64.5	63.9	63.9	64.1
	(4.5)	(4.4)	(4.4)	(4.5)
Survey year	1983.3	1982.8	1982.8	1983.0
	(4.5)	(4.4)	(4.4)	(4.4)
Age at entry	18.1	18.2	18.2	18.2
	(1.0)	(1.0)	(1.0)	(1.0)
Years of Schooling < 11 (%)	12.4	15.4	14.9	14.2
	(32.9)	(36.1)	(35.6)	(34.9)
Years of Schooling $= 11$ (%)	42.5	39.5	38.7	38.9
	(49.4)	(48.9)	(48.7)	(48.8)
Years of Schooling > 11 (%)	45.1	45.1	46.4	46.9
	(49.8)	(49.8)	(49.9)	(49.9)
pH		7.5	7.2	7.3
		(0.6)	(0.6)	(0.6)
$pH \leq 7.0$ (%)		25.4	28.4	39.3
		(43.5)	(45.1)	(48.8)
$pH \geq 8.0$ (%)		33.2	13.3	17.0
		(47.1)	(34.0)	(37.6)
Feature population (000s), 1990		1336.2	778.5	338.4
		(1709.0)	(787.4)	(462.7)
Sample selection				
Reliability groups		1, 2	1, 2	1, 2
Piping type		Lead	Mixed	Other
N(Obs.)	6,442,125	550,420	344,931	354,134
N(Features)		117	83	127
N(Zipcodes)		715	499	572

Notes: Sample for full AFQT data are men aged 17 to 20 at entry and born between 1957 and 1974 who take the AFQT test between 1976 and 1991. Piping data are matched to pH data from this sample. Full AFQT sample refers to the enlistees meeting the same sample selection criteria prior to matching. Standard deviations are in parentheses.

value of using this second model is that allows us to directly test the effects of extreme pH in the absence of lead piping where we should not expect to see effects. We estimate this model both unweighted and using IPW to adjust for the likelihood of having lead pipes. This allows us to see whether there are any direct effects of lead piping alone that remain after our controls and inverse propensity score weighting. In both models we cluster at the feature (i.e. location) level. This differenced specification is as follows:

$$\begin{aligned} Y_{ijbt} = \alpha + \gamma_0 D(\text{Extreme } pH)_j + \gamma_1 D(\text{Lead pipes})_j \\ + \beta D(\text{Lead pipes})_j \times D(\text{Extreme } pH)_j + \theta X_{ij} + \delta_b + \delta_t + \delta_{s(j)} + u_{ijbt} \end{aligned} \tag{4.2}$$

As discussed in the next section, these controls account for differences in test score determinants between cities that may be correlated with pH and piping type. Moreover, including these controls drastically improves the precision of our estimates by absorbing residual determinants of AFQT scores. In addition, Appendix Table A7 shows that cities with extreme pH and lead piping have more individuals applying to the military and taking the AFQT, but that these differences disappear once using these regression controls, indicating that our covariate adjustments are sufficient to account for differential selection into AFQT test taking. Consistent with the absence of differential selection into the test-taking sample in Table A7, Figure A8 shows county-level test-taking rates by piping type and pH. While there are secular trends in test-taking rates over time, they are nearly identical across piping types, and enlistment rates are similar by pH among lead-pipe cities.

5 Results

5.1 Preview: Extreme *pH* and AFQT scores

Before presenting our main results we show further justification for our main specifications utilizing the interaction of lead piping with extreme pH. We do this by plotting test score residuals against pH values for our estimation samples. AFQT test scores are first regressed against our baseline set of controls and then the residuals are used for the figures.[23] In panel (a) of Figure 3 we show the plots unweighted.[24] As was the case with modern day data on blood lead levels and water pH we see a striking non linear relationship between test scores and pH. For the sample of military applicants who grew up in cities with lead pipes, the mean residuals are 2 to 6 percentiles lower when the pH of water is below 7 or above 8 compared to when pH is 7.5. The line is much flatter and never statistically significantly different from a pH of 7.5 when we use applicants who grew up in cities without lead pipes. Panel (b) shows this just for lead piped cities when using IPW to account for the likelihood of having extreme pH among lead pipe cities. Appendix Figure A5 shows the non-parametric relationship separately by race (unweighted), while

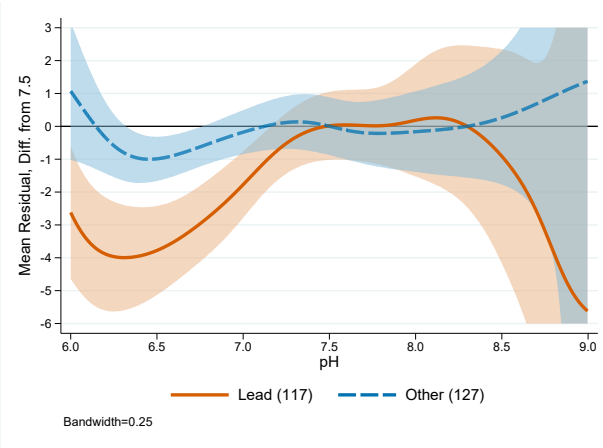
[23] We normalize the residuals by their difference relative to their values when pH is 7.5.

[24] The number of cities in Figure 3 (117 lead, 127 other) differs from Figure 1 (123 lead, 98 other) because the latter requires merging to available blood lead level data, which only some states report, whereas Figure 3 and later analyses require merging to the 1900 census data used for propensity score adjustment.

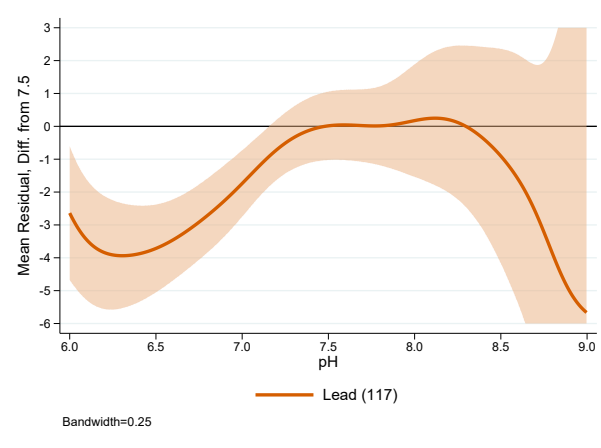
Appendix Figure A6 shows the pattern separately by race using IPW.

Figure 3: Non-parametric relationship between AFQT residuals and pH

(a) Unweighted



(b) IPW



Notes: Figure plots smoothed feature effects from a regression of AFQT on feature dummies and the baseline controls in Table 2, weighted by the precision of each estimate (the inverse of the squared standard error of each feature's estimated effect). Panel (a) shows this relationship when estimating feature effects on a combined lead and other piping sample of 244 cities (904,554 individuals) without IPW weighting, analogous to column (2) of Table 4. Panel (b) focuses on the smaller lead-piping sample of 117 cities (550,420 individuals) using weighting by the $\hat{P}(\text{Extreme}|\text{Lead})$ inverse-propensity weights, analogous to column (4) of Table 2. See Appendix Figure A7 for the combined lead and other piping sample when weighting by the probability of a city having lead pipes, analogous to column (4) of Table 4, which shows similar results to Panel (a). Shaded areas are 95 percent confidence intervals. Legend shows the number of feature effects used to estimate these relationships. Smoothing uses a Gaussian kernel and a bandwidth of 0.25 points of pH.

Panel (a) of Figure 3 also compares the relationship for both lead and other piping types. In Appendix Figure A7 we show an analogous figure using IPW to weight for the likelihood of a city having lead pipes. These figures are striking in that we show that the curve for other pipes is almost completely flat especially for pH values between 7.5 and 8.25.

5.2 Main results of Lead Exposure

The findings from our first model using only locations with lead piping is shown in Table 2. The first two columns show the estimates without propensity score weights. With no controls in column (1), we find that growing up with lead pipes and extreme pH leads to about 2.5 lower percentile scores on the AFQT. However, the estimate is statistically insignificant. In column (2) once we control for basic racial demographics and population size, our estimate on extreme pH is now -1.8 percentiles (rounded) and is statistically significant at the 1 percent level. In columns (3) and (4), we rerun the same specifications but we now use inverse probability weights (IPW) for the probability that a location had extreme pH. In column (4) we now find that our coefficient of interest is reduced a bit to an estimate of -1.3 percentiles.

In Table 3, we show how our results vary by the quality of our data on water pH. In particular our estimates in Table 2 appear to be significantly attenuated relative to using only readings where our pH readings are most reliable (coded as 1). This is evident in columns (1) and (2) where our coefficients are now -2.6 percentiles, unweighted and -2.4 percentiles when using IPW. Our baseline estimates from Table

Table 2: Main estimates: Lead pipes only

	Unweighted		*IPW*	
	(1)	(2)	(3)	(4)
D(Extreme pH)	-2.492	-1.788***	-2.506	-1.286**
	(1.778)	(0.461)	(2.245)	(0.537)
D(Black)		-18.72***		-19.17***
		(0.799)		(0.850)
D(Other Non-White)		-14.17***		-14.61***
		(0.392)		(0.586)
log(Feature pop.), 1990		-0.285*		-0.188
		(0.150)		(0.196)
Piping type	Lead	Lead	Lead	Lead
Weights	None	None	$\hat{P}$(Extreme|Lead)	$\hat{P}$(Extreme|Lead)
Controls				
Birth year FE		X		X
State FE		X		X
Survey FE		X		X
N(Obs.)	550,420	550,420	550,420	550,420
N(Features)	117	117	117	117
Extreme $\times$ *Lead*	50	50	50	50
Adjusted R^2	0.003	0.211	0.003	0.216
Outcome Mean	40.8	40.8	40.7	40.7
Outcome S.D.	23.9	23.9	23.8	23.8

Notes: See Appendix Table A8 to robustness to trimming the propensity score, Appendix Table A13 for alternate definitions of normal/extreme pH, and Appendix Table A15 for implied differences from polynomial controls for pH. Standard errors are clustered at the feature-level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2 are reproduced in columns (3) and (4). We show that our estimates are even further attenuated when we include locations where the reliability is less reliable (coded 1, 2 or 3). This is shown in columns (5) and (6) where our estimates drop to between -0.7 and -1 percentiles but are statistically insignificant. Clearly, measurement error can attenuate the true effect of lead exposure. Nevertheless, to be conservative, we focus on the baseline estimates for our cost-benefit analysis and our main conclusions.

Table 3: Lead pipe estimates by pH measurement reliability

	Most Reliable		*Baseline*		*Less Reliable*	
	(1)	(2)	(3)	(4)	(5)	(6)
	Unweighted	IPW	Unweighted	IPW	Unweighted	IPW
D(Extreme pH)	-2.641***	-2.391**	-1.788***	-1.286**	-0.979*	-0.684
	(0.848)	(1.102)	(0.461)	(0.537)	(0.529)	(0.518)
D(Black)	-17.78***	-17.74***	-18.72***	-19.17***	-19.02***	-19.84***
	(1.161)	(1.055)	(0.799)	(0.850)	(0.704)	(0.639)
D(Other Non-White)	-13.67***	-13.78***	-14.17***	-14.61***	-14.26***	-14.83***
	(0.425)	(0.407)	(0.392)	(0.586)	(0.407)	(0.536)
log(Feature pop.), 1990	-0.681***	-0.696**	-0.285*	-0.188	-0.442***	-0.308*
	(0.239)	(0.302)	(0.150)	(0.196)	(0.162)	(0.179)
Piping type	Lead	Lead	Lead	Lead	Lead	Lead
Propensity weights	None	$\hat{P}$(Extreme|Lead)	None	$\hat{P}$(Extreme|Lead)	None	$\hat{P}$(Extreme|Lead)
N(Obs.)	324,938	324,938	550,420	550,420	665,675	664,649
N(Features)	57	57	117	117	147	146
Extreme × Lead	21	21	50	50	61	60
Adjusted R^2	0.198	0.198	0.211	0.216	0.214	0.220
Outcome Mean	39.2	39.1	40.8	40.7	41.6	41.7
Outcome S.D.	23.3	23.3	23.9	23.8	24.1	24.2

Notes: Propensity score weights are computed on the baseline sample of pipes and extrapolated for "Most Reliable" and "Less Reliable" pipes. See Appendix Tables A11 and A12 for complete weighted and unweighted results across pH reliability and piping types. Standard errors are clustered at the feature-level.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Our second empirical approach is to include military applicants who lived in locations with "other" piping that did not include lead. In this specification, we include an added level of differencing, allowing us to control for any direct effects of extreme pH. This approach also greatly expands our sample from 550,000 to 900,000, potentially leading to more precise estimates. The results are shown in Table 4. In column (2) when our regressions are unweighted and we add our main controls, we find that the coefficient on the interaction of having lead pipes and extreme pH is to lower AFQT scores by about 1.7 percentiles which is quite similar to the unweighted results in Table 2. In column (4) we use IPW to deal with any potential determinants of test scores that are related to the probability that a city adopted lead pipes. In this specification, we find a slightly more negative estimated effect of lead exposure on test scores. The coefficient on the interaction between lead pipes and extreme pH is now -1.8 percentiles.

A useful feature of this approach is we can show that there is no statistically significant direct effect of extreme pH on test scores, which holds both before and after propensity score adjustment. We consider this null result across piping types as encouraging for the plausible exogeneity of pH with respect to test score unobservables, especially after further adjusting for baseline observables.

On the other hand, we estimate a positive effect of a city having lead pipes on test scores, even after IPW adjustment. As we do not know of any direct mechanism by which lead pipes would have a causal

positive effect on test scores, this suggests that our estimates on the full sample are not eliminated of all the unobserved determinants of test scores related to the historical adoption of lead pipes. Consistent with this, Appendix Table A4 shows that lead-pipe cities have school enrollment rates that are about 4 percentage points higher than cities with other piping types in 1900 even after IPW adjustment. As our coefficient of interest in Table 4 is the interaction between lead and extreme pH, the gap in levels is differenced out. Our estimates would be biased only if the imbalance in unobservables between lead-pipe and other cities were systematically different among extreme-pH cities than among normal-pH cities. That the combined sample with both piping types gives estimates similar to the lead-only sample is consistent with this assumption and encouraging for our research design. Nevertheless, for these reasons, we prefer the within-lead estimates by pH from Table 2 as our baseline.

Table 4: Main differenced estimates: Lead and other pipes

	Unweighted		*IPW*	
	(1)	(2)	(3)	(4)
D(Extreme pH)	0.00411	0.00638	1.717	0.444
	(1.957)	(0.367)	(2.224)	(0.644)
D(Lead pipes)	-1.640	0.979***	0.989	1.203**
	(2.185)	(0.374)	(2.318)	(0.528)
D(Lead pipes) × D(Extreme pH)	-2.497	-1.725***	-4.073	-1.819**
	(2.641)	(0.518)	(2.788)	(0.833)
D(Black)		-19.51***		-19.91***
		(0.547)		(0.395)
D(Other Non-White)		-13.65***		-13.79***
		(0.472)		(0.485)
log(Feature pop.), 1990		-0.0259		0.0303
		(0.117)		(0.145)
Piping type	Other, Lead	Other, Lead	Other, Lead	Other, Lead
Weights	None	None	$\hat{P}$(Lead)	$\hat{P}$(Lead)
Controls				
Birth year FE		X		X
State FE		X		X
Survey FE		X		X
N(Obs.)	904,554	904,554	904,554	904,554
N(Features)	244	244	244	244
Extreme × Lead	50	50	50	50
Adjusted R^2	0.005	0.221	0.002	0.225
Outcome Mean	42.0	42.0	41.9	41.9
Outcome S.D.	24.2	24.2	24.2	24.2

Notes: See Appendix Table A9 to robustness to trimming the propensity score, Appendix Table A14 for alternate definitions of normal/extreme pH, and Appendix Table A16 for implied differences from polynomial controls for pH. Standard errors are clustered at the feature-level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We also examine how measurement error in our pH measure may impact our findings in the differenced model in Table 5. We find that our estimates are roughly 3 times larger when we use the sample of cities where we have the most reliable measures of water pH. For example, in our IPW weighted regression the effect of lead exposure is now estimated to lead to a 5 percentile reduction in test scores.

However, these specifications also imply very large positive direct effects of extreme pH and having lead piping. This is likely related to identification relying on a relatively small number of locations. Given that our estimates when using a sample of cities without lead pipes, sometimes yield estimates that appear to be due to unobservable confounders, we prefer to emphasize estimates based on the subsample of cities with lead piping and rely on the design exploiting water chemistry differences within these cities.

Table 5: Differenced estimates by pH measurement reliability

	Most Reliable		*Baseline*		*Less Reliable*	
	(1)	(2)	(3)	(4)	(5)	(6)
	Unweighted	IPW	Unweighted	IPW	Unweighted	IPW
D(Extreme pH)	2.974***	2.777***	0.00638	0.444	-0.344	0.243
	(0.947)	(0.614)	(0.367)	(0.644)	(0.402)	(0.500)
D(Lead pipes)	4.732***	4.534***	0.979***	1.203**	-0.0392	-0.0157
	(0.980)	(0.895)	(0.374)	(0.528)	(0.607)	(0.684)
D(Lead pipes) × D(Extreme pH)	-5.542***	-5.045***	-1.725***	-1.819**	-0.517	-0.794
	(1.204)	(0.928)	(0.518)	(0.833)	(0.661)	(0.762)
D(Black)	-18.03***	-18.58***	-19.51***	-19.91***	-20.00***	-20.23***
	(1.031)	(0.890)	(0.547)	(0.395)	(0.448)	(0.335)
D(Other Non-White)	-13.54***	-13.83***	-13.65***	-13.79***	-13.34***	-13.26***
	(0.460)	(0.468)	(0.472)	(0.485)	(0.480)	(0.460)
log(Feature pop.), 1990	-0.669***	-0.660***	-0.0259	0.0303	-0.105	-0.0119
	(0.198)	(0.184)	(0.117)	(0.145)	(0.115)	(0.132)
Piping type	Other, Lead	Other, Lead	Other, Lead	Other, Lead	Other, Lead	Other, Lead
Propensity weights	None	$\hat{P}$(Lead)	None	$\hat{P}$(Lead)	None	$\hat{P}$(Lead)
N(Obs.)	380,786	380,786	904,554	904,554	1,259,960	1,243,471
N(Features)	81	81	244	244	356	352
Extreme × Lead	21	21	50	50	61	60
Adjusted R^2	0.202	0.207	0.221	0.225	0.222	0.221
Outcome Mean	40.0	40.3	42.0	41.9	43.2	43.3
Outcome S.D.	23.6	23.8	24.2	24.2	24.4	24.4

Notes: Propensity score weights are computed on the baseline sample of pipes and extrapolated for "Most Reliable" and "Less Reliable" pipes. See Appendix Tables A11 and A12 for complete weighted and unweighted results across pH reliability and piping types. Standard errors are clustered at the feature-level.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.3 Heterogeneity by Race and Poverty Status

We also consider differential effects for Black individuals and by whether an area has a high rate of poverty by interacting our baseline model (using cities with lead piping), with indicator variables for being in these categories. The results are shown in Appendix Table A5. We find no statistically significant differences. The point estimates are consistently more negative in higher poverty areas but we lack power to identify a significant difference. We also note that the difference between the share of Black individuals living in extreme pH cities is about 8 percentage points higher than that for white individuals. Nevertheless, a simple "back of the envelope" calculation suggests that a policy that removes LSLs from extreme pH cities would have a very small effect on reducing the Black-White test score gap.[25]

[25]It is worth noting that Black individuals are more likely to live in high pH cities. Therefore if lead exposure in very high pH areas had a more deleterious effect on test scores than exposure in very low pH areas then it might be the case that a policy focused on removing lead from high pH cities could help reduce the racial test score gap. However, we lack sufficient power to detect differences in effects between very low pH and very high pH lead exposure.

5.4 Mismeasurement of Lead Exposure during Childhood

An important caveat to our analysis is that we lack data on the location of military applicants throughout their childhood. Instead, we proxy for their location based on their residence at the time that they apply to the military (between the ages of 17 and 20). To the extent that some applicants may have moved locations between early childhood and adolescence/young adulthood, our analysis may be adding measurement error to the right hand side variable. It could be that some children with true exposure to lead pipes and extreme pH in childhood would be incorrectly classified as having no exposure because they moved, or vice versa. Thus our estimates are likely an underestimate of the true effects. Indeed Fletcher and Noghanibehambari [2025] find significantly worse effects on longevity from lead exposure in one's city of birth than exposure to lead later in life.

To get an idea of the severity of the measurement error, we compiled several statistics that are relevant for our sample. First, the share of individuals in the 1980 and 1990 Census who moved from their state of birth by age 18 was 22.9%. We also find that the fraction who experienced any migration in the last 5 years, either across or within their MSA in the 1980 and 1990 Census was 39.9%. This is a broader statistic that captures smaller geographic moves, but during a shorter time window. Finally, using the NLSY79 [U.S. Bureau of Labor Statistics, 2025], the fraction moving from their city/town/county of birth (excluding foreign born) was 55.1%.[26] Overall, these statistics suggest that there may be a substantial degree of mismeasurement of childhood lead exposure further suggesting our estimates may be quite conservative.

In Appendix Table A6, we directly estimate the effects of mismeasurement by excluding areas which have historically had a high rate of migration and enlistees are likely to have grown up elsewhere. Specifically, we drop counties with a large share of residents in 1980 born outside the state or have a large share of recent migrants.[27] While these estimates appear to imply that there is no attenuation due to migration, it is also possible that these high migration counties happen to be places that had larger negative effects of lead exposure, such that their removal from the sample weakened the overall relationship. Overall, we suspect that our estimates are attenuated due to migration but we continue to use our estimates without a correction factor for measurement error, as our baseline estimates.

5.5 Additional Robustness Checks

A possible threat to our research design is if there is differential selection of lower scoring individuals into the military applicant sample in extreme pH lead cities. If this were the case, there would likely be

[26] This uses the following NLSY79 variable `FAM-1_4_1979` (reference number R00030.00): "Some people live in the same place all of their lives, while others move from time to time. How about you—have you lived here in this (city/town/county) all of your life?"

[27] Based on the 1980 census county tables [Manson et al., 2024], we drop counties where a large share of residents lived in a different county five years prior (Columns 2 and 5 of Appendix Table A6), which drops 9 cities (40,916 individuals), or where a large share of residents are born outside the state (Columns 3 and 6 of Appendix Table A6), which drops 27 cities (35,754 individuals). For our baseline estimates that only use cities with lead piping, we find that the effects of dropping high migration counties actually lowers our point estimates slightly from -1.3 to about -1.2 for both exercises. For our broader sample that includes cities without lead piping, the estimates of lead exposure also fall from -1.8 to either -1.7 or -1.4.

a correlation between our exposure measure and the total number of test takers. In Table A7, we show that extreme pH lead cities do indeed have larger sample sizes but we find that this difference is entirely explained by controlling for 1990 population. In other words, we find no differential application rates across cities with different piping and pH. Selection into test-taking could only drive our results if it were entirely due to compositional changes — an increased application rate of low-scoring individuals that is entirely offset by a decreased application rate of high scoring individuals, with no direct effect of lead exposure on cognitive ability. Such a scenario seems unlikely. Moreover, in Section 5.6 below, we find little impact of lead exposure on the lower-tail of the test score distribution. In fact the impacts are largely concentrated in the upper to middle part of the test score distribution, which is inconsistent with an influx of lower scoring applicants.

Our estimates are similar across different sample selection criteria and trimming of the propensity scores. While our baseline estimates do not exclude any data based on the p-score, the estimates are robust to excluding data by trimming at the 5 or 10 percent tails of the p-score, as shown in Tables A8 and A9. This indicates our results are not driven by a lack of overlap in observables between treated and control cities. Results are slightly attenuated but similar when controlling for deciles of the propensity score instead of weighting. Given the small number of cities with lead piping, we opt to include all the available data as our preferred specification so as not to underpower our analysis.

In addition, Table A10 shows that our results are robust to different choices when merging geographies across datasets. The main exception is when we only use zip codes that are entirely contained within a 1990 place's geography, which significantly cuts the sample. Finally, Tables A11 and A12 show the robustness of our main results across different reliability groups and piping types. The differencing results in these tables are similar when we included mixed piping to identify the coefficients on controls and main effects.

We also show that our results are robust to changing the pH cutoffs used for "normal" pH as well as using polynomials in pH. While our baseline approach is to define pH outside $(7, 8)$ as extreme, there is no reason for this cutoff to be sharp. Tables A13 and A14 keep our main treatment discrete and considers alternate definitions of normal pH, re-estimating the propensity score on each alternate treatment. Across most alternate cutoffs, our estimates increase. The main exception is that the differencing analysis using lead and other piping attenuates when using wide cutoffs. Similarly, Tables A15 and A16 use polynomials up to the fifth-degree to flexibly estimate relationships between pH and test scores, computing the implied impacts of extreme pH levels relative to 7.5. While sometimes imprecise, we consistently find negative impacts of extreme pH even for the most flexible fifth-degree polynomial, with stronger negative effects at higher (i.e., more basic) pH-levels. This may be suggestive that the ingestion of lead as a solid particle due to flaking in high pH water could be more harmful for cognition than lead ingestion due to a dissolved ion in low pH water.

5.6 Quantile Treatment Effects

Figure 4 shows the impact of extreme-pH lead exposure among cities with lead pipes on the distribution of test scores by conducting a quantile treatment effects analysis (QTE) on test score residuals. To make these figures, we take test score residuals from a regression on our baseline controls and compute percentiles of this residual in both normal and extreme-pH cities. Differences between the corresponding quantiles of the residualized AFQT distributions give the QTE estimates. Importantly, this analysis looks at impacts on treated versus control outcome distributions, not differences in treatment effects.[28]

The results in Figure 4 show that the most significant impacts of lead exposure occur in the middle of the distribution of test scores, with effects on the order of 1 percentiles. There are few impacts in the bottom third or very top of the distribution. In addition, there are also large but noisy impacts in the upper-middle part of the distribution on the order of 1 to 2 percentiles. Such distributional effects may arise if very low AFQT scorers are already disadvantaged and have low counterfactual scores in the absence of extreme pH. Similarly, very high scorers may not be as impacted if they have less exposure or more chances for remediation, for example, if they are from higher socioeconomic status backgrounds. This may be why we see impacts in the middle and upper-middle parts of the distribution, where individuals both have more room to fall as well as less opportunity for remediation.

Finally, we consider the absence of significantly positive effects as reassuring for our research design. As there is little reason to expect lead exposure in extreme pH cities to *increase* test scores, positive effects would be difficult to reconcile with our preferred interpretation and mechanisms. Across the distribution, we find null or significantly negative effects of lead exposure on AFQT scores. The one exception to this rule is the positive impacts below the 30th percentile, but these effects are not statistically significant and confidence intervals include large negative effects.

6 Cost-Benefit Analysis

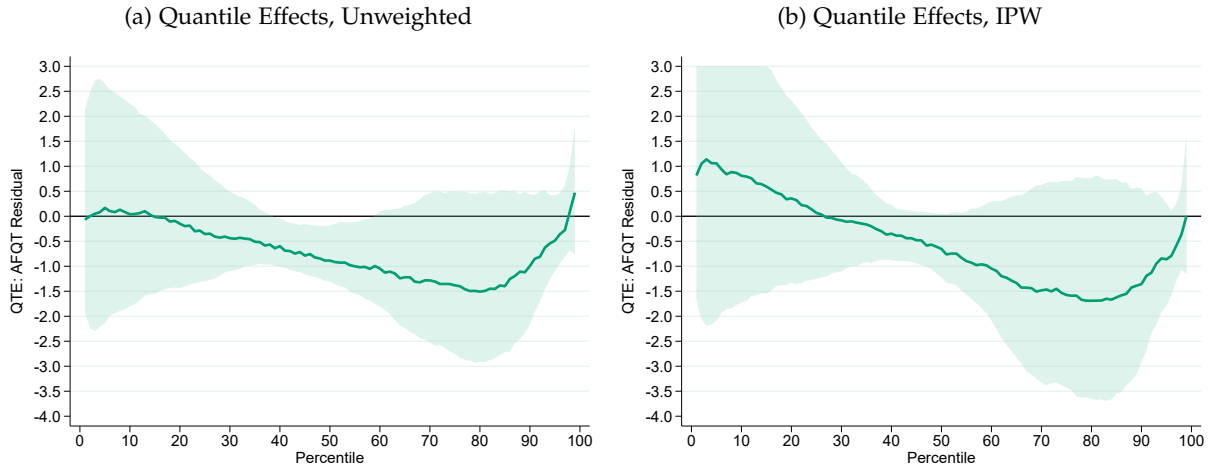
In Appendix Table A18 we show that a one percentile increase in AFQT scores translates into a $45,811 increase in earnings over the life cycle.[29] Combining this with our main estimates of the impact of lead exposure in extreme pH cities on AFQT scores of -1.286 (0.537), indicates a decrease in earnings of $58,913 over the life cycle. Discounting back to birth at a 3% rate, this translates into a decrease of $14,013 in present value earnings. As long as the upfront costs of lead piping replacement is less than this number, it will be cost effective in the sense that it generates earnings increases that outweigh its cost. Moreover, assuming a tax and transfer rate of 18% that we use below, the policy will pay for itself through increased tax revenues if the upfront costs are less than $2,522 ($= 0.18 \times 14,013$).

A good case study in the costs of water piping replacement comes from the city of Madison, Wisconsin,

[28] For example, it could be that high counterfactual AFQT scorers are most exposed to lead and fall into the bottom tail, while those with low counterfactual AFQT somehow benefit and move to the top of the distribution. This would result in the same observed distribution as both high and low counterfactual AFQT scorers having no response.

[29] See Appendix Table A17 for a replication and extension of Neal and Johnson [1996] to the full NLSY79.

Figure 4: Impact of pH on AFQT distribution: Lead pipes only



Notes: At each percentile of the residualized AFQT distribution for extreme pH places, we take the difference in that quantile from the corresponding quantile in the normal pH range distribution (7.0-8.0). Differences in these values give us quantile treatment effects. 95 percent confidence intervals come from the 2.5th and 97.5th quantiles of these effects resulting from 1,000 bootstrap samples, sampling features and then enlistees with replacement. The AFQT distribution is residualized on controls in each bootstrap sample.

which replaced all of the city's remaining lead pipes between 2000 and 2011 at a nominal cost to the city of $15.5 million [City of Madison, 2026]. The program also mandated that homeowners pay at least 50% of the total cost of their private lines, the other half of which was paid by the city, which results in additional private costs of replacement of $3.75 million.[30] We calculate that this results in an upfront cost of about $30.4 million to replace all of Madison's lead water pipes, in 2025 dollars and discounted to the year of first intervention, 2000.[31] In addition, we estimate that only about 15.1 percent of Madison's piping was affected by the program, which we use as a proxy for the share of the number of children affected.[32] Using census data to compute birth cohort sizes, we estimate a cost-per-child of $9,853 when using all children in Madison under the age of 10 in 2000. In order to account for impacts on unborn future cohorts, we convert this into a cost-per-effective child of $2,343, by forecasting birth cohort sizes and discounting over a fifty year horizon[33] to the present day.[34]

[30] Over 5,600 homeowners had to replace their private side of the lines. At an average reported cost of $670 after reimbursement, this results in additional private costs of $3.75 million of the program. We add these private costs to the total social cost.

[31] While the costs were paid over 12 years, the city reports most of the costs were paid in the earlier years of the program, averaging only $100K per year after 2006. We assume these reported numbers are divided equally and are reported in nominal dollars in the year the costs were incurred.

[32] Only 8,000 total service lines were affected [City of Madison, 2026]. The Madison water utility annual report lists the number of water service lines in 2000 at 53,138 [Madison Water Utility, 2000], which gives us the 15.1% of pipes that were affected ($\frac{8,000}{53,138}$). Since lead piping may be more prevalent in older buildings or denser areas, it is likely that this share is an underestimate.

[33] The horizon of fifty years is chosen to be consistent with typical water pipe service life as reported by the American Water Works Association [2012], spanning 50 to 135 years depending on the material and region.

[34] To compute the effective number of exposed children, we look at birth cohorts born between 2001 and 2020. We estimate birth cohort sizes between 2021 and 2050 using the average size between 2016 and 2020. Then, we discount all post-2000 birth cohorts at a 3% rate to get an effective number of children treated, which down-weights the influence of children in the future. As this is a denominator in producing average cost estimates per child, our final MVPF using effective child costs is numerically equivalent to dividing total benefits – for both current and future cohorts – by total costs, where total benefits include discounting forecast earnings benefits.

An alternative to using costs from this case study is to use estimates from the EPA's economic analysis of lead service line replacement, which reports an average cost of $8,571, converted to 2025 dollars [U.S. Environmental Protection Agency, 2024b].[35] We treat this as a cost-per-structure of replacement, and convert this to a cost-per-child of $23,227 by dividing by 0.369, the average number of children under 10 per structure in the 2024 American Community Survey (ACS). In order to account for impacts on future cohorts of children, we then convert this to a cost-per-effective child of $6,677 by forecasting birth cohort sizes and discounting over a fifty year horizon to the present day.

Conducting a cost-benefit analysis in the style of Hendren and Sprung-Keyser [2020] indicates that replacing lead piping is generally cost effective and may even end up paying for itself, at least in the cities with extreme-pH. However, this result is sensitive to the exact cost estimates used and whether these estimates account for impacts on future generations. This cost-benefit analysis is displayed in Table 6, which reports marginal value of public funds (MVPF) estimates, dividing present values of estimated average post-tax benefits by tax-adjusted upfront costs per affected individual. Below the MVPF estimates, we report the tax-adjusted present value of upfront costs-per-child.

Lead piping replacement is always valued more than its cost in calculations that account for future affected cohorts. We only find that replacing lead pipes is valued at less than its cost when using the higher EPA cost estimates as well as ignoring impacts on future children. However, if we include future birth cohorts and use the lower cost-per-child based on Madison's reported costs, lead pipe removal pays for itself through increased future taxes, resulting from higher earnings children spared from lead exposure. In panel C of Table 6, we adjust for attenuation from migration by dividing by 0.45, the share of individuals staying in their city of birth according to the NLSY79. We also emphasize that these MVPF estimates do not account for hard-to-quantify externalities that are likely to make lead piping removal even more cost effective, such as impacts on crime, reductions in the usage of public programs, or increases in property values accruing to homeowners.

Our preferred MVPF is 2.77, reflecting the more conservative EPA cost estimate but incorporating impacts on future cohorts. Comparing these MVPFs to those reported in Hendren and Sprung-Keyser [2020] indicates a return on investment to replacing lead piping that is lower (but not uniformly so) than other early childhood education and health interventions, but a much higher cost effectiveness than other adult-focused programs and policies. For example, MVPFs range from 10.24 to ∞ for Medicaid expansions to children and pregnant women, from 0.72 to ∞ for early-childhood programs (such as Head Start, Perry Preschool, and Abecedarian), and from 0.65 to ∞ for K–12 school spending. The Moving to Opportunity experiment, which similarly to lead piping replacement affects a child's neighborhood environment, has a reported MVPF of infinity (i.e., pays for itself) but with a wide confidence interval spanning -2.8 to ∞. In contrast, lead service line replacement is much more cost effective when compared to common adult

[35] This can be found in Table A-1 in Appendix A to the EPA economic analysis, which reports lead service line replacement cost estimates based on the Drinking Water Infrastructure Needs Survey and Assessment.

policy interventions, such as job training (-0.23 to 1.48), housing vouchers and food stamps (0.65 to 1.04), and adult health insurance expansions (0.40 to 1.63).

Table 6: Marginal Value of Public Funds Estimates

	A. Upfront Cost-per-Child	*B. MVPF, Unadjusted*			*C. MVPF, Adjusted for Migration*		
		Baseline	*Differenced*	*Most Reliable*	*Baseline*	*Differenced*	*Most Reliable*
I. Earnings Impacts		14,013	19,820	26,053	31,208	44,143	58,024
SE		5,851	9,077	12,008	10,619	16,473	21,792
II. Madison Cost Estimate							
Per Child							
MVPF		1.57	2.59	4.14	6.04	18.97	∞
Cost	9,853	7,331	6,286	5,164	4,236	1,908	-591
Per Effective Child							
MVPF		∞	∞	∞	∞	∞	∞
Cost	2,343	-179	-1,224	-2,346	-3,274	-5,602	-8,101
III. EPA Cost Estimate							
Per Child							
MVPF		0.55	0.83	1.15	1.45	2.37	3.72
Cost	23,227	20,705	19,660	18,538	17,610	15,282	12,783
Per Effective Child							
MVPF		2.77	5.23	10.75	24.16	∞	∞
Cost	6,677	4,154	3,109	1,987	1,059	-1,269	-3,768

Notes: Presents estimates of the net costs and marginal value of public funds from replacing lead pipes in extreme-pH cities. All calculations use a 3% discount rate and an 18% tax and transfer rate [Hendren and Sprung-Keyser, 2020]. Column panel A reports the main upfront cost estimates, both for all children under 10 as well as adjusting for additional impacts on birth cohorts 50 years into the future, discounted to time costs are realized. Column panel B reports the MVPF for various estimates of extreme-pH lead piping on AFQT scores. Column panel C adjusts these earnings impact estimates for migration by dividing the coefficients and standard errors by 0.45, the share of NLSY79 respondents who had not moved from their city of birth (see Section 5.4). "Baseline" reports our main lead-only estimates from Table 2; "Differenced" reports our estimates combining lead and other piping from Table 4; "Most Reliable" reports the estimates using only the highest reliable groups, reported in Table 3. Row panel I shows the estimated pre-tax impacts on lifetime earnings for the various estimates reported in the text, using Column (2) of Table A18 for the relationship between AFQT and earnings. Earnings impacts are discounted to birth. Estimates from row panel II ("Madison Cost Estimate") uses the costs from Madison, Wisconsin's lead service line replacement and row panel III ("EPA Cost Estimate") use costs from the Environmental Protection Agency, both discussed in the main text. Net up front costs after accounting for increased tax revenue are reported below MVPFs, with upfront costs reported in column panel A.

7 Conclusion

Despite a huge amount of knowledge about the dangers of lead exposure on the cognitive development of children, waterborne exposure to lead through lead service lines (LSLs) continues to be a reality in many American cities. We quantify a portion of the economic benefits that would accrue from removing this last remaining vector of lead exposure by measuring the impacts of lead exposure on AFQT scores used to admit applicants to the US military. We do this by linking a huge sample of between 500,000 and 1 million applicants to the military between 1976 and 1991 to a proxy for their lead exposure during their childhood. We exploit differences in water chemistry across cities with lead piping to identify locations where lead is likely to contaminate drinking water either through very low or very high pH levels.

We find that exposure to lead leads to at least a 1.3 percentile decrease in AFQT scores. This translates into a reduction in the present discounted value of lifetime earnings of about $14,000. This implies that the economic benefits of removing waterborne lead exposure is extremely high and that removing LSLs in cities with extreme pH levels would be a highly cost effective policy. By comparison, the effects of the Clean Air Act on lifetime earnings [Isen et al., 2017] is estimated to be less than half of what we

find. Under conservative assumptions, we find that the marginal value of public funds (MVPF) for such a policy would be about 2.8. It is worth noting that this does not take into account other health effects of reducing lead exposure or other externalities such as reductions in crime that would likely accrue from such a policy.

References

Anna Aizer and Janet Currie. Lead and juvenile delinquency: New evidence from linked birth, school, and juvenile detention records. *Review of Economics and Statistics*, 101(4):575–587, 2019.

Anna Aizer, Janet Currie, Peter Simon, and Patrick Vivier. Do low levels of blood lead reduce children's future test scores? *American Economic Journal: Applied Economics*, 10(1):307–341, 2018.

American Water Works Association. Buried no longer: Confronting america's water infrastructure challenge. Technical report, American Water Works Association, Denver, CO, 2012. URL https://www.awwa.org/wp-content/uploads/Buried-No-Longer.pdf.

Randolph K. Byers and Elizabeth E. Lord. Late effects of lead poisoning on mental development. *American Journal of Diseases of Children*, 66(5):471–494, 1943.

City of Madison. Information for utilities on lead service replacement. https://www.cityofmadison.com/water/water-quality/lead-copper-in-water/information-for-utilities-on-lead-service-replacement, 2026.

Federal Register. National primary drinking water regulations for lead and copper: Improvements LCRI. https://www.federalregister.gov/documents/2024/10/30/2024-23549/national-primary-drinking-water-regulations-for-lead-and-copper-improvements-lcri, 2024. Accessed: September 4, 2026.

James J. Feigenbaum and Christopher Muller. Lead exposure and violent crime in the early twentieth century. *Explorations in Economic History*, 62:51–86, 2016. ISSN 0014-4983. doi: https://doi.org/10.1016/j.eeh.2016.03.002. URL https://www.sciencedirect.com/science/article/pii/S0014498316300109.

Joseph P. Ferrie, Karen Rolf, and Werner Troesken. 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 & Human Biology*, 10(1):98–111, 2012. ISSN 1570-677X. doi: https://doi.org/10.1016/j.ehb.2011.09.003. URL https://www.sciencedirect.com/science/article/pii/S1570677X11000943.

Jason Fletcher and Hamid Noghanibehambari. Early-life lead exposure and male longevity: Evidence from historical municipal water systems. *Explorations in Economic History*, 97:101701, 2025. ISSN 0014-4983. doi: https://doi.org/10.1016/j.eeh.2025.101701. URL https://www.sciencedirect.com/science/article/pii/S0014498325000488.

Philippe Grandjean. Even low-dose lead exposure is hazardous. *The Lancet*, 376(9744):855–856, 2010.

Hans Grönqvist, J. Peter Nilsson, and Per-Olof Robling. Understanding how low levels of early lead exposure affect children's life trajectories. *Journal of Political Economy*, 128(9):3376–3433, 2020.

James Grout. Lead poisoning and Rome. *Encyclopaedia Romana*, pages 11–14, 2008. URL https://penelope.uchicago.edu/encyclopaedia_romana/wine/leadpoisoning.html.

DZ Hambrick, AP Burgoyne, EM Altmann, and TJ Matteson. Explaining the validity of the ASVAB for job-relevant multitasking performance: The role of placekeeping ability. *Journal of Intelligence*, 11(12), 2023.

James J. Heckman, Jora Stixrud, and Sergio Urzua. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–482, 2006. doi: https://doi.org/10.1086/504455.

Nathaniel Hendren and Ben Sprung-Keyser. A unified welfare analysis of government policies. *The Quarterly Journal of Economics*, 135(3):1209–1318, 2020. doi: 10.1093/qje/qjaa006.

Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences.* Cambridge University Press, 2015.

Adam Isen, Maya Rossin-Slater, and W. Reed Walker. Every breath you take—every dollar you'll make: The long-term consequences of the clean air act of 1970. *Journal of Political Economy*, 125(3):848–902, 2017. doi: 10.1086/691465.

Haena Lee, Mark W. Lee, John Robert Warren, and Joseph Ferrie. Childhood lead exposure is associated with lower cognitive functioning at older ages. *Science Advances*, 8(45), 2022.

Madison Water Utility. Annual report of Madison water utility for the year ended December 31, 2000. Water, Electric, or Joint Utility Annual Report Utility No. 3280, Public Service Commission of Wisconsin, Madison, WI, 2000.

Steven Manson, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. IPUMS national historical geographic information system: Version 19.0 [dataset]. Minneapolis, MN: IPUMS, 2024. URL http://doi.org/10.18128/D050.V19.0.

J. Montgomery, J.A. Evans, S.R. Chenery, V. Pashley, and K. Killgrove. Gleaming white and deadly: using lead to track human exposure and geographic origins in the roman period in britain. *Journal of Roman Archaeology*, 78:199–226, 2010.

M. R. Moore. Plumbosolvency of waters. *Nature*, 243(5404):222–223, 1973.

Derek A. Neal and William R. Johnson. The role of premarket factors in black-white wage differences. *Journal of Political Economy*, 104(5):869–895, 1996. doi: 10.1086/262045. URL https://doi.org/10.1086/262045.

Herbert Needleman. Lead poisoning. *Annual Review of Medicine*, 55(1):209–222, 2004.

NRDC. The epa now says there aren't 9 million lead pipes—there are 4 million. be skeptical. https://www.nrdc.org/media/epa-now-says-there-arent-9-million-lead-pipes-there-are-4-million-be-skeptical, 2025. Accessed: October 14, 2026.

Richie Poulton, Terrie E. Moffitt, and Phil A. Silva. The Dunedin multidisciplinary health and development study: Overview of first 40 years, with an eye to the future. *Social Psychiatry and Psychiatric Epidemiology*, 50:679–693, 2015.

Michael R. Schock. Causes of temporal variability of lead in domestic plumbing systems. *Environmental Monitoring and Assessment*, 15(1):59–82, 1990.

Michael R. Schock and Marvin C. Gardels. Plumbosolvency reduction by high pH and low carbonate—solubility relationships. *Journal AWWA*, 75(2):87–91, February 1983. Curbing Corrosion Issue.

J. Sherlock, G. Smart, G.I. Forbes, M.R. Moore, W.J. Patterson, W.N. Richards, and T.S. Wilson. Assessment of lead intakes and dose-response for a population in ayr exposed to a plumbosolvent water supply. *Human Toxicology*, 1(2):115–122, 1982.

U.S. Bureau of Labor Statistics. National longitudinal survey of youth 1979 cohort, 1979–2022 (rounds 1–30). Produced and distributed by the Center for Human Resource Research (CHRR), The Ohio State University. Columbus, OH, 2025. URL https://www.nlsinfo.org/content/cohorts/nlsy79.

U.S. Environmental Protection Agency. Updated 7th drinking water infrastructure needs survey assessment may 2024, 2024a. URL https://www.epa.gov/system/files/documents/2024-05/fact-sheet_one-time-update_2024.04.30_508_compliant_1.pdf.

U.S. Environmental Protection Agency. Economic analysis appendices for the final lead and copper rule improvements. Technical Report EPA 810-R-24-005, Office of Water, Washington, DC, 10 2024b. URL https://www.epa.gov/system/files/documents/2024-10/508_lcri_final-ea_appendices_10_23_24.pdf.

U.S. Environmental Protection Agency. Lead service lines, 2026a. URL https://www.epa.gov/ground-water-and-drinking-water/lead-service-lines.

U.S. Environmental Protection Agency. Learn about lead. https://www.epa.gov/lead/learn-about-lead, 2026b. Accessed: September 4, 2026.

Christian Warren. *Brush with death: a social history of lead poisoning*. JHU Press, 2001.

Appendix A: Additional Material

A.1 Description of Piping Data

The material used for the water service lines of 4,090 cities and towns was obtained from eight sources:

1. Croes, J.J.R. Statistical Tables from the History and Statistics of American Water Works. Engineering News: 1883.
2. Croes, J.J.R. Statistical Tables from the History and Statistics of American Water Works. Engineering News: 1885.
3. Croes, J.J.R. Statistical Tables of American Water Works. Engineering News: 1887.
4. Baker, M.N. Manual of American Water-Works. Vol. 1. Engineering News: 1888.
5. Baker, M.N. Manual of American Water-Works. Vol. 2. Engineering News: 1889/90.
6. Baker, M.N. Manual of American Water-Works. Vol. 4. Engineering News: 1897.
7. Massachusetts, State Board of Health. "Materials Used for Service Pipes in Massachusetts." Thirty-Seventh Annual Report (1906): 197-205.
8. Public Works. "Water Works Services." 52 (19): 352-355 and 52 (23): 430-433.

Typical entries are "lead," "galvanized iron," "cast iron," "wrought iron," and combinations of these. These reports reflect the predominant material in place at a point in time. Cities occasionally shifted from one material or mix of materials to another. If they did, however, they seldom required customers to replace existing service lines.

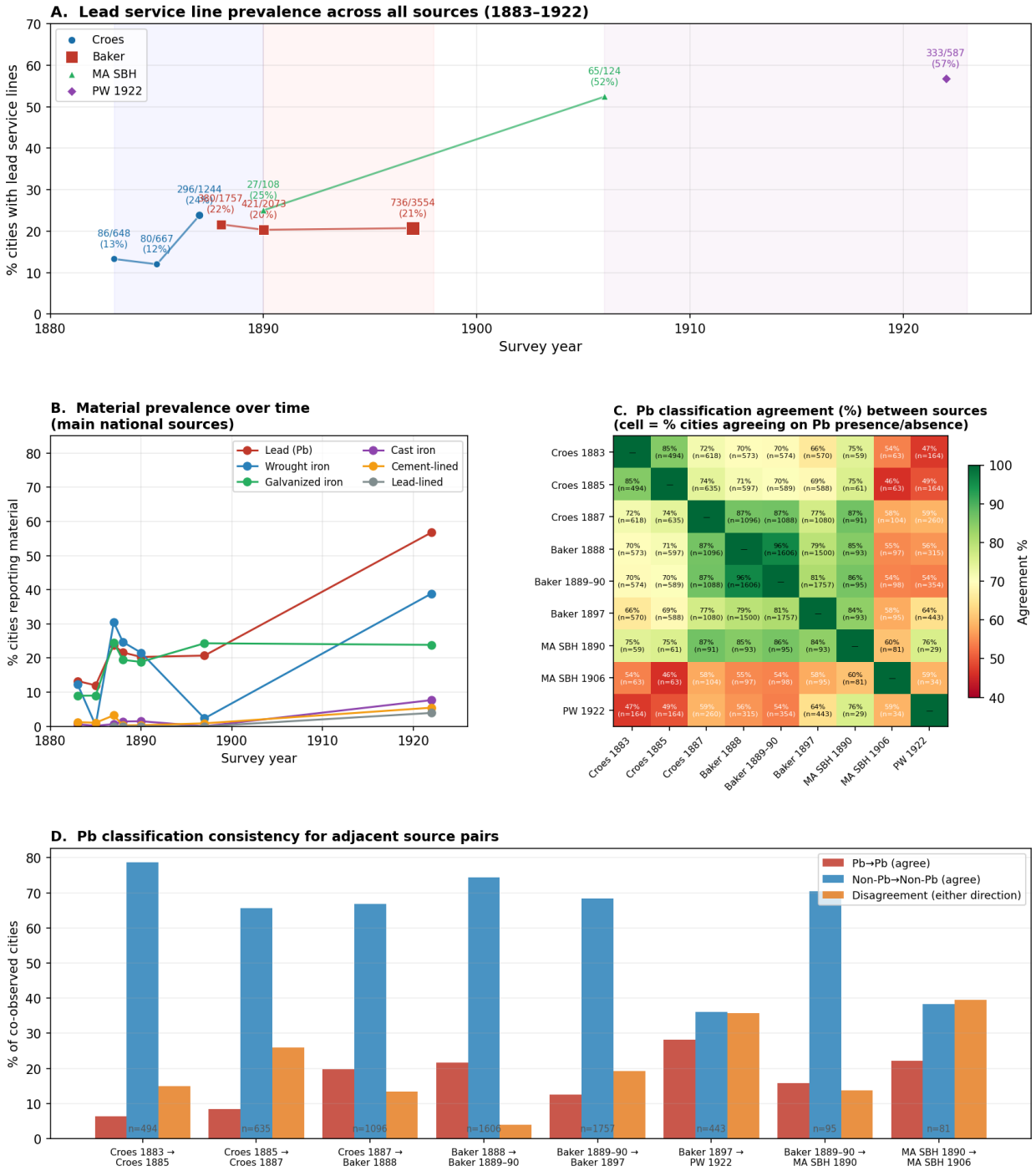
Figure A1 analyzes the consistency of the service line material reports. Panel A shows an increasing prevalence of lead over time as cities switched from other materials to lead but seldom switched in the other direction. Panel B shows that the "non-lead" cities were also switching toward cast iron and away from galvanized iron. Panels C and D show pairwise agreement and disagreement for cities that appeared in multiple editions of the same source or in two different sources in close temporal proximity. The disagreements appear to reflect actual adoption of new materials, particularly lead, over the four decades for which we have data. We can document many of these changes through municipal plumbing codes.

In practical terms, this means that the data will tend to understate lead prevalence and cause attenuation bias in any measured lead exposure effect. In order to ensure that our lead exposure measure most accurately reflects the risk posed to the individuals in our data, we have prioritized the lead reports as: 1922 > 1906 > 1897 > 1889/90 > 1888 > 1887 > 1885 > 1883.

A.2 Description of Water Chemistry Data

We have data on the water supplies of 2,838 U.S. cities and towns from 1923 to 1984. For each place, we know most of the following characteristics: water source (specific surface-water or ground-water location from which water was drawn), method(s) of treatment, and the chemical analysis of the raw and treated water.

Figure A1: Consistency of service-line material reports across sources, 1883–1922

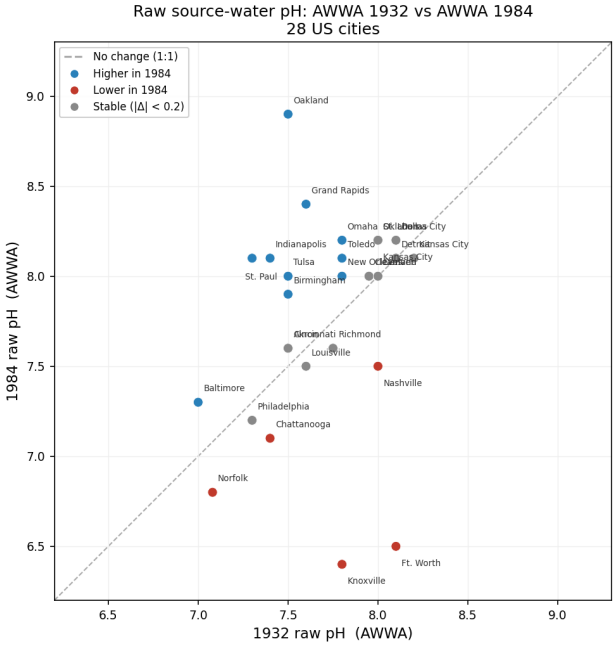


Notes: Panel A shows the share of cities reporting lead service lines by source and year. Panel B shows the share reporting each material by year. Panel C shows pairwise agreement on the presence or absence of lead for cities appearing in two sources. Panel D decomposes agreement and disagreement for adjacent source pairs. Sources are listed in Appendix A.1.

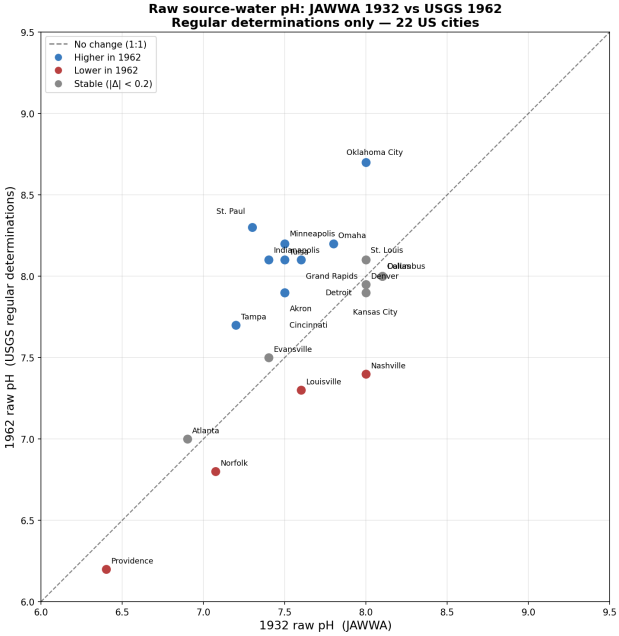
Figure A2 compares the pH of raw source water for (1) 28 U.S. cities over 52 years in two reports by the American Water Works Association (1932 and 1984) and (2) 22 U.S. cities over 30 years in both the 1932 AWWA report and a 1962 report by the U.S. Geological Survey.

Figure A2: Persistence of raw source-water *pH*

(a) 1932 to 1984 (28 cities)



(b) 1932 to 1962 (22 cities)



Notes: Panel (a) compares raw source-water *pH* for the 28 cities reported in both the 1932 and 1984 AWWA surveys. Panel (b) compares the 22 cities in both the 1932 AWWA survey and the 1962 USGS report. Points are colored by whether *pH* increased (blue), fell (red), or changed by less than 0.2 (gray) between the two readings. The dashed line is the 45-degree line.

The sources for water chemistry are:

1. Hardin, Eugene A. "Design and Operation Data on Large Rapid Sand Filtration Plants in the United States and Canada." Journal (American Water Works Association) 24, no. 8 (1932): 1190–207.
2. Hardin, Eugene A. "Design and Operation Data on Large Rapid Sand Filtration Plants in the U.S. and Canada." Journal (American Water Works Association) 34, no. 12 (1942): 1847–79.
3. Cosens, Kenneth W. "Design and Operation Data on Large Rapid Sand Filtration Plants in the United States and Canada." Journal (American Water Works Association) 48, no. 7 (1956): 819–53.
4. American Water Works Association. "1984 Water Utility Operating Data." Report 6 (unpublished, 1984).
5. Collins, William Dennis. The industrial utility of public water supplies in the United States. U.S. Geological Survey Water Supply Paper No. 496. U.S.Government Printing Office, 1923.
6. Collins, William Dennis, William Luther Lamar, and Edwin Wallace Lohr. The industrial utility of public water supplies in the United States, 1932. U.S. Geological Survey Water Supply Paper No. 658. U.S. Government Printing Office, 1934.

7. Lohr, Edwin Wallace, and Samuel Kenneth Love. The Industrial Utility of Public Water Supplies in the United States, 1952: States east of the Mississippi River. U.S. Geological Survey Water Supply Paper No. 1299. U.S. Government Printing Office, 1954.
8. Lohr, Edwin Wallace, and Samuel Kenneth Love. The Industrial Utility of Public Water Supplies in the United States, 1952: States west of the Mississippi River. U.S. Geological Survey Water Supply Paper No. 1300. U.S. Government Printing Office, 1954.
9. Durfor, Charles Norman, and Edith Becker. Public water supplies of the 100 largest cities in the United States, 1962. U.S. Geological Survey Water Supply Paper No. 1812. U.S. Government Printing Office, 1964.
10. Massachusetts, Department of Public Health. "Condition of Public Water Supplies." Twenty-Sixth Annual Report (1940): 168-181.
11. Lamar, William Luther. Industrial quality of public water supplies in Georgia, 1940. U.S. Geological Survey Water Supply Paper No. 912. U.S. Government Printing Office, 1942.
12. Sundstrom, Raymond W., and Warren W. Hastings. Public water supplies in eastern Texas. U.S. Geological Survey Water Supply Paper No. 1047. U.S. Government Printing Office, 1948.
13. Broadhurst, William Leslie, Raymond W. Sundstrom, and D. E. Weaver. Public water supplies in western Texas. U.S. Geological Survey Water Supply Paper No. 1106. U.S. Government Printing Office, 1951.
14. Broadhurst, William Leslie, Raymond W. Sundstrom, and J. H. Rowley. Public water supplies in southern Texas. U.S. Geological Survey Water Supply Paper No. 1070. US Government Printing Office, 1950.
15. Sundstrom, Raymond W., William Leslie Broadhurst, and B. C. Dwyer. Public water supplies in central and north-central Texas. U.S. Geological Survey Water Supply Paper No. 1069. US Government Printing Office, 1949.

The AWWA reports and the MDPH report contain chemistry only for raw source water. The AWWA reports are based on measurements reported by each city's water authority, while the MDPH reports are based on samples analyzed by the state laboratory. The USGS reports contain chemistry for both raw and treated water from samples analyzed by the USGS. They also report for many cities "regular determinations" of water chemistry, both raw and treated, based on measurements reported by each city's water authority.

In our assignment of pH values to each city's water supply, we have prioritized the available measures as: "regular determinations" > USGS or MDPH analysis, and 1984 > 1962 > 1956 > 1952 > 1948-1951 > 1942 > 1940 > 1932 > 1923. The first ranking reflects our assessment of the precision of the measurements: assessment made over an extended period of time will be better indicators of the overall lead exposure risk faced by individuals than assessment made on the basis of a sample drawn on a single day. The second ranking reflects our desire to have measurements closest in time to when the individuals in our

data were under age 5. Where the indicated treatment of the water supply did not significantly alter the pH (e.g. aeration, filtration) we categorized the reported pH as both raw and treated. Where multiple sources were reported for a city we used an unweighted average of the source values.

A.3 Description of Propensity Score Analysis

We construct the propensity score using census information from 1900. We identify covariates to include as follows: First, we force linear effects of log-population, share black, share urban, share literate, and occupational income to be incorporated. Second, from the remaining possible variables (share foreign, share renting, share with a mortgage, share over the age of 65, share under the age of 21, county-level fertility rate, and county-level manufacturing capital expenditures, average wages, employment share, and output-per-worker), we iteratively add linear effects to the model by searching for the largest marginal increase in the log-likelihood subject to the constraint that the added variable is significant with a likelihood ratio statistic above 1.0.[36] The coefficients from these propensity score models are shown in Appendix Table A2. We focus on two propensity score models in the main analysis: (a) the probability of having lead piping, estimated over the universe of lead and other piping, and (b) the probability of having extreme pH, estimated over the universe of lead pipes. These are shown in columns (1) and (3) of Appendix Table A2, respectively.[37]

Appendix Tables A3 and A4 show inverse-propensity score weighting adjusts for observable differences between treatment and control cities. Appendix Table A3 shows that, among cities with lead pipes, extreme pH tends to be more common in urban areas with a higher penetration of manufacturing as measured by capital expenditures, population share, and wages. Adjusting for these observable differences results in complete balance on these observed covariates, with a joint test for balance resulting in an F-statistic of 0.7.

Appendix Table A4 shows that, among cities with lead and other piping, cities with lead piping have higher baseline education levels as measured by literacy rates and share of children in school, have occupations associated with higher incomes, have a smaller share of the population who is black, and have a higher penetration of manufacturing. However, they are not significantly more urban or have a larger population. Again, linear probability models show that inverse-propensity score weighting mostly adjusts for these observable differences, with a joint F-statistic of 1.1.

Owing to our small sample sizes, we opt not to trim the propensity score models and present baseline results using the largest possible sample. We present in the Appendix results that show our baseline results are robust to trimming and controlling for deciles of the propensity score.[38] In some robustness analyses that focus on smaller samples or alternate definitions of extreme pH we are not able to estimate

[36] This procedure follows from Section 13.3 of Imbens and Rubin [2015].

[37] We estimate these models using all pipes with reliability of 2 or less. We extrapolate this propensity score based on covariates for other reliability levels. In particular, we have trouble getting our propensity score models to converge using only the most reliable piping data due to the small number of cities with lead piping with extreme pH measures.

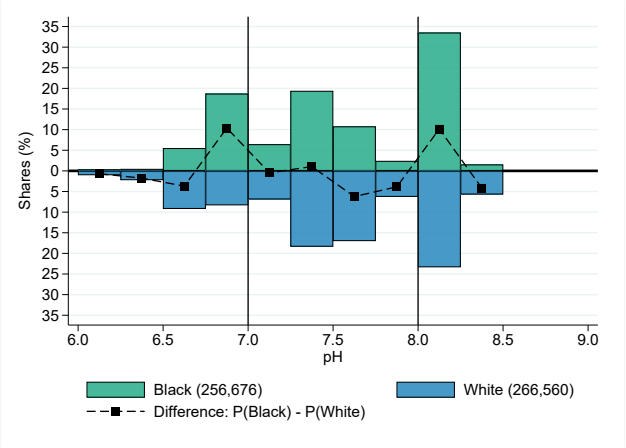
[38] Specifically, these results are shown in Appendix Tables A8 and A9.

the propensity score. For example, there are only 19 cities with lead pipes and a pH measure outside of $(6.75, 8.25)$, as shown in Appendix Table A13. The fact that our samples are balanced indicates that any issues related to overlap (or lack thereof) are not significant enough to generate residual lack of balance.

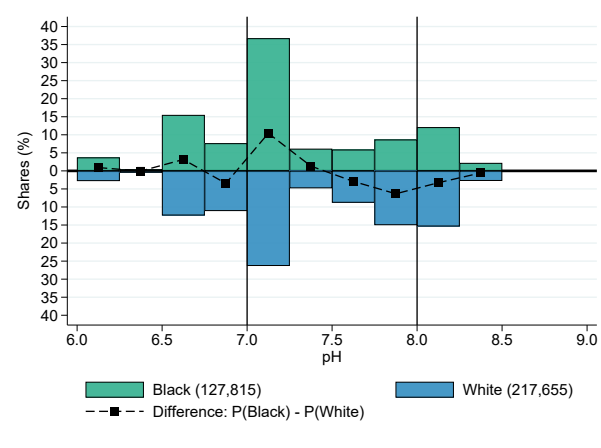
A.4 Appendix Figures and Tables

Figure A3: Distribution of pH and piping by race

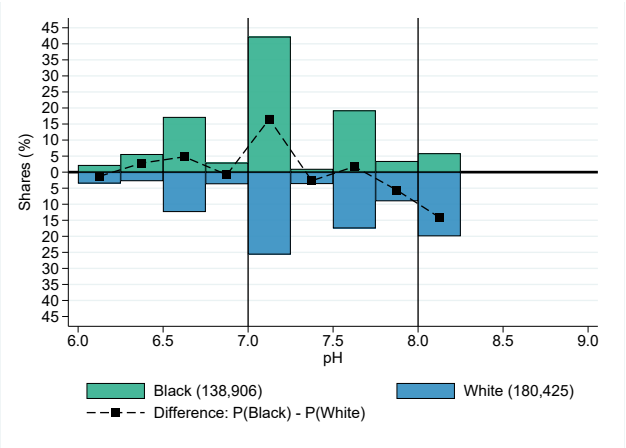
(a) Lead pipes



(b) Other pipes



(c) Mixed pipes



(d) By piping type

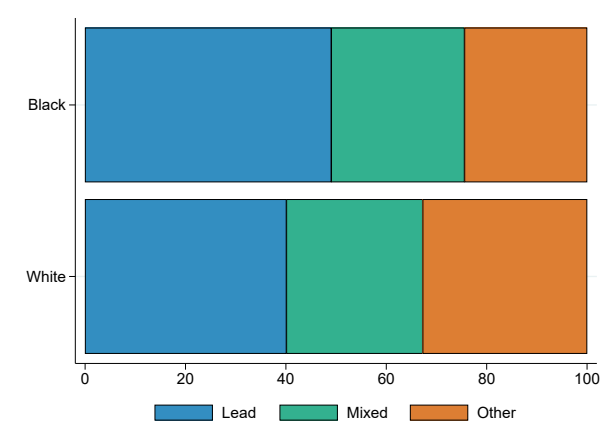


Figure A4: Distribution of piping types over predicted probability of lead

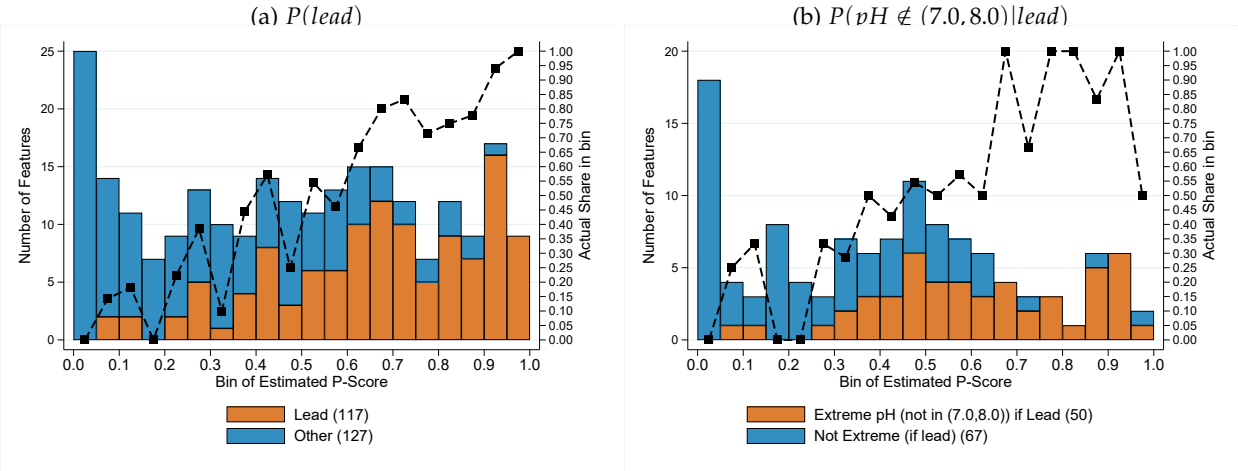
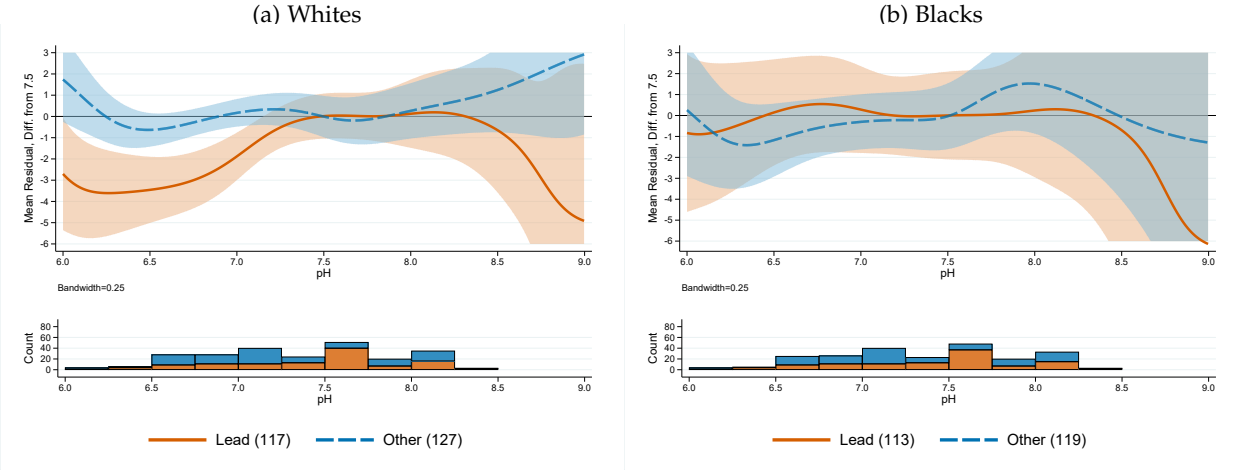
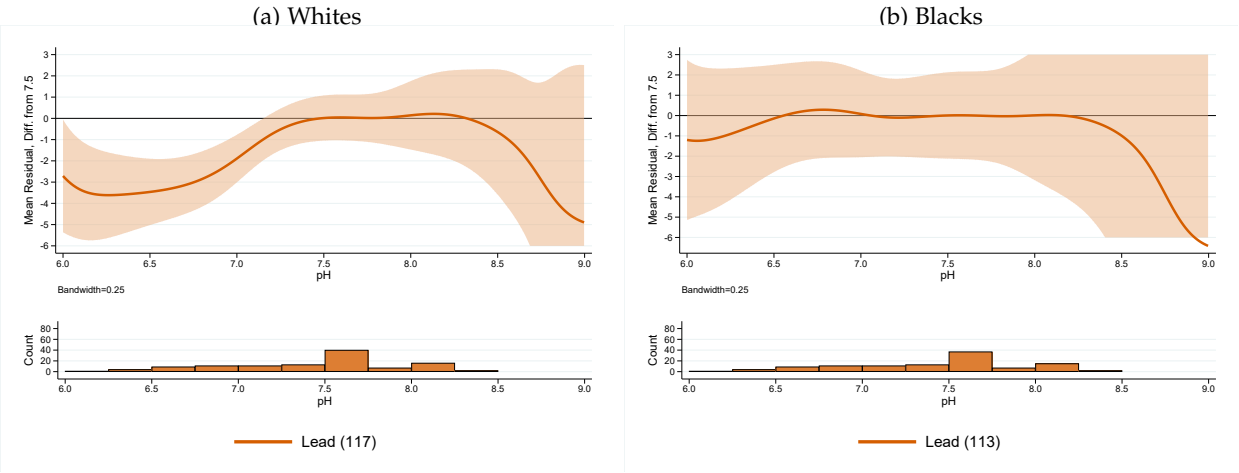


Figure A5: Non-parametric relationship by race, Unweighted



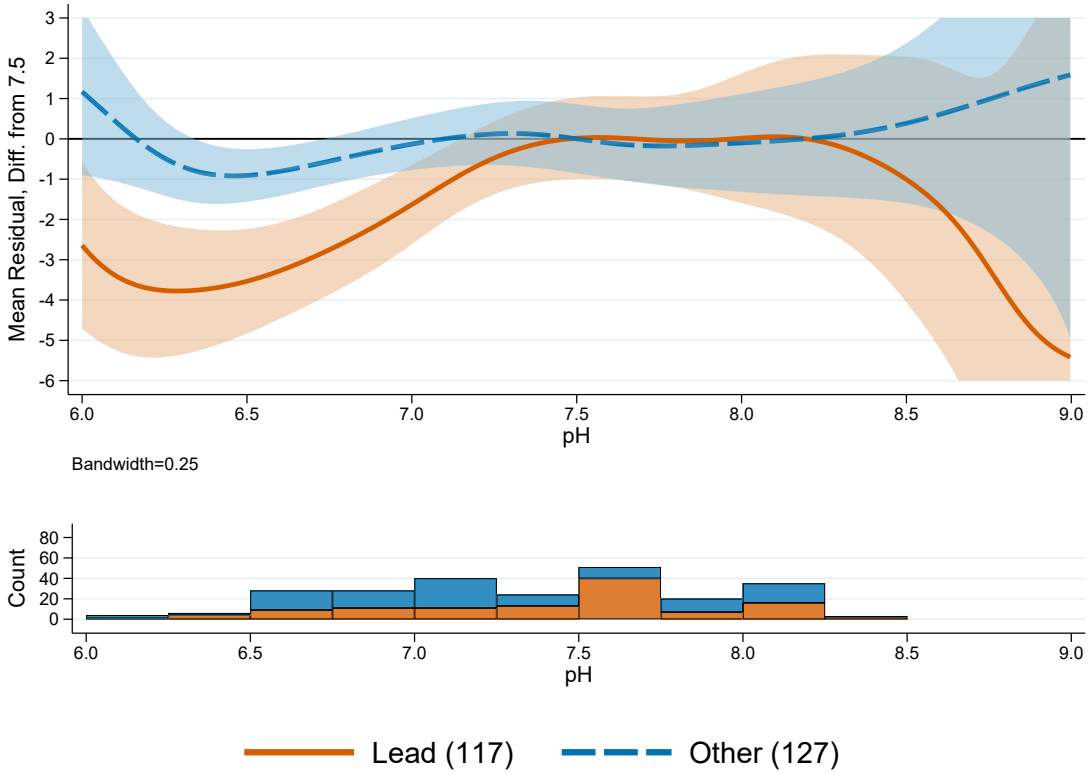
Notes: See Figure 3 for the main results. See Figure A6 for weighted estimates.

Figure A6: Non-parametric relationship by race, Weighted



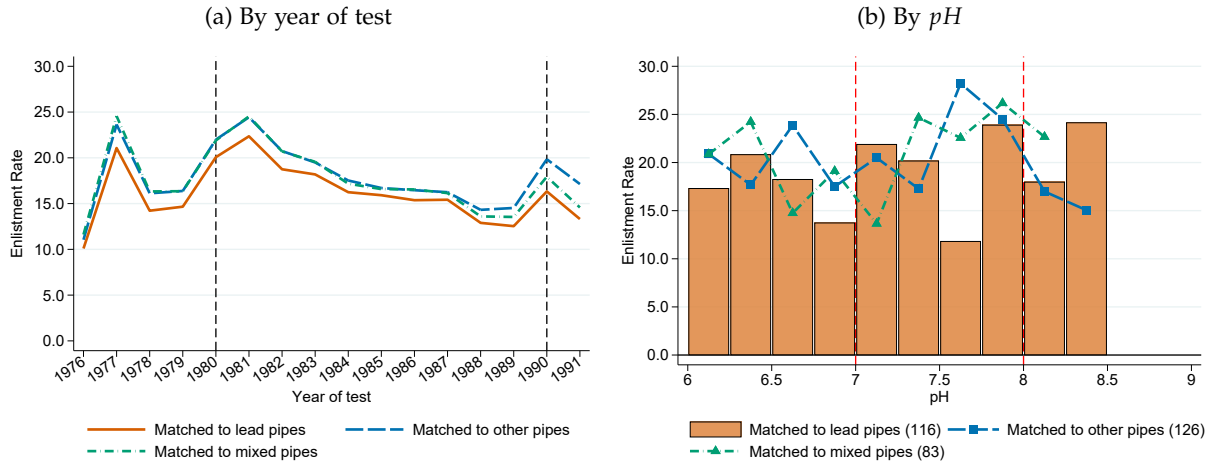
Notes: IPW weights constructed using $P(Extreme|Lead)$. See Figure A5 for unweighted results.

Figure A7: Non-parametric relationship, IPW: Lead and Other Pipes



Notes: IPW weights constructed using $\hat{P}(\text{Lead})$, analogous to column (4) of Table 4. See notes to Figure 3

Figure A8: Selection into AFQT Test Taking Sample



Notes: Propensity to take the AFQT is computed at the county-level using the number of 18 year old men in a county using the 1970, 1980, and 1990 census. We compute the enlistment rate for the years 1976 to 1991 using linear interpolation between census years. Average enlistment rates by piping type are computed by weighting across counties using the imputed number of 18 year old men. Multiple pH readings can take place in multiple counties. These are duplicates in this figure, but the results are similar when looking at counties with only one pH reading.

Table A1: 1990 census places and pH readings by piping type

	All pipes		*Lead pipes*		*Mixed pipes*		*Other pipes*	
	City	*pH*	City	*pH*	City	*pH*	City	*pH*
Highest pH readings								
1	Marietta, OH	9.3	Saginaw, MI	9.0	Marietta, OH	9.3	Redlands, CA	8.7
2	Saginaw, MI	9.0	Springfield, IL	8.4	San Angelo, TX	8.2	Orlando, FL	8.4
3	Redlands, CA	8.7	South Holland, IL	8.3	Pueblo, CO	8.2	Winchester, KY	8.2
4	Springfield, IL	8.4	Saint Paul, MN	8.3	Los Angeles, CA	8.2	Rock Springs, WY	8.2
5	Orlando, FL	8.4	Denver, CO	8.3	Lima, OH	8.1	Pocatello, ID	8.2
Middle pH readings								
1	Adrian, MI	7.4	East Rutherford, NJ	7.5	Butte, MT	7.4	Demopolis, AL	7.3
2	Yorkville, IL	7.3	Dumont, NJ	7.5	Adrian, MI	7.4	Decatur, AL	7.3
Median	Wausau, WI	7.3	DeKalb, IL	7.5	Rome, GA	7.3	Boise, ID	7.3
4	Victoria, TX	7.3	Connersville, IN	7.5	Piqua, OH	7.3	Alexandria, LA	7.3
5	Trenton, NJ	7.3	Cleveland, OH	7.5	Murfreesboro, TN	7.3	Aberdeen, WA	7.3
Lowest pH readings								
1	New Brunswick, NJ	5.7	Manchester, NH	6.2	Lebanon, PA	6.0	Newark, DE	6.1
2	Brockton, MA	5.7	Cranston, RI	6.2	Bethlehem, PA	6.0	Fayetteville, NC	6.0
3	Sumter, SC	5.2	Camden, NJ	6.0	New Brunswick, NJ	5.7	Tuscaloosa, AL	5.9
4	McKeesport, PA	4.6	Weymouth, MA	5.8	Brockton, MA	5.7	Greenwich, CT	5.9
5	Altoona, PA	3.5	McKeesport, PA	4.6	Sumter, SC	5.2	Altoona, PA	3.5

Table A2: Propensity score estimates by pH and lead piping

	(1)	(2)	(3)	(4)
		Extreme pH:	Extreme pH	Extreme pH
	Lead pipes	pH $\notin (7.0, 8.0)$	if Lead	and Lead
Log(Population)	-0.014	-0.020	0.006	0.002
	(0.014)	(0.019)	(0.018)	(0.013)
Black (%)	0.010***	-0.007*	-0.014**	-0.002
	(0.003)	(0.004)	(0.006)	(0.003)
Urban (%)	0.001	-0.000	0.001	0.001*
	(0.001)	(0.001)	(0.001)	(0.001)
Literacy (%)	0.026***	-0.006	0.008	0.016***
	(0.005)	(0.006)	(0.010)	(0.005)
Non-Farmer Occ. Income, County, (100s 2015 Dollars)	0.007***	-0.004*	-0.005*	-0.001
	(0.002)	(0.003)	(0.003)	(0.002)
Renters (%), County		0.006	0.009	0.006*
		(0.004)	(0.005)	(0.003)
Mortgage if Non-renter (%), County	0.006**	-0.010***	-0.009**	-0.003
	(0.003)	(0.003)	(0.004)	(0.003)
Age 0-21 (%)		0.006		
		(0.005)		
School (% $\geq$ 6 mos.), County	0.008***			0.005*
	(0.003)			(0.003)
Fertility Rate, County		0.191*	0.325**	0.145**
		(0.113)	(0.156)	(0.071)
Manuf. Capital Expenditures, County		0.001*	0.001	0.000**
		(0.001)	(0.001)	(0.000)
Manuf. Average Wages, County	-0.002***	0.001		
	(0.000)	(0.000)		
Manuf. Establishments / 1000, County	-0.032**	0.026	0.030	
	(0.013)	(0.018)	(0.025)	
Manuf. Population Share, County		0.019***	0.017*	0.009*
		(0.007)	(0.009)	(0.005)
Manuf. Output-per-worker, County	0.000***	-0.000*		
	(0.000)	(0.000)		
Number of Features	244	244	117	244
Lead	117	117	117	117
Other	127	127	0	127
Counties	193	193	82	193
Pseudo-R^2	0.309	0.176	0.290	0.239
Adj. Pseudo-R^2	0.244	0.088	0.140	0.142

Notes: Probit marginal effects. Columns (1), (2), and (4) use the sample for all lead and other pipes; column (3) shows results only for lead pipes. Covariates to be included are selected using an iterative procedure that iteratively adds main and interaction effects for each variable based on sequential increases in the probit log-likelihood. Linear effects for log-population, share black, share urban, share literate, and occupational income are forced to be included in each specification. Then, we sequentially include additional linear effects that give the largest increase in the log-likelihood subject to an increase in the likelihood ratio test statistic of greater than 1. Standard errors are clustered at the county-level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Balance by pH-level for lead pipes, $\hat{P}(Extreme|Lead)$

	Unweighted, Full Sample			*IPW*		
	Extreme pH: $pH \notin (7.0, 8.0)$	*Normal pH:* $pH \in (7.0, 8.0)$	Test	*Extreme pH:* $pH \notin (7.0, 8.0)$	*Normal pH:* $pH \in (7.0, 8.0)$	Test
Urban (%)	89.2	70.7	2.77***	85.0	81.9	0.48
	(3.9)	(5.4)		(4.6)	(4.6)	
Age 65+ (%)	3.5	4.0	-1.78*	3.8	3.5	0.81
	(0.2)	(0.2)		(0.2)	(0.2)	
Age 0-21 (%)	41.2	40.5	0.48	40.8	41.2	-0.27
	(1.3)	(1.0)		(1.2)	(0.9)	
Literacy (%)	93.9	92.7	1.16	93.1	93.6	-0.47
	(0.7)	(0.7)		(0.8)	(0.6)	
School (% $\geq$ 6 mos.), County	80.6	80.5	0.04	80.0	81.9	-1.21
	(1.3)	(1.1)		(1.2)	(0.9)	
Renters (%), County	60.8	54.7	2.51**	59.4	58.0	0.55
	(2.1)	(1.4)		(2.1)	(1.4)	
Non-Farmer Occ. Income, County, (100s 2015 Dollars)	222.2	217.4	1.30	218.9	220.5	-0.45
	(2.8)	(2.4)		(2.8)	(2.3)	
Manuf. Establishments / 1000, County	8.3	7.2	2.60**	7.7	7.3	1.08
	(0.4)	(0.2)		(0.4)	(0.2)	
Manuf. Capital Expenditures, County	105.6	25.7	3.68***	70.1	95.8	-1.04
	(20.7)	(6.5)		(16.4)	(18.4)	
Manuf. Population Share, County	13.1	8.9	3.71***	11.6	11.4	0.19
	(1.0)	(0.6)		(0.9)	(0.7)	
Manuf. Average Wages, County	450.5	426.3	2.01**	440.8	446.5	-0.42
	(9.9)	(6.9)		(10.9)	(8.3)	
Manuf. Output-per-worker, County	2721.5	2473.0	1.11	2557.9	2734.8	-0.80
	(151.7)	(164.7)		(140.1)	(169.3)	
Log(Population)	9.7	8.9	1.67*	9.8	9.2	1.27
	(0.4)	(0.3)		(0.4)	(0.2)	
Black (%)	4.1	7.7	-1.78*	7.0	5.2	0.79
	(1.2)	(1.6)		(1.8)	(1.4)	
Foreign born (%)	0.1	0.1	0.34	0.1	0.1	-0.07
	(0.1)	(0.1)		(0.1)	(0.1)	
Mortgage if Non-renter (%), County	40.0	42.2	-0.93	38.1	40.9	-1.30
	(1.6)	(1.6)		(1.7)	(1.4)	
Fertility Rate, County	2.3	2.2	1.76*	2.3	2.3	-0.86
	(0.0)	(0.0)		(0.0)	(0.0)	
Joint F-test from OLS			7.2***			0.7
N (Features)	50	67		50	67	

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Balance by type of piping, $\hat{P}(Lead)$

	Unweighted, Full Sample			*IPW*		
	Lead pipes	*Other pipes*	Test	*Lead pipes*	*Other pipes*	Test
Urban (%)	78.6	77.2	0.27	80.3	81.2	-0.20
	(3.6)	(3.3)		(3.4)	(3.0)	
Age 65+ (%)	3.8	4.3	-1.19	3.9	4.1	-0.56
	(0.1)	(0.4)		(0.1)	(0.3)	
Age 0-21 (%)	40.8	42.5	-1.76*	41.2	42.0	-1.01
	(0.8)	(0.6)		(0.7)	(0.5)	
Literacy (%)	93.2	87.9	5.75***	91.1	89.9	1.21
	(0.5)	(0.8)		(0.6)	(0.7)	
School (% $\geq$ 6 mos.), County	80.6	64.3	7.17***	76.1	71.8	2.02**
	(0.8)	(2.1)		(1.2)	(1.8)	
Renters (%), County	57.3	54.0	2.04**	57.2	56.2	0.57
	(1.2)	(1.1)		(1.3)	(1.0)	
Non-Farmer Occ. Income, County, (100s 2015 Dollars)	219.4	196.1	7.75***	210.7	206.6	1.25
	(1.8)	(2.4)		(2.2)	(2.3)	
Manuf. Establishments / 1000, County	7.7	6.8	2.56**	7.4	7.3	0.16
	(0.2)	(0.3)		(0.2)	(0.3)	
Manuf. Capital Expenditures, County	59.9	20.9	3.42***	42.4	54.6	-0.86
	(10.2)	(5.1)		(8.4)	(11.5)	
Manuf. Population Share, County	10.7	8.7	2.47**	9.5	10.0	-0.61
	(0.6)	(0.6)		(0.6)	(0.6)	
Manuf. Average Wages, County	436.6	392.7	3.85***	426.6	413.2	1.26
	(5.9)	(9.8)		(6.3)	(8.6)	
Manuf. Output-per-worker, County	2579.2	2095.8	3.74***	2354.1	2276.7	0.66
	(114.6)	(59.5)		(94.2)	(70.9)	
Log(Population)	9.3	9.2	0.46	9.5	9.5	-0.07
	(0.2)	(0.1)		(0.2)	(0.1)	
Black (%)	6.1	15.3	-4.64***	11.4	12.5	-0.49
	(1.1)	(1.6)		(1.6)	(1.4)	
Foreign born (%)	0.1	0.3	-1.14	0.1	0.2	-0.56
	(0.1)	(0.1)		(0.1)	(0.1)	
Mortgage if Non-renter (%), County	41.2	29.9	6.89***	35.8	34.5	0.77
	(1.2)	(1.2)		(1.3)	(1.2)	
Fertility Rate, County	2.3	2.6	-5.66***	2.4	2.5	-1.91*
	(0.0)	(0.1)		(0.0)	(0.0)	
Joint F-test from OLS			18.7***			1.1
N (Features)	117	127		117	127	

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Interactions: Lead pipes only

	Unweighted				*IPW*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		By pH	By Race	By Poverty		By pH	By Race	By Poverty
D(Extreme pH)	-1.788***	-1.866***	-2.097***	-1.547***	-1.286**	-1.142*	-1.026	-1.035*
	(0.461)	(0.546)	(0.679)	(0.525)	(0.537)	(0.595)	(0.736)	(0.584)
× D(pH ≥ 8.0)		0.204				-0.351		
		(0.800)				(0.898)		
× D(Black)			0.824				-0.725	
			(1.342)				(1.471)	
× D(High 1970 Poverty)				-0.256				-0.716
				(0.901)				(0.857)
D(Black)	-18.72***	-18.72***	-19.18***	-18.70***	-19.17***	-19.17***	-18.77***	-19.15***
	(0.799)	(0.799)	(0.541)	(0.797)	(0.850)	(0.852)	(0.524)	(0.851)
D(Other Non-White)	-14.17***	-14.18***	-14.05***	-14.14***	-14.61***	-14.60***	-14.69***	-14.56***
	(0.392)	(0.393)	(0.445)	(0.390)	(0.586)	(0.586)	(0.630)	(0.590)
log(Feature pop.), 1990	-0.285*	-0.279*	-0.284*	-0.251*	-0.188	-0.200	-0.187	-0.134
	(0.150)	(0.148)	(0.147)	(0.143)	(0.196)	(0.191)	(0.195)	(0.177)
D(High 1970 Poverty)				-0.801				-0.583
				(0.766)				(0.721)
Piping type	Lead	Lead	Lead	Lead	Lead	Lead	Lead	Lead
Weights	None	None	None	None	$\hat{P}$(Extreme|Lead)	$\hat{P}$(Extreme|Lead)	$\hat{P}$(Extreme|Lead)	$\hat{P}$(Extreme|Lead)
Controls								
Birth year FE	X	X	X	X	X	X	X	X
State FE	X	X	X	X	X	X	X	X
Survey FE	X	X	X	X	X	X	X	X
N(Obs.)	550,420	550,420	550,420	550,420	550,420	550,420	550,420	550,420
N(Features)	117	117	117	117	117	117	117	117
Extreme × Lead	50	50	50	50	50	50	50	50
Adjusted R^2	0.211	0.211	0.211	0.211	0.216	0.216	0.216	0.217
Outcome Mean	40.8	40.8	40.8	40.8	40.7	40.7	40.7	40.7
Outcome S.D.	23.9	23.9	23.9	23.9	23.8	23.8	23.8	23.8

Notes: Standard errors are clustered at the feature-level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Robustness to Removing High-Migration Counties

	Lead piping only			*Lead and other piping*		
	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline	Exclude High 5-Year Mig.	Exclude High Birth State Mig.	Baseline	Exclude High 5-Year Mig.	Exclude High Birth State Mig.
D(Extreme pH)	-1.286**	-1.230**	-1.187**	0.444	0.281	0.345
	(0.537)	(0.533)	(0.539)	(0.644)	(0.764)	(0.737)
D(Lead pipes)				1.203**	1.318**	1.238**
				(0.528)	(0.593)	(0.563)
D(Lead pipes) × D(Extreme pH)				-1.819**	-1.432	-1.737*
				(0.833)	(0.925)	(0.937)
D(Black)	-19.17***	-18.98***	-19.21***	-19.91***	-19.64***	-19.90***
	(0.850)	(0.832)	(0.913)	(0.395)	(0.418)	(0.419)
D(Other Non-White)	-14.61***	-14.96***	-14.99***	-13.79***	-14.13***	-14.26***
	(0.586)	(0.510)	(0.525)	(0.485)	(0.465)	(0.441)
log(Feature pop.), 1990	-0.188	-0.176	-0.119	0.0303	-0.0578	0.0543
	(0.196)	(0.197)	(0.208)	(0.145)	(0.154)	(0.153)
Piping type	Lead Only	Lead Only	Lead Only	Other, Lead	Other, Lead	Other, Lead
Weights	$\hat{P}$(Extreme|Lead)	$\hat{P}$(Extreme|Lead)	$\hat{P}$(Extreme|Lead)	$\hat{P}$(Lead)	$\hat{P}$(Lead)	$\hat{P}$(Lead)
Controls						
Birth year FE	X	X	X	X	X	X
State FE	X	X	X	X	X	X
Survey FE	X	X	X	X	X	X
N(Obs.)	550,420	509,504	514,666	904,554	789,123	808,940
N(Features)	117	108	90	244	215	205
Extreme × Lead	50	47	46	50	47	46
Adjusted R^2	0.216	0.212	0.217	0.225	0.220	0.225
Outcome Mean	40.7	40.2	40.3	41.9	41.1	41.6
Outcome S.D.	23.8	23.7	23.7	24.2	24.0	24.1

Notes: Columns (1) and (4) reproduce the baseline estimates. Columns (2) and (5) exclude counties in the top quartile of the share of residents who lived in a different county five years earlier. Columns (3) and (6) exclude counties in the top quartile of the share of residents born in a different state. Both measures come from 1980 Census county tables. Standard errors are clustered at the feature level.

Table A7: Sample Size and Lead Exposure: Feature-Level

	Lead Only			*Lead and Other*		
	(1) No Controls	(2) Population	(3) All Controls	(4) No Controls	(5) Population	(6) All Controls
D(Lead pipes)				-0.459*	-0.0721	-0.0762
				(0.258)	(0.0666)	(0.106)
D(Extreme pH)	0.858***	-0.0351	-0.105	-0.182	0.00501	-0.140
	(0.290)	(0.0711)	(0.0832)	(0.211)	(0.0633)	(0.124)
D(Lead) × D(Extreme pH)				1.041***	-0.0169	0.121
				(0.358)	(0.0924)	(0.137)
ln(Feature pop, 1990)		0.763***	0.742***		0.743***	0.740***
		(0.0183)	(0.0219)		(0.0173)	(0.0256)
Piping type	Lead	Lead	Lead	Other, Lead	Other, Lead	Other, Lead
Weights	None	None	$\hat{P}$(Extreme|Lead)	None	None	$\hat{P}$(Lead)
State FE			X			X
N(Features)	117	117	117	244	244	244
Adjusted R^2	0.060	0.958	0.964	0.034	0.941	0.926

Notes: Shows feature-level regressions where the dependent variable is the log-number of AFQT test takers. Columns (1) and (4) show no controls, columns (2) and (5) show only controls for log-population, and Columns (3) and (6) include our baseline controls and weight by the listed propensity scores. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Robustness to propensity score/trimming: Lead pipes only

	(1) Unweighted	(2) IPW, No Trim	(3) IPW, 5% Trim	(4) IPW, 10% Trim	(5) P-Score Deciles
D(Extreme pH)	-1.788***	-1.286**	-1.436**	-1.492**	-1.023*
	(0.461)	(0.537)	(0.549)	(0.601)	(0.517)
D(Black)	-18.72***	-19.17***	-18.87***	-18.42***	-18.76***
	(0.799)	(0.850)	(0.765)	(0.700)	(0.807)
D(Other Non-White)	-14.17***	-14.61***	-13.97***	-13.81***	-14.13***
	(0.392)	(0.586)	(0.923)	(0.938)	(0.383)
log(Feature pop.), 1990	-0.285*	-0.188	-0.315	-0.359*	-0.208
	(0.150)	(0.196)	(0.203)	(0.204)	(0.185)
Piping type	Lead	Lead	Lead	Lead	Lead
Controls					
Birth year FE	X	X	X	X	X
State FE	X	X	X	X	X
Survey FE	X	X	X	X	X
N(Obs.)	550,420	550,420	470,478	433,917	550,420
N(Features)	117	117	97	87	117
Extreme × Lead	50	50	49	42	50
Adjusted R^2	0.211	0.216	0.216	0.211	0.211
Outcome Mean	40.8	40.7	41.0	41.3	40.8
Outcome S.D.	23.9	23.8	23.9	24.0	23.9

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Robustness to propensity score/trimming: Lead and Other pipes

	(1) Unweighted	(2) IPW, No Trim	(3) IPW, 5% Trim	(4) IPW, 10% Trim	(5) P-Score Deciles
D(Extreme pH)	0.00638	0.444	0.468	0.169	0.341
	(0.367)	(0.644)	(0.667)	(0.467)	(0.372)
D(Lead pipes)	0.979***	1.203**	1.269**	0.812	1.254***
	(0.374)	(0.528)	(0.553)	(0.497)	(0.450)
D(Extreme pH) × D(Lead Pipes)	-1.725***	-1.819**	-1.847**	-1.025	-1.957***
	(0.518)	(0.833)	(0.841)	(0.701)	(0.495)
D(Black)	-19.51***	-19.91***	-19.94***	-20.24***	-19.50***
	(0.547)	(0.395)	(0.421)	(0.287)	(0.551)
D(Other Non-White)	-13.65***	-13.79***	-13.88***	-13.52***	-13.63***
	(0.472)	(0.485)	(0.487)	(0.507)	(0.467)
log(Feature pop.), 1990	-0.0259	0.0303	-0.00361	0.230*	0.0251
	(0.117)	(0.145)	(0.147)	(0.138)	(0.139)
Piping type	Other, Lead	Other, Lead	Other, Lead	Other, Lead	Other, Lead
Controls					
Birth year FE	X	X	X	X	X
State FE	X	X	X	X	X
Survey FE	X	X	X	X	X
N(Obs.)	904,554	904,554	832,427	756,433	904,554
N(Features)	244	244	210	179	244
Extreme × Lead	50	50	47	40	50
Adjusted R^2	0.221	0.225	0.224	0.229	0.221
Outcome Mean	42.0	41.9	41.9	41.9	42.0
Outcome S.D.	24.2	24.2	24.2	24.3	24.2

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A10: Robustness to geographic merging, IPW

	Lead piping only				*Lead and other piping*			
	(1) Baseline	(2) All Zips	(3) Interior Zips	(4) 10 % of Area	(5) Baseline	(6) All Zips	(7) Interior Zips	(8) 10 % of Area
D(Extreme pH)	-1.286**	-1.230**	1.906	-1.225**	0.444	0.409	1.071	0.347
	(0.537)	(0.547)	(1.267)	(0.559)	(0.644)	(0.647)	(0.683)	(0.676)
D(Lead pipes)					1.203**	1.179**	0.719	1.138**
					(0.528)	(0.532)	(0.718)	(0.542)
× D(Extreme pH)					-1.819**	-1.760**	-1.160	-1.628*
					(0.833)	(0.839)	(1.065)	(0.855)
D(Black)	-19.17***	-19.12***	-17.56***	-19.09***	-19.91***	-19.88***	-18.47***	-19.85***
	(0.850)	(0.836)	(1.006)	(0.877)	(0.395)	(0.393)	(0.721)	(0.413)
D(Other Non-White)	-14.61***	-14.60***	-13.75***	-14.65***	-13.79***	-13.80***	-13.73***	-13.83***
	(0.586)	(0.586)	(0.624)	(0.548)	(0.485)	(0.483)	(0.481)	(0.470)
log(Feature pop.), 1990	-0.188	-0.174	-1.128***	-0.208	0.0303	0.0290	-0.727***	-0.00848
	(0.196)	(0.199)	(0.369)	(0.200)	(0.145)	(0.146)	(0.191)	(0.148)
Weights	$\hat{P}$(Extreme|Lead)	$\hat{P}$(Extreme|Lead)	$\hat{P}$(Extreme|Lead)	$\hat{P}$(Extreme|Lead)	$\hat{P}$(Lead)	$\hat{P}$(Lead)	$\hat{P}$(Lead)	$\hat{P}$(Lead)
N(Obs.)	550,420	560,853	339,841	531,622	904,554	915,157	455,460	868,288
N(Features)	117	117	71	117	244	244	121	244
Extreme × Lead	50	50	28	50	50	50	28	50
Adjusted R^2	0.216	0.217	0.197	0.217	0.225	0.225	0.213	0.226
Outcome Mean	40.7	40.6	37.4	40.4	41.9	41.9	38.5	41.6
Outcome S.D.	23.8	23.8	22.9	23.8	24.2	24.2	23.5	24.2

Notes: This table assesses the robustness of our baseline results to different approaches to merging zip codes with 1990 places. Zip codes do not merge perfectly with places and some zip codes are linked to multiple places. Our baseline approach removes the multiple-linked zip codes from the analysis, whereas "All Zips" includes them by allocating a zip code the 1990 place which has the largest coverage of the zip codes area. Our baseline approach uses any zip code that has at least 1% of its area overlapping with the area of the 1990 place. "Interior Zips" instead only use zip codes that lie fully within the zip code. "10% of Area" changes this threshold to 10% from 1%. Standard errors are clustered at the feature-level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: Robustness to piping sample and reliability measure, IPW

	Lead piping only			*Lead and other piping*			*Lead, other, and mixed piping*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
D(Extreme pH)	-0.684	-1.286**	-2.391**	0.243	0.444	2.777***	-0.121	0.251	4.048***
	(0.518)	(0.537)	(1.102)	(0.500)	(0.644)	(0.614)	(0.460)	(0.445)	(1.028)
D(Lead pipes)				-0.0157	1.203**	4.534***	-0.457	0.624	5.322***
				(0.684)	(0.528)	(0.895)	(0.605)	(0.452)	(0.873)
× D(Extreme pH)				-0.794	-1.819**	-5.045***	-0.636	-1.557**	-6.415***
				(0.762)	(0.833)	(0.928)	(0.704)	(0.647)	(1.356)
D(Mixed pipes)							-1.151**	-0.574	1.443**
							(0.556)	(0.527)	(0.566)
× D(Extreme pH)							1.759**	1.710**	-0.164
							(0.774)	(0.737)	(0.979)
D(Black)	-19.84***	-19.17***	-17.74***	-20.23***	-19.91***	-18.58***	-19.51***	-19.13***	-17.07***
	(0.639)	(0.850)	(1.055)	(0.335)	(0.395)	(0.890)	(0.386)	(0.473)	(0.984)
D(Other Non-White)	-14.83***	-14.61***	-13.78***	-13.26***	-13.79***	-13.83***	-13.56***	-13.93***	-12.88***
	(0.536)	(0.586)	(0.407)	(0.460)	(0.485)	(0.468)	(0.312)	(0.322)	(0.503)
log(Feature pop.), 1990	-0.308*	-0.188	-0.696**	-0.0119	0.0303	-0.660***	-0.109	-0.00535	-0.819***
	(0.179)	(0.196)	(0.302)	(0.132)	(0.145)	(0.184)	(0.104)	(0.111)	(0.180)
Reliability groups	1,2,3	1,2	1	1,2,3	1,2	1	1,2,3	1,2	1
Propensity weights	$\hat{P}$(Extreme|Lead)	$\hat{P}$(Extreme|Lead)	$\hat{P}$(Extreme|Lead)	$\hat{P}$(Lead)	$\hat{P}$(Lead)	$\hat{P}$(Lead)	$\hat{P}$(Lead)	$\hat{P}$(Lead)	$\hat{P}$(Lead)
N(Obs.)	664,649	550,420	324,938	1,243,471	904,554	380,786	1,687,928	1,233,616	506,022
N(Features)	146	117	57	352	244	81	457	325	103
Extreme × Lead	60	50	21	60	50	21	60	50	21
Adjusted R^2	0.220	0.216	0.198	0.221	0.225	0.207	0.211	0.211	0.190
Outcome Mean	41.7	40.7	39.1	43.3	41.9	40.3	43.3	42.5	40.6
Outcome S.D.	24.2	23.8	23.3	24.4	24.2	23.8	24.4	24.2	23.6

Notes: Propensity scores weights are estimated using groups 1 and 2 (the baseline) and uses the propensity score model to construct weights for the pooled analysis on groups 1 to 3 and the group 1-only analysis. Standard errors are clustered at the feature-level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A12: Robustness to piping sample and reliability measure, no IPW

	Lead piping only			*Lead and other piping*			*Lead, other, and mixed piping*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
D(Extreme pH)	-0.979*	-1.788***	-2.641***	-0.344	0.00638	2.974***	-0.128	0.276	4.324***
	(0.529)	(0.461)	(0.848)	(0.402)	(0.367)	(0.947)	(0.393)	(0.400)	(1.042)
D(Lead pipes)				-0.0392	0.979***	4.732***	-0.216	0.549	5.547***
				(0.607)	(0.374)	(0.980)	(0.578)	(0.419)	(0.887)
× D(Extreme pH)				-0.517	-1.725***	-5.542***	-0.579	-1.556***	-6.702***
				(0.661)	(0.518)	(1.204)	(0.642)	(0.541)	(1.368)
D(Mixed pipes)							-0.846	-0.548	1.577**
							(0.522)	(0.517)	(0.619)
× D(Extreme pH)							1.091	1.062	-0.420
							(0.739)	(0.754)	(1.001)
D(Black)	-19.02***	-18.72***	-17.78***	-20.00***	-19.51***	-18.03***	-19.58***	-19.04***	-16.64***
	(0.704)	(0.799)	(1.161)	(0.448)	(0.547)	(1.031)	(0.456)	(0.572)	(1.105)
D(Other Non-White)	-14.26***	-14.17***	-13.67***	-13.34***	-13.65***	-13.54***	-13.52***	-13.88***	-12.74***
	(0.407)	(0.392)	(0.425)	(0.480)	(0.472)	(0.460)	(0.299)	(0.326)	(0.542)
log(Feature pop.), 1990	-0.442***	-0.285*	-0.681***	-0.105	-0.0259	-0.669***	-0.160	-0.0537	-0.904***
	(0.162)	(0.150)	(0.239)	(0.115)	(0.117)	(0.198)	(0.102)	(0.115)	(0.191)
Reliability groups	1,2,3	1,2	1	1,2,3	1,2	1	1,2,3	1,2	1
Propensity weights	None	None	None	None	None	None	None	None	None
N(Obs.)	665,675	550,420	324,938	1,259,960	904,554	380,786	1,733,394	1,249,485	506,022
N(Features)	147	117	57	356	244	81	465	327	103
Extreme × Lead	61	50	21	61	50	21	61	50	21
Adjusted R^2	0.214	0.211	0.198	0.222	0.221	0.202	0.217	0.215	0.188
Outcome Mean	41.6	40.8	39.2	43.2	42.0	40.0	43.0	42.0	40.4
Outcome S.D.	24.1	23.9	23.3	24.4	24.2	23.6	24.4	24.1	23.6

Notes: Standard errors are clustered at the feature-level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: Robustness to alternate pH cutoffs: Lead pipes only

	(7.25,7.75)		(7,8)		(7,8), Drop (6.75,7]U[8,8.25)		(6.75,8.25)		(6.75,8)		(7,8.25)	
	(1) Unweighted	(2) IPW	(3) Unweighted	(4) IPW	(5) Unweighted	(6) IPW	(7) Unweighted	(8) IPW	(9) Unweighted	(10) IPW	(11) Unweighted	(12) IPW
D(Extreme pH)	-1.436***	-1.002**	-1.788***	-1.286**	-2.297***	-1.531*	-2.158***	-1.403*	-1.702***	-1.362**	-2.006***	-1.623***
	(0.447)	(0.398)	(0.461)	(0.537)	(0.726)	(0.796)	(0.653)	(0.766)	(0.573)	(0.604)	(0.504)	(0.500)
D(Black)	-18.73***	-18.75***	-18.72***	-19.17***	-19.24***	-19.01***	-18.71***	-18.82***	-18.72***	-19.36***	-18.71***	-18.57***
	(0.800)	(0.627)	(0.799)	(0.850)	(0.440)	(0.451)	(0.799)	(0.645)	(0.799)	(0.928)	(0.798)	(1.163)
D(Other Non-White)	-14.19***	-14.28***	-14.17***	-14.61***	-13.95***	-14.30***	-14.24***	-14.40***	-14.19***	-14.78***	-14.21***	-14.61***
	(0.392)	(0.496)	(0.392)	(0.586)	(0.753)	(1.112)	(0.391)	(0.441)	(0.390)	(0.487)	(0.393)	(0.472)
log(Feature pop.), 1990	-0.352**	-0.322**	-0.285*	-0.188	-0.0911	0.0665	-0.449***	-0.432**	-0.397**	-0.341*	-0.320**	-0.337*
	(0.151)	(0.145)	(0.150)	(0.196)	(0.249)	(0.358)	(0.156)	(0.172)	(0.163)	(0.188)	(0.142)	(0.190)
Piping type	Lead	Lead	Lead	Lead	Lead	Lead	Lead	Lead	Lead	Lead	Lead	Lead
Controls												
Birth year FE	X	X	X	X	X	X	X	X	X	X	X	X
State FE	X	X	X	X	X	X	X	X	X	X	X	X
Survey FE	X	X	X	X	X	X	X	X	X	X	X	X
N(Obs.)	550,420	550,420	550,420	550,420	307,110	307,110	550,420	550,420	550,420	550,420	550,420	550,420
N(Features)	117	117	117	117	86	86	117	117	117	117	117	117
Extreme × Lead	64	64	50	50	19	19	19	19	35	35	34	34
Adjusted R^2	0.210	0.211	0.211	0.216	0.201	0.194	0.210	0.205	0.210	0.232	0.211	0.211
Outcome Mean	40.8	40.9	40.8	40.7	42.6	42.8	40.8	41.6	40.8	40.3	40.8	40.4
Outcome S.D.	23.9	23.8	23.9	23.8	24.1	24.2	23.9	24.1	23.9	24.1	23.9	23.7

Notes: Each pair of columns redefines our definition of normal pH (i.e., not extreme) on the baseline sample of lead pipes, re-estimating the propensity score to target these alternate treatments. The column "(7,8), Drop (6.75,7]U[8,8.25)" uses the same propensity score as "(7,8)".* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A14: Robustness to alternate pH cutoffs: Lead and Other pipes only

	(7.25,7.75)		(7,8)		(7,8), Drop (6.75,7]U[8,8.25)		(6.75,8.25)		(6.75,8)		(7,8.25)	
	(1) Unweighted	(2) IPW	(3) Unweighted	(4) IPW	(5) Unweighted	(6) IPW	(7) Unweighted	(8) IPW	(9) Unweighted	(10) IPW	(11) Unweighted	(12) IPW
D(Lead pipes)	1.094*	1.738*	0.979***	1.203**	0.725*	0.607	0.390	0.349	0.580*	0.639	0.539	0.552
	(0.622)	(0.946)	(0.374)	(0.528)	(0.388)	(0.465)	(0.328)	(0.376)	(0.337)	(0.423)	(0.344)	(0.424)
D(Extreme pH)	0.188	1.054	0.00638	0.444	-0.308	-0.490	-0.426	-0.685	-0.103	0.424	-0.441	-0.616
	(0.619)	(0.991)	(0.367)	(0.644)	(0.487)	(0.615)	(0.394)	(0.464)	(0.370)	(0.743)	(0.353)	(0.408)
D(Extreme pH) × D(Lead pipes)	-1.426**	-2.016*	-1.725***	-1.819**	-1.576**	-0.397	-1.311**	-0.294	-1.121**	-0.960	-1.426***	-0.934
	(0.704)	(1.059)	(0.518)	(0.833)	(0.717)	(0.854)	(0.640)	(0.710)	(0.564)	(0.910)	(0.515)	(0.646)
D(Black)	-19.51***	-19.87***	-19.51***	-19.91***	-19.81***	-19.99***	-19.51***	-19.89***	-19.52***	-19.92***	-19.49***	-19.88***
	(0.548)	(0.394)	(0.547)	(0.395)	(0.325)	(0.319)	(0.549)	(0.398)	(0.549)	(0.400)	(0.548)	(0.399)
D(Other Non-White)	-13.68***	-13.80***	-13.65***	-13.79***	-13.44***	-13.73***	-13.71***	-13.80***	-13.69***	-13.82***	-13.66***	-13.76***
	(0.471)	(0.483)	(0.472)	(0.485)	(0.687)	(0.680)	(0.468)	(0.485)	(0.473)	(0.485)	(0.465)	(0.477)
log(Feature pop.), 1990	-0.0859	-0.0548	-0.0259	0.0303	-0.0176	0.0528	-0.145	-0.0611	-0.0970	-0.0339	-0.0819	-0.0435
	(0.118)	(0.135)	(0.117)	(0.145)	(0.174)	(0.185)	(0.124)	(0.149)	(0.132)	(0.150)	(0.109)	(0.139)
Piping type	Lead,Other	Lead,Other	Lead,Other	Lead,Other	Lead,Other	Lead,Other	Lead,Other	Lead,Other	Lead,Other	Lead,Other	Lead,Other	Lead,Other
Controls												
Birth year FE	X	X	X	X	X	X	X	X	X	X	X	X
State FE	X	X	X	X	X	X	X	X	X	X	X	X
Survey FE	X	X	X	X	X	X	X	X	X	X	X	X
N(Obs.)	904,554	904,554	904,554	904,554	534,062	534,062	904,554	904,554	904,554	904,554	904,554	904,554
N(Features)	244	244	244	244	162	162	244	244	244	244	244	244
Extreme × Lead	64	64	50	50	19	19	19	19	35	35	34	34
Adjusted R^2	0.221	0.225	0.221	0.225	0.215	0.222	0.221	0.224	0.221	0.224	0.221	0.225
Outcome Mean	42.0	41.9	42.0	41.9	43.0	42.4	42.0	41.9	42.0	41.9	42.0	41.9
Outcome S.D.	24.2	24.2	24.2	24.2	24.3	24.2	24.2	24.2	24.2	24.2	24.2	24.2

Notes: Each pair of columns redefines our definition of normal pH (i.e., not extreme) on the baseline sample of lead and other pipes, re-estimating the propensity score to target these alternate treatments. The column "(7,8), Drop (6.75,7]U[8,8.25)" uses the same propensity score as "(7,8)". * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A15: Robustness to polynomial in pH: Lead pipes only

	Unweighted				*IPW*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
pH^1	-0.463	0.526	0.615	0.920	-0.855	0.188	1.434	1.386*
	(0.489)	(0.575)	(0.840)	(0.763)	(0.534)	(0.618)	(0.870)	(0.712)
pH^2	-0.961***	-1.636***	-1.543*	-4.671***	-0.593	-1.086***	0.178	-3.253***
	(0.341)	(0.268)	(0.899)	(1.323)	(0.387)	(0.309)	(0.780)	(1.216)
pH^3		-0.515***	-0.585	-2.331***		-0.551**	-1.458***	-2.770***
		(0.129)	(0.560)	(0.881)		(0.245)	(0.517)	(0.495)
pH^4			-0.0325	1.572**			-0.460**	1.177**
			(0.257)	(0.608)			(0.228)	(0.500)
pH^5				0.634**				0.592***
				(0.246)				(0.161)
D(Black)	-18.70***	-18.71***	-18.71***	-18.72***	-19.18***	-19.18***	-19.15***	-19.16***
	(0.799)	(0.799)	(0.799)	(0.799)	(0.854)	(0.853)	(0.852)	(0.852)
D(Other Non-White)	-14.18***	-14.19***	-14.18***	-14.19***	-14.57***	-14.58***	-14.56***	-14.58***
	(0.387)	(0.389)	(0.389)	(0.392)	(0.580)	(0.583)	(0.583)	(0.587)
log(Feature pop.), 1990	-0.475***	-0.433***	-0.432***	-0.383**	-0.327	-0.303	-0.262	-0.226
	(0.159)	(0.152)	(0.150)	(0.150)	(0.204)	(0.195)	(0.184)	(0.185)
Implied Effect:								
6.0 v. 7.5	-1.467	-2.732***	-2.585	-0.877	-0.050	-0.868	0.841	1.415
	(1.055)	(1.020)	(1.821)	(1.405)	(1.155)	(1.203)	(1.470)	(0.950)
6.5 v. 7.5	-0.498	-1.647**	-1.606*	-2.322***	0.263	-0.724	-0.258	-1.284*
	(0.593)	(0.692)	(0.849)	(0.793)	(0.644)	(0.690)	(0.699)	(0.721)
8.5 v. 7.5	-1.424**	-1.625***	-1.545**	-3.875***	-1.448**	-1.449***	-0.307	-2.868***
	(0.601)	(0.291)	(0.745)	(0.955)	(0.674)	(0.360)	(0.771)	(0.960)
9.0 v. 7.5	-2.857***	-4.630***	-4.686***	-4.222***	-2.616**	-4.020***	-4.702***	-4.136***
	(1.070)	(0.396)	(0.461)	(0.308)	(1.210)	(0.486)	(0.543)	(0.369)
Piping type	Lead	Lead	Lead	Lead	Lead	Lead	Lead	Lead
Weights	None	None	None	None	$\hat{P}$(Extreme|Lead)	$\hat{P}$(Extreme|Lead)	$\hat{P}$(Extreme|Lead)	$\hat{P}$(Extreme|Lead)
Controls								
Birth year FE	X	X	X	X	X	X	X	X
State FE	X	X	X	X	X	X	X	X
Survey FE	X	X	X	X	X	X	X	X
N(Obs.)	550,420	550,420	550,420	550,420	550,420	550,420	550,420	550,420
N(Features)	117	117	117	117	117	117	117	117
Adjusted R^2	0.210	0.211	0.211	0.211	0.216	0.217	0.217	0.217
Outcome Mean	40.8	40.8	40.8	40.8	40.7	40.7	40.7	40.7
Outcome S.D.	23.9	23.9	23.9	23.9	23.8	23.8	23.8	23.8

Notes: pH measures are normalized to be relative to 7.5, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A16: Robustness to polynomial in pH: Lead and Other pipes

	Unweighted				*IPW*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
pH^1	0.512	0.952*	-1.124	-1.054	0.854	0.926	-1.225	-1.386
	(0.458)	(0.534)	(0.846)	(0.839)	(0.562)	(0.604)	(1.024)	(0.979)
pH^2	0.331	-0.505	-2.169**	-2.403	0.554	0.752	-0.459	2.608
	(0.373)	(1.059)	(1.019)	(2.860)	(0.536)	(1.646)	(1.794)	(5.107)
pH^3		-0.707	4.079***	3.852**		0.147	5.326**	6.677**
		(0.753)	(1.352)	(1.887)		(1.101)	(2.131)	(2.830)
pH^4			3.363***	3.674			3.486***	-0.687
			(0.962)	(3.113)			(1.279)	(5.534)
pH^5				0.228				-2.484
				(1.792)				(3.168)
D(Lead pipes)	0.485	0.552	0.372	0.686	0.549	0.760	0.521	1.145
	(0.346)	(0.383)	(0.368)	(0.490)	(0.393)	(0.497)	(0.476)	(0.756)
Lead × pH^1	-0.529	-0.110	1.659*	1.987**	-0.680	-0.194	2.012*	2.459**
	(0.563)	(0.609)	(0.935)	(0.959)	(0.743)	(0.781)	(1.201)	(1.180)
Lead × pH^2	-1.223**	-1.143	0.292	-2.215	-1.179*	-2.072	-0.603	-6.047
	(0.572)	(1.064)	(1.189)	(3.068)	(0.685)	(1.604)	(1.788)	(5.232)
Lead × pH^3		0.132	-4.475***	-5.939***		-0.676	-6.062***	-8.818***
		(0.752)	(1.349)	(2.077)		(1.093)	(2.119)	(2.924)
Lead × pH^4			-3.287***	-2.164			-3.598***	1.839
			(0.935)	(3.155)			(1.248)	(5.566)
Lead × pH^5				0.355				2.991
				(1.803)				(3.178)
D(Black)	-19.49***	-19.49***	-19.49***	-19.49***	-19.88***	-19.87***	-19.83***	-19.83***
	(0.550)	(0.549)	(0.549)	(0.550)	(0.397)	(0.397)	(0.398)	(0.399)
D(Other Non-White)	-13.68***	-13.67***	-13.66***	-13.66***	-13.80***	-13.79***	-13.75***	-13.77***
	(0.469)	(0.468)	(0.468)	(0.468)	(0.482)	(0.479)	(0.476)	(0.475)
log(Feature pop.), 1990	-0.148	-0.132	-0.148	-0.110	-0.0624	-0.0638	-0.0881	-0.0549
	(0.124)	(0.118)	(0.117)	(0.116)	(0.147)	(0.140)	(0.137)	(0.130)
Implied Effect:								
6.0 v. 7.5	-1.960	-2.852***	-3.367*	-1.571	-1.633	-2.088*	-2.131	-0.934
	(1.211)	(1.011)	(1.906)	(1.977)	(1.301)	(1.249)	(1.914)	(1.846)
6.5 v. 7.5	-0.695	-1.165	-0.179	-0.782	-0.499	-1.202	-0.151	-0.839
	(0.614)	(0.718)	(0.903)	(0.863)	(0.672)	(0.942)	(1.013)	(1.117)
8.5 v. 7.5	-1.752*	-1.120	-5.810***	-7.976***	-1.859	-2.942	-8.251**	-7.575***
	(0.955)	(1.703)	(1.471)	(1.711)	(1.261)	(2.624)	(3.204)	(2.789)
9.0 v. 7.5	-3.546*	-2.289	-28.595***	-30.306	-3.673	-7.234	-37.014***	-7.652
	(1.810)	(4.621)	(6.907)	(19.223)	(2.352)	(7.034)	(12.146)	(34.371)
Piping type	Other, Lead	Other, Lead	Other, Lead	Other, Lead	Other, Lead	Other, Lead	Other, Lead	Other, Lead
Weights	None	None	None	None	$\hat{P}(\text{Lead})$	$\hat{P}(\text{Lead})$	$\hat{P}(\text{Lead})$	$\hat{P}(\text{Lead})$
Controls								
Birth year FE	X	X	X	X	X	X	X	X
State FE	X	X	X	X	X	X	X	X
Survey FE	X	X	X	X	X	X	X	X
N(Obs.)	904,554	904,554	904,554	904,554	904,554	904,554	904,554	904,554
N(Features)	244	244	244	244	244	244	244	244
Adjusted R^2	0.221	0.221	0.221	0.221	0.224	0.225	0.225	0.225
Outcome Mean	42.0	42.0	42.0	42.0	41.9	41.9	41.9	41.9
Outcome S.D.	24.2	24.2	24.2	24.2	24.2	24.2	24.2	24.2

Notes: pH measures are normalized to be relative to 7.5, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A17: AFQT in NLSY

	Neal and Johnson Replication		*Full NLSY79*			
	(1)	(2)	(3)	(4)	(5)	(6)
Z_{AFQT}		0.200***	0.178***	0.162***	0.158***	0.130***
		(0.017)	(0.010)	(0.006)	(0.006)	(0.007)
Z^2_{AFQT}		-0.043**				
		(0.017)				
D(Black)	-0.239***	-0.039				
	(0.035)	(0.037)				
D(Hispanic)	-0.129***	0.008				
	(0.043)	(0.044)				
Age	0.073***	0.047**				
	(0.020)	(0.019)				
D(College)						0.137***
						(0.018)
Years Used	1990, 1991	1990, 1991	1991	1979-2022	1979-2022	1979-2022
Birth year FE					X	X
Survey Year FE					X	X
N(Obs.)	1,534	1,534	3,480	92,777	92,777	92,777
N(Individuals)	1,534	1,534	3,480	5,782	5,782	5,782
Adjusted R^2	0.035	0.142	0.116	0.073	0.254	0.260

Notes: Columns (1) and (2) replicate Neal and Johnson's (1996) methodology: only test-takers born after 1961 are used, the outcome is the log of the average of non-missing 1990 and 1991 wages measured in 1991 dollars, and wage-observations $<$ \$1 or $>$\$75 are discarded. AFQT is studentized using this subsample of the NLSY. Columns (4)-(6) use all birth years and survey years with non-missing test and wage data. Wages are measured in 2025 dollars. Like Neal and Johnson, we discard wage-observations $<$\$1 or $>$\$130. AFQT in this sample is studentized using sample moments from our full sample of AFQT test takers (column 1 of Table 1). "D(College)" is measured as ever completing college across sample rounds. The sample is restricted to men older than 18 who have completed at least 8th grade with non-missing education, AFQT score and earnings across all columns. In all regressions, year-specific sample weights are used and robust standard errors are clustered at the individual-level.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A18: Present Value of AFQT in NLSY

	(1) Baseline	(2) Add Age Restriction	(3) Add Zeros Restriction
AFQT Perc.	181.8	185.2	134.4
	(12.1)	(15.1)	(17.0)
$\times$ Age	68.9	74.6	77.6
	(3.8)	(4.7)	(5.4)
$\times$ Age^2/100	-83.0	-99.4	-112.2
	(12.7)	(14.4)	(16.5)
Implied PV Difference of AFQT Perc			
$r = 0.00$	44810.3	45811.3	42922.9
	(2060.1)	(2272.7)	(2700.3)
$r = 0.03$	10630.1	10896.2	10052.4
	(462.5)	(521.5)	(617.9)
$r = 0.05$	4341.9	4448.6	4023.6
	(185.9)	(213.0)	(251.5)
Years Used	1979-2022	1979-2022	1979-2022
Year-by-Cohort Effects	X	X	X
N(Obs.)	117,706	78,200	55,420
N(Individuals)	5,889	2,877	2,014
Adjusted R^2	0.243	0.238	0.241

Notes: Shows associated impacts of an increase in AFQT score percentile on total earnings with interactions between AFQT and age profiles. Total earnings includes zeros for years where the respondent was interviewed. "Implied PV Difference" shows the differences in lifetime earnings between age 18 and 65 implied by the AFQT-by-age coefficients, discounted back to birth using a 3 percent discount rate, and computing standard errors using the Delta-method. Age coefficients are centered at age 25. Column (1) presents our baseline estimates from all observations. Column (2) restricts the sample to respondents who have one non-missing earnings observation in each 5-year band from age 20 through age 59. Column (3) extends this restriction to consider only respondents who have one non-missing and non-zero observation in each 5-year band from age 20 through age 59. The sample is restricted to men older than 18 who have completed at least 8th grade with non-missing education, AFQT score and earnings across all columns. In all regressions, year-specific sample weights are used and robust standard errors are clustered at the individual-level.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.