NBER WORKING PAPER SERIES

DRINKING WATER CONTAMINANT CONCENTRATIONS AND BIRTH OUTCOMES

Richard DiSalvo
Elaine L. Hill

Working Paper 31567
http://www.nber.org/papers/w31567

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2023

For generous feedback, we thank Andrew Boslett, Alina Denham, Dan Grossman, Lala Ma, Michelle Marcus, Steven Martin, David Slusky, Devin Sonne, Mary Willis, Alexis Zavez, participants at the iHEA 2017 and the APPAM 2017 conferences, and participants at the applied reading group seminar at University of Rochester. We thank Karla Zhendejas, Evan Volkin, Mira Chaskes, Max Morrow, and Grace Sventek for excellent research assistance. These data were supplied by the Bureau of Health Statistics & Research, Pennsylvania Department of Health, Harrisburg, Pennsylvania. The Pennsylvania Department of Health specifically disclaims responsibility for any analyses, interpretations or conclusions. Thank you to Amy Farrell and James Rubertone of Pennsylvania Department of Health for facilitating access to the data. This research was supported by the Office of the Director of the National Institutes of Health under award number DP5OD021338 (PI Hill). We also gratefully acknowledge funding from the University of Rochester Environmental Health Sciences Center (EHSC), an NIH/NIEHS-funded program (P30 ES001247). The study was reviewed and provided expedited determination by the University of Rochester RSRB. The authors have no relevant or material financial interests that relate to the research described in this paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research, NIH, or the New York City Independent Budget Office.

Drinking Water Contaminant Concentrations and Birth Outcomes
Richard DiSalvo and Elaine L. Hill
NBER Working Paper No. 31567
August 2023
JEL No. I1,Q53

ABSTRACT

Previous research in the US has found negative health effects of contamination when it triggers regulatory violations. An important question is whether levels of contamination that do not trigger a health-based violation impact health. We study the impact of drinking water contamination in community water systems on birth outcomes using drinking water sampling results data in Pennsylvania. We create an overall water quality index and an index specific to reproductive health. We focus on the effects of water contamination for births not exposed to regulatory violations. Our most rigorous specification employs mother fixed effects and finds changing from the 10th to the 90th percentile of water contamination (among births not exposed to regulatory violations) increases low birth weight by 12% and preterm birth by 17%.

Richard DiSalvo
New York City Independent Budget Office
110 William Street #1401
New York, NY 10038
rdisalv2@gmail.com

Elaine L. Hill
Department of Public Health Sciences
Department of Economics
University of Rochester
265 Crittenden Boulevard
Rochester, NY 14642
and NBER
elaine_hill@urmc.rochester.edu

A data appendix is available at http://www.nber.org/data-appendix/w31567

1 Introduction

Most residential settings in the US, including homes, apartments, and mobile park homes, are served by community water systems (CWS), representing 94% of the nation's population (US Environmental Protection Agency, 2019a). The Safe Drinking Water Act (SDWA) was enacted in 1974 to protect communities from unsafe drinking water in CWS. The act currently regulates over 90 contaminants (e.g., lead, arsenic, disinfection byproducts). Regulatory drinking water violations[1] under SDWA occur when a maximum contaminant level (MCL), action level, or maximum residual disinfectant level (MRDL) is exceeded (US Environmental Protection Agency, 2018).[2] Despite successfully reducing drinking water contamination in CWS, up to 45 million people were affected by health-based violations from 1982-2015 (Allaire et al., 2018). In 2017, 22 million people, representing 7 percent of the US population, relied on a CWS with a health-based drinking water violation (US Environmental Protection Agency, 2019a). Recent lead-related drinking water crises in Newark, New Jersey (Dave and Yang, 2022) and Flint, Michigan (Grossman and Slusky, 2019) have highlighted the importance of these regulations to protect public health.

The vast majority of the literature studying SDWA has focused on *non-compliant* drinking water. SDWA violations disproportionately impact minority and low-income communities (Allaire et al., 2018; McDonald and Jones, 2018), egregious cases of water contamination, such as the case of Flint, have major community and health impacts (Grossman and Slusky, 2017; Danagoulian and Jenkins, 2021; Danagoulian et al., 2022; Christensen et al., 2023), and health-based violations adversely impact infant health (Currie et al., 2013). Public reporting of these health-based violations is an important mechanism to encourage compliance within the law (Bennear and Olmstead, 2008; Bennear et al., 2009; Baker et al., 2022; Grooms, 2016). Prior research has shown that this public reporting has resulted in fewer violations over time (Bennear and Olmstead, 2008; Baker et al., 2022), however, other work using sampling concentrations suggests that systems may not return to compliance following health-based violations (Grooms, 2016). Consumers appear to respond to public reporting of violations with avoidance behavior, such as bottled water purchases (Zivin et al., 2011; Allaire et al., 2019; Marcus, 2022). In this paper, we provide new evidence of the health impacts of drinking water quality that is not subject to a health-based violation, which we call *compliant* drinking water contamination.[3]

[1]Throughout this paper, we use "regulatory violation", "MCL" and "MCL violation" interchangeably to mean any action level or health based violation, respectively, even for those where the violation is based upon an action level or MRDL.

[2]These thresholds are based on single contaminant concentrations, as opposed to mixtures of contaminants or co-occurring contaminants. Most contaminants have a MCL, but lead has an action level, and disinfectants have MRDL. Moreover, regulations vary by contaminant over whether and how samples are averaged to determine violations. Treatment technique (TT) violations also occur when the water system does not implement corrective actions in response to contamination.

[3]We discuss in the Discussion section issues that have been described in the literature regarding compliance,

A large and growing literature focused on the health impacts of air pollution that is compliant with the Clean Air Act has shown that even low levels of air quality can still have meaningful impacts on health, labor market outcomes, and cognition (Currie and Walker, 2019). Much like air quality, where particulate matter is actually a grouping of many air pollutants and not isolating particular pollutant constituents, we propose to measure this compliant contamination using a water quality index. Using an index addresses the multiple comparison issues we would face if we separately estimated the effects of each contaminant, or the multicollinearity issues we would face if we attempted to include the over 90 contaminants that are regulated under SDWA in the same regression.[4] Furthermore, while current toxicological and epidemiological studies of the associations between water quality and health generally reflect a narrow focus on single contaminants (Murphy et al., 2012), researchers have noted that low concentrations of contaminants may not be associated with health outcomes when considered individually but may produce effects when examined as mixtures (Gennings et al., 2018). For these reasons, we create an overall water quality index using contaminant concentrations and regulatory standards, creating a single measure for testing our hypothesis. Following prior literature, we also construct an aggregate "chemical only" measure (Currie et al., 2013). Our aggregation strategy is discussed further in Section 2.3.

Scientific literature assessing the association between each of the more than 90 contaminants regulated by SDWA and birth outcomes is lacking beyond a subset of contaminants.[5] Still, our overall water quality index may not be specific to our outcomes of interest (e.g., low birth weight, preterm birth), and so we create a second index specific to 5 contaminants with the most associational evidence linking exposure to these contaminants with the outcomes studied in this paper.[6] While a reproductive-specific index can be supportive of the theoretical link between water quality and birth outcomes, lack of empirical evidence for the other contaminants regulated by SDWA is not indicative of no biological effect. Our overall index can be used for hypothesis generation regarding other contaminants impacting birth outcomes. Furthermore, and we consider this most important, the overall measure can be validated on other populations (children, older adults) and health outcomes (cancer), and allow for a useful general index that can be used across systems and susceptible populations, analogous to the EPA's Air Quality Index (AQI).

We employ the universe of address-specific birth records in Pennsylvania from 2003-2014 geocoded to CWS boundaries, and match births to water samples measured at the CWS-level

enforcement, and how our results shed light on particular policy responses.

[4]We do include all 94 contaminants in this summary measure of water quality. See Table A2 for the full list of contaminants included.

[5]We discuss in Section 2.3.1 and the Appendix Section A1.3 the literature on drinking water and birth outcomes. Quasi-experimental literature is virtually non-existent except in the case of lead (Grossman and Slusky, 2019; Dave and Yang, 2022) or any MCL violation as in Currie et al. (2013).

[6]We discuss in Section 2.3.1 and Appendix Section A1.3 creation of this index. The five contaminants include arsenic, atrazine, nitrate, Di(2-ethylhexyl) Phthalate (DEHP), and Tetrachloroethene (PERC).

that underlie regulatory violations during gestation.[7] Using these novel data on contamination *concentrations* summarized in an omnibus index, we find evidence of precise harmful effects of water contamination on birth outcomes after removing regulatory violations. Our empirical strategy relies on within-CWS variation (CWS fixed effects) such that our approach ignores correlations between water quality and infant health driven by cross-sectional differences. We also include birth year-month fixed effects to control for seasonality and state-wide time shocks. Using fixed-effects models that isolate only within-mother across-sibling variation, we find that changes in overall contamination from the 10th to the 90th percentile – conditional on *no violation* – leads to an increase in low-birth weight (LBW) of 12% and preterm birth (PTB) of 17%. Using a reproductive-specific index, we find slightly larger impacts of water quality on our birth outcomes – a 19% increase in LBW and 22% increase in PTB. And in sensitivity analyses, we remove births with samples above the regulatory threshold, which is more stringent than how regulatory violations are determined, and continue to find a strong impact on birth outcomes.[8]

Given the recent findings of fertility effects of contamination and the consequent issues of selection (e.g., for lead, see Grossman and Slusky (2019) and Clay et al. (2021)), we additionally fit CWS panel models of births per person served by the system on various measures of recent water contamination. Our estimates suggest small negative effects of contamination on fertility. We also analyze whether mothers move out of their current CWS in response to contamination. While maternal response would be of concern for our identification strategy, we find no significant effects on mobility, which is consistent with mothers not being aware of variation in contamination that does not trigger regulatory violations. Moreover, our results are robust to excluding mothers who move across pregnancies. Finally, we use the bounding methods of Altonji et al. (2005) and Oster (2016) to assess the importance of selection on unobservables, finding in most specifications that it would have to be greater than selection on observables to conclude that there are no damaging effects of water contamination.

Our work contributes to a well-established fetal origins literature in economics that links shocks, such as pollution, experienced *in utero* to poorer birth outcomes (Almond et al., 2018). Extensive research in economics establishes that poor birth outcomes lead to worse short- and long-term health and human capital outcomes (see e.g. Black et al. (2007), Currie and Moretti (2007), Royer (2009), Oreopoulos et al. (2008) and Almond et al. (2005)). Air pollution has attracted particular attention, as findings of harmful effects appear robust, even at levels compliant with air advisory standards (Knittel et al., 2016; Currie and Walker, 2011; Currie et al., 2015; Yang and Chou, 2018; Beatty and Shimshack, 2014). Although federal and state regulation of public drinking water in

[7]Samples are taken at the system level prior to distribution and are not taken at the residence. Exceptions include lead and copper which are measured at a random selection of residences, however, it is a small subsample within the system and we do not have information on which residences.

[8]See Appendix Section A2.4.

the US has recently come under scrutiny, in part because of the highly publicized Flint water crisis (Grossman and Slusky (2019), Allaire et al. (2018)), few papers have studied how the quality of drinking water in the US affects health outcomes using quasi-experimental designs (Keiser and Shapiro, 2019b).[9] The most closely related study to our work is Currie et al. (2013), who find that exposure to a water quality *violation* during gestation increases the chance of low birth weight and preterm birth, based on 1997-2007 data from NJ.[10] Importantly, their estimates suggest that non-compliant contamination has negative health consequences. Our analysis addresses health impacts of compliant contamination levels.

Consumers currently learn about the quality of their water either by reviewing the annual Consumer Confidence Reports (CCR), which contain the average concentrations for contaminants in their water system, or through notifications about extreme contamination events that constitute regulatory violations. We argue that, based on the reports as currently designed, consumers are unlikely to respond to concentrations that are compliant with current regulations. Given our findings of adverse birth outcomes from overall contamination concentrations removing regulatory violations, using our index to describe these risks for consumers could greatly improve the available information. Specifically, we propose using our overall water quality index to create a letter grade as a summative measure of water quality that could be included in CCRs to provide more actionable information to consumers. In Section 6, we evaluate this potential policy finding that it could lead to economically meaningful improvements in public health as measured by birth outcomes. We find that a policy improving water quality for births served by systems with letter grades B-F to the average quality of systems with letter grade A would reduce the number of LBW infants by 3,960 and reduce the number of PTB infants by 6,930 over 11 years. Such a policy could save around $280 million in medical costs for these avoided adverse birth outcomes. This analysis could serve as a foundation for future refinements to providing a summary measure of water quality for CCRs.

2 Data and Summary Statistics

To conduct our analysis, we assembled data on birth outcomes from confidential birth records, temperature, weather, and pollution data derived from weather stations and Toxics Release Inventory (TRI) locations and emissions, and water quality and service areas from community water systems (CWS). We describe each of these in turn. We join these three data sources together using spatial and temporal attributes.

[9]Exceptions include Currie et al. (2013); Grossman and Slusky (2019); Marcus (2019); Hill and Ma (2022); Dave and Yang (2022); fly (2199).

[10]In the Online Appendix Section A3, we estimate similar models to Currie et al. (2013) using our data in Pennsylvania from 2003-2014 and find consistent, but not statistically significant, effects.

2.1 Birth Records and Control Variables

Data on birth certificates spanning the years 2003-2014 were provided by the Pennsylvania Department of Health. These data contain detailed information on birth outcomes, clinical characteristics and demographic characteristics. We describe in detail our control variables obtained from these birth records in Section 3. Our two main outcomes of interest are binary indicators for low birth weight[11] and preterm birth.[12] In addition, we study two birth outcome measures often used by epidemiologists that better capture growth rates during pregnancy (see e.g., Govarts et al. (2018)). These measures are small for gestational age (SGA) and term birth weight (TBW; in grams).[13]

We obtained daily weather statistics from the website of Schlenker (Schlenker and Roberts, 2009). These data provide estimates of daily precipitation and maximum and minimum temperatures for a grid of points across the United States. We match mothers to their closest grid point using exact location of birth, and calculate a set of temperature and weather controls over the mom's pregnancy. We use these as control variables in our regression specifications described in Section 3. Note that weather and temperature have been found to be important in previous studies relating health to drinking water contamination (Currie et al., 2013), as well as to air pollution (Beatty and Shimshack, 2014; Knittel et al., 2016). Finally, we also obtained Toxics Release Inventory (TRI) controls (locations and emissions) from the US Environmental Protection Agency (EPA) (US EPA, 2013). Using maternal address, we calculate the number of TRI facilities operating within 1, 3, and 5 kilometers in the year the infant was born, and the reported total onsite, offsite, and overall releases for TRI facilities weighted by the squared inverse distance between the facility and the mother's residential location.[14]

2.2 Community Water System Data

Our data on CWS comes from two sources: the US EPA and the Pennsylvania Department of Environmental Protection (PADEP).

We obtained Safe Drinking Water Information System (SDWIS) data from the EPA through a Freedom of Information Act request, which contains information on CWS drinking water violations as well as system characteristics such as recently reported population served. These data include violation reports 2000-2014. We observe intervals of time ("compliance periods") for these vio-

[11]I.e., the baby's weight at birth was less than 2,500 grams.

[12]I.e., gestation length of 36 weeks or less.

[13]SGA is an indicator for birth weight being below the 10th percentile of the infant's gestational week, while TBW is the birth weight (in grams) of babies that are not preterm, i.e., babies who "came to term."

[14]Currie et al. (2015) show evidence that the impacts of TRI facilities are strongest within 1 km, however, emissions likely persist beyond 1 km. Hill and Ma (2022) include these additional distances to account for air pollution near the maternal residence. We follow this literature here given we are concerned about air pollution confounding our estimates of the impact of drinking water contamination.

lations that are primarily reported quarterly; we define a birth to have been exposed to a violation if its gestation period overlaps with the compliance period of a violation for at least one day.[15] Further discussion of these data is provided in the Online Appendix Section A1. For our fertility analysis, we make use of the system population served measure from SDWIS, both to normalize the dependent variable across systems (we divide births per month by system population served to calculate a monthly fertility rate), and as weights for the regression model.

Most importantly for our study, we obtained public drinking water sampling information for CWS from PADEP.[16] These data contain the sample results underlying Pennsylvania's SDWA violations recorded in the SDWIS data. This drinking water sampling data contains the date of the sample, the contaminant tested, and the result of the test. These data are recorded at a much higher frequency than water quality violations, and they provide information on water contamination at levels that do not trigger a violation.[17]

These data were augmented with a shapefile obtained from the Pennsylvania Spatial Data Access (PASDA) website containing the service area boundaries of CWS.[18] Figure 1 is a map of the CWS service area boundaries for the systems in our data. As is apparent, CWS vary substantially in size. There are 1,768 CWS in the boundary file; the average system reports serving about 5,700 people, but the smallest systems (10th percentile) serve only around 60 each, while the top 10 largest systems serve 41% of people served in this data (the Philadelphia Water Department alone serves 16%, reporting about 1.6 million people served).

Most water quality samples are measured at entry points to the distribution system, or somewhere within the distribution system (e.g., at the tap for lead and copper samples).[19] In the simplest case, a treatment facility pipes water from a single source through its operations, then out into a distribution system. In more complicated cases, a distribution system can be served by several source locations.[20] Analysis requires us to aggregate up to the level of a complete CWS, as that is the level at which we can match the water sample results to birth records using the water system

[15]Compliance periods vary by contaminant with more than 2/3 being ~90 days. About 20% are 30 days and a few are annual or longer (min 27 days and max 1095 days). Importantly, we do not have the *date* of the violation.

[16]We received the data in database form from the department. Interested readers can download data from the PA Drinking Water Reporting System (DWRS) here `http://www.drinkingwater.state.pa.us/dwrs/HTM/Welcome.html`

[17]The frequency of these data are dependent on the monitoring requirements for the systems and contaminants of interest. Some contaminants are measured monthly, quarterly or annually (lead and copper are measured every three years) based upon the regulations in the Safe Drinking Water Act (US Environmental Protection Agency, 2019b). These data provide exact date of sample to match to pregnancy gestation periods.

[18]Find these data here: `https://www.pasda.psu.edu/download/dep/historic/PublicWaterSupply/`. We used vintage 2016 month 4 for these analyses. We discuss concerns about boundaries changing over time in Section 5.4 and Appendix Section A1.5.

[19]Entry points are locations in the CWS after treatment or chemical addition, if any, but prior to the distribution system. About 1% of the reported samples are reported as being of raw water; we remove those from the analysis.

[20]We discuss data limitations and this issue in Section 5.4 and Appendix Section A1.5.

boundary shapefile and geocoded mother addresses.[21]

2.3 Water Quality Index Construction Details

In this subsection, we describe how we construct our measure of drinking water quality. For each CWS, contaminant, and day, we use the samples of the contaminant on that day to calculate the average contamination in the CWS's distributed water. This measure of average daily contamination is in the units in which the contaminant is measured, typically parts-per-billion (ppb) or parts-per-million (ppm). Then, for each birth and contaminant, we average this daily measure of contamination for the CWS where the mother's residence was at the time of birth. The average is taken across all samples within the birth's gestation period.[22]

The raw Pennsylvania sampling data contains the 94 contaminants regulated under the Safe Drinking Water Act (SDWA).[23] We also obtained the health based regulatory violations from EPA by Freedom of Information Act (FOIA) request which includes the maximum contaminant level (MCL), action level or maximum residual disinfectant level (MRDL) for each contaminant and whether the system received a violation during each compliance period. (Throughout this paper, we use "MCL" and "MCL violation" to mean any action level or health based violation, respectively, even for those where the violation is based upon an action level or MRDL.) Instead of studying each contaminant separately, in our main analysis we aggregate by normalizing and averaging. Our goal is to construct a simple composite measure of water contamination, in order to avoid the statistical pitfalls of very noisy data combined with multiple hypothesis testing. The measure we choose to employ for this purpose is the average "result relative to MCL" (RRMCL).

$$RRMCL = SamplingResult/MCL \tag{1}$$

This approach accomodates the fact that what is atypically large for one contaminant may not be for another contaminant. For example, the MCL for nitrate and nitrite are 10 ppm and 1ppm, respectively. In order to create a single composite measure of contamination, we divide nitrate samples by 10, divide nitrite samples by 1, and divide all other contaminants by their respective MCL.

[21]Other work has aggregated to the county and then controlled for the percent of the population served by public water (Almberg et al., 2018). While there remains potential measurement error in our approach, we believe it is an improvement.

[22]Our method for measuring the water contamination faced by a particular birth is similar to an earlier epidemiological study on disinfectant byproducts, Bove et al. (1995), who measure contamination monthly and then take averages over the gestation period, weighting months equally. Our approach is the same, except we average over daily averages.

[23]The original Pennsylvania water sample data has 248 unique contaminant identifiers. We restrict the data to the 94 contaminants that are regulated under the Safe Drinking Water Act and for which we have regulatory MCL. These are listed in Appendix Table A2.

We then average over all contaminants sampled during pregnancy for each infant to have a measure of "overall" contamination based upon contaminants regulated by the Safe Drinking Water Act (SDWA), weighting contaminants (not samples) equally. Given the large number of chlorine samples (over 4.8 million samples; the next most common in our data is lead at 93,000 samples), we take a 10% random subsample of the chlorine samples before we build the index, to ensure computational feasibility.

Sample nondetects, i.e., samples with results that cannot be distinguish from zero, are recorded as zero in the data and included in our index calculations. This will introduce measurement error into our index measure. In general, when the contamination in the water sample is small, methods that estimate the level of contamination can have substantial relative error. We take the position expressed in the EPA internal document (US Environmental Protection Agency (2006)) that measured samples below the MCL still contain (albeit noisy) information. We further discuss this issue in Online Appendix Section A1.4.

We make two notes about our "average RRMCL" construction process. First, it maximizes sample size, but the measure is not composed of the same contaminants across births. For example, one birth may only be matched to total coliform samples, while another birth may include nitrate and arsenic samples. Each birth will have a measure of average RRMCL, but the contaminant used in the first birth is just total coliform, while the RRMCL in the second birth is a simple average of the RRMCL for nitrate and RRMCL for arsenic. To be clear, this is just a stylized example for illustration: 2.47% of births in our data have only one or two unique contaminants measured. Indeed, across births, an average of 39.31 unique contaminants are used to calculate our overall contamination index. Second, and as this example makes clear, the step in this process that likely makes the strongest assumptions is weighting the contaminants available for each birth equally. Moreover, while maximum contaminant level *goals* (MCLG) may be a more natural denominator given the focus on protecting health, MCLG are often zero, making them not as useful for aggregation. In addition to our main overall measure of contamination, we repeat this exercise, but removing coliform, heterotrophic bacteria, and turbidity. This additional measure we call a measure of "chemical" (i.e., non-bacterial) contamination. Splitting between overall and chemical contamination is consistent with previous literature (Currie et al., 2013). On average, 38.50 unique contaminants are used to calculate this chemical-only index; 3.05% of these births have only one or two contaminants in the index. Finally, we also construct a "reproductive index" measure, using five contaminants that have been found linked to reproductive health outcomes; further discussion of its construction is postponed to Section 2.3.1.

There is one important contaminant that we treat as an exception to Equation 1: total coliform. Specifically, we define the Result Relative to MCL (RRMCL) for total coliform, for each birth, as

$$\text{RRMCL} = \frac{\text{Share of Positive Tests}}{\max\left\{\dfrac{2}{\left(\dfrac{\text{Number of tests}}{\dfrac{273}{30}}\right)}, 0.05\right\}} \tag{2}$$

In this expression, the ratio (Number of Tests) / ((273/30)) gives the approximate number of tests per month (per 30 days) over the gestation period, assuming full gestation length (273 days = 39 weeks). We use this expression for the RRMCL for this contaminant, since the effective MCL for coliform is 2 tests when the number of tests per month is less than 40, and 5% of tests when the number of tests per month is greater than 40. This implies that the MCL is the maximum in the denominator of the above expression. Indeed, this Equation can be interpreted as the generalization of Equation 1 for total coliform, given that the MCL for total coliform varies based on the number of tests conducted by the system (the number of tests required to be conducted depends on system size; see Bennear et al. (2009) for a detailed analysis of the Total Coliform Rule).

Taken together, our approach to defining an water quality index is the "sampling data analogue" of the drinking water violations measure employed in Currie et al. (2013). To see this, consider that we are combining across contaminants with very different properties and health effects by normalizing by their *regulatory* thresholds. This is very similar to Currie et al. (2013), who also group all contaminants into a single summary measure, weighting equally, but instead define treatment as exposure to a regulatory violation during gestation. The advantage of our approach is that it allows us to use variation in contamination that doesn't trigger a health-based violation, which is important for providing policy-relevant evidence.

Throughout our analysis, we refer to and use what we call our "analytic sample," which restricts births as follows. First, births in our analytic sample are not exposed to any MCL violations; second, they have some measure of contamination within their gestation period; third, they are not plural births; and fourth, their measured contamination does not exceed the 99th percentile across all births in the data for the specific contamination measure under study.[24] In addition, when we provide summary statistics, we remove singleton observations based on CWS and birth year-month fixed-effects, or mother and birth year-month fixed-effects, in order to calculate statistics for the

[24]The last restriction depends on the contamination measure we are using as our independent variable, e.g., when we take overall contamination as the independent variable of interest, we trim at the 99th percentile for that measure.

exact samples used in our regressions.[25]

Table 1 describes the distribution of our average RRMCL measures across births; throughout, births exposed to MCL violations are excluded. In preliminary analyses, we noticed that our RRMCL index measures are very right-skewed, even after removing violation-exposed births. Thus, we decided to trim out the top 1% of births based on RRMCL when we formed our analytic sample.[26] In panel A, we show the distribution of RRMCL across births, by contaminant group, after our trimming. To illustrate why we trimmed outliers, panel B shows the distribution before trimming. Finally, since our main specification for identifying causal effects is a within-mother design, Panel C shows the distribution of these index measures after removing mother fixed-effects. We see that the standard deviation in Panel C is about two-thirds the size, as is the spread between the 10th and 90th percentiles.

To facilitate comparing the overall, chemical, and reproductive index contamination index measures, in our regressions, we standardize such that a 1 unit change is equivalent to a change from the 10th to the 90th percentile of the average RRMCL, for the RRMCL measure of interest (overall or chemical) as the case may be. These movements are equivalent to about 1.875 standard deviations for the overall measure, about 1.625 standard deviations for the chemical measure, and about 1.889 standard deviations for the reproductive index (five contaminant) measure (Table 1 contains the quantiles of the distribution). We chose the 10th and 90th percentiles to standardize in order to simulate a large, though not impossible, change in contamination exposure.[27] For comparison, a movement from the 25th to the 75th percentile is about 33% of the 10th-to-90th percentile movement for the overall contamination measure, and about 31% for the chemical-only measure. The reader can rescale our estimates to represent effects of these movements by multiplying by 0.33 or 0.31, respectively.[28] Finally, our most rigorous specifications use within-mother across-sibling variation only; Table 1 includes statistics describing the spread of our contamination indices across births within mother in Panel C, in case the reader wishes to rescale our estimates based on that distribution instead.[29]

[25]See Correia (2015) for a discussion of the importance of eliminating singleton observations in order to obtain the correct standard errors in high-dimensional fixed-effects regressions.

[26]Trimming the top 1% substantially improves the precision of our estimates, and without this trimming we would generally conclude that there is no significant effect of water contamination on birth outcomes, in contrast to our present conclusion of a precise effect (estimates without trimming are not shown).

[27]Note that it is a large movement in the conditional distribution of birth exposure after removing MCL violations and after trimming outliers, which are both data manipulations that limit this difference. See Table 1 for the quantiles of the underlying RRMCL variable.

[28]The regression models we use are linear, so the reader can multiply our estimates by these ratios (or different ones based on Table 1) to obtain the estimated effects for smaller changes as desired. To obtain the effect of a change in our overall contamination index of Δ units of RRMCL, simply multiply our estimates by $\Delta/0.15$; similarly for chemical, multiply our estimates by $\Delta/0.13$. For instance, to consider the effects of a change from 10th-to-90th within mother for the overall contamination index, multiply our estimates by 0.10/0.15 = 66%.

[29]Based on this panel, the movement from the 10th to the 90th within-mother is about 0.10 RRMCL for our overall index (the difference between the 90th percentile of 0.05 and the 10th percentile of -0.05; note that since Panel C is the

In the Online Appendix Section A1.1, we further describe the correlation between our index measures and the samples for selected contaminants of interest in Table A1, as well as provide a full table listing all the contaminants included in each of our index measures (94 for the all index, 91 for the chemical index) in Table A2.

2.3.1 Reproductive Health Water Quality Index

EPA lists 13 contaminants associated with "reproductive difficulties" and another 6 associated with harms to "infants" (US Environmental Protection Agency, 2018). See Appendix Table A3 for the specific contaminants. We performed a literature review (see Appendix Section A1.3 for discussion of this literature) and the epidemiology literature primarily focuses on arsenic, nitrate, atrazine, and disinfectant byproducts (Haloacetic Acids (HAA5), Total Trihalomethanes (TTHM)) and only nitrate and atrazine overlap with the EPA's list. We also found some epidemiological literature associating tetrachloroethylene (PERC) and Di(2-ethylhexyl)phthalate (DEHP) and a recent study in Virgina included these plus lead, radon, and fecal bacteria (Young et al., 2022).[30] We assessed the availability of samples for these contaminants to create an index that would have more uniform coverage in our sample population and limited to 5 with frequent sampling in Pennsylvania for our reproductive-health specific water quality index: arsenic, atrazine, nitrate, Di(2-ethylhexyl) Phthalate (DEHP), and Tetrachloroethene (PCE or PERC). On average, 3.27 unique contaminants are used in this more specific water quality index. This index could be used to target water alerts communications to pregnant persons and other studies could create water quality indices for specific health outcomes to target water alerts to specifically vulnerable populations.

2.4 Summary Statistics

Summary statistics for births are provided in Table 2. In column (1) we display statistics for all non-plural births matched to CWS without exposure to MCL violations; in column (2) we restrict to our analytic sample for overall contamination.[31] These births include those to mothers who switch CWS between some of their births in the panel. There is very little difference in mother characteristics between columns (1) and (2), suggesting that samples are consistent after addressing data availability and trimming. Columns (3) and (4) examine the births in column (2), split by the tails of average contamination by CWS. These columns suggest that mothers of

distribution of residuals after removing mother effects, the distribution is centered around zero but the units are still in RRMCL), or about 66% of the 10th-to-90th percentile movement we use, and 0.09 for our chemical index, or about 69% of the movement we use in that case. See prior footnote for scaling instructions.

[30]They only find an association between MCL violations for nitrate-nitrite in this study.

[31]We keep only births with contamination measures, trim at the 99th percentile for overall contamination, and drop singletons based on CWS and birth year-month fixed-effects (these restrictions were discussed in the previous subsection).

lower socioeconomic status, and worse birth outcomes on average, tend to give birth in places with higher levels of water contamination (that doesn't trigger a health-based violation). Column (5) restricts the sample to the mothers with multiple births that are included in our mother fixed-effects regressions. In general, differences between column (5) and column (2) are not large, suggesting that sample composition does not substantially change when we restrict to mothers with multiple births (i.e., siblings).[32]

To reiterate, column (1) of Table 2 already restricts to births matched to CWS; in Online Appendix Table A8 we compare this restricted sample to the sample of all births, including those not matched to CWS. The most notable difference is that the sample matched to CWS is more urban, which is expressed most strongly by being of greater share black (as rural Pennsylvania is predominantly non-black).

3 Empirical Methods

We employ two regression models to study the effects of drinking water contamination on birth outcomes, one with community water system (CWS) fixed effects (FE) and the other with mother FE (within-mother across-birth variation).

The first model takes the following form:

$$\text{Outcome}_{iwmt} = \beta \text{RRMCL}_{iwmt} + \theta' X_{iwmt} + \eta' \text{Temp\&Poll}_{iwmt} + \alpha_w + \gamma_t + \epsilon_{iwmt} \tag{3}$$

In this model, i denotes birth to mother m living in CWS w in year-month t at the time of the birth; Outcome_{iwmt} is a birth outcome; the birth outcomes we consider are: low birth weight (LBW), preterm birth (PTB), small for gestational age (SGA) and term birth weight (TBW). See Section 2.1 for information about how these outcomes are defined.

This regression specification employs a fixed effects estimator, where α_w and γ_t are water system and birth year-month fixed-effects, respectively. The independent variable of interest is RRMCL_{iwmt}, "mean result relative to MCL," a measure of overall or chemical drinking water contamination which we construct from the sample concentrations as described in subsection 2.3. The identification assumption is that, after adjusting for an array of individual mother, weather and pollution characteristics in X_{iwmt} and $\text{Temp\&Poll}_{iwmt}$, trends in birth outcomes across water systems would be the same on average except for differences due to differences in water contamination trends.

[32]In exploratory regressions (not shown), we found that systems with a greater share of non-Hispanic white mothers, and lower share served by WIC/Medicaid, are more likely to have violations. This is surprising and differs from findings of Currie et al. (2013). However, our sampling data-based measure of contamination exhibits different relationships: systems with lower share non-Hispanic white experience higher average levels of contamination by these measures, as do larger systems. These last results hold both before and after removing births exposed to violations.

We control rigorously for a wide array of mother, birth, and weather measures. Whenever a variable is missing, we set it to zero and include an indicator for missingness. We include a full set of mother demographic and socioeconomic controls, and risk factors for the pregnancy, in X_{iwmt}.[33] These control variables were curated from the birth records.[34] The variables in Temp&Poll$_{iwmt}$ are weather and temperature controls obtained from the daily weather statistics provided by Schlenker (Schlenker and Roberts, 2009), as well as Toxics Release Inventory (TRI) facility statistics from the EPA (US EPA, 2013).[35]

The CWS fixed-effects model also naturally lends itself to an aggregate water-system level fertility regression,

$$\begin{aligned}(\text{Birth Rate})_{wt} = {} & \beta\,(\text{RRMCL last 9 months}_{wt}) \\ & + \theta'\,(\text{average } X \text{ last 9 months}_{wt}) \\ & + \eta'\,\left(\text{average Temp\&Poll last 9 months}_{wt}\right) + \alpha_w + \gamma_t + \epsilon_{wt}\end{aligned} \tag{4}$$

This model is motivated by the concern that observed fertility rates may respond to drinking water contamination, either due to behavior such as mobility or selection into pregnancy, or due to health issues, such as infertility, miscarriage, or stillbirth. In this model, (Birth Rate)$_{wt}$ is the birth rate for the CWS in the given year-month, defined as the number of births divided by the system's population served. We also estimate a version using $log(births)$ as the dependent variable to alleviate concerns that we are not accurately measuring CWS-population. Our population served measure is time-invariant, so we cannot ignore the fact that variations in birth rates may be due to mothers leaving or entering CWS over our panel; we address maternal mobility below. The variable of interest (RRMCL last 9 months$_{wt}$) is a measure of mean contamination in the water system over

[33] These control variables include: mother's age categories (19-24, 25-34, 35+); mother race (non-Hispanic white, non-Hispanic black, Hispanic); mother's education (less than high school, some college, college or more); risk factors for the pregnancy (including pre-pregnancy diabetes, gestational diabetes, pre-pregnancy hypertension, gestational hypertension, previous pre-term birth, previous poor pregnancy outcomes, vaginal bleeding, infertility treatment, previous cesarean); maternal smoking (counts of cigarettes smoked by term); parity indicators (number of previous successful live births and unsuccessful live births); mother is married; child is male; WIC participation; and payment type (Medicaid, private, self-pay).

[34] These covariates are commonly used in the birth outcome literature. See Almond et al. (2018) for a review and Currie et al. (2015), Currie et al. (2013), and Hill and Ma (2022) for specific examples.

[35] We are concerned about other environmental confounders that may be co-occurring with drinking water contamination, such as weather, temperature and air pollution. Our weather and temperature controls include maximum and minimum daily temperature, percentage of days in which the daily maximum is above $29.4°C$, percentage of days in which daily minimum is below $0°C$, average daily precipitation, percentage of days in which precipitation is over 0, percentage of days in which precipitation is over 0.25 millimeters, and maximum and minimum daily precipitation. Our TRI facility statistics, which proxy for air pollution (a potential confounder, even in our mother fixed-effects models), include the number of TRI facilities operating within 1, 3, and 5 kilometers of the maternal address in the year the infant was born, and the reported total onsite, offsite, and overall releases for TRI facilities weighted by the squared inverse distance between the facility and the mother's residential location. We include these weather and air pollution controls following Knittel et al. (2016) and Currie et al. (2015).

the last 9 months (by aggregating the RRMCL measures described in section 2, weighting months equally), (average X last 9 months$_{wt}$) is a measure of average birth characteristics listed earlier taken over the last 9 months weighting births equally, and (average Temp&Poll last 9 months$_{wt}$) is the same battery of weather and TRI facility-based controls we use in models at the individual birth level, except again averaged over the last 9 months, weighting births equally.[36] In these regressions, we weight systems by their population served.

In light of the possibility that any fertility effects we identify may be due to mothers simply moving out of CWS in response to contamination, it is useful to repeat regression (3), but with a different outcome variable: whether the mom is in a different CWS in her next birth (restricting to mothers with at least two births). These models directly calculate whether mothers can be seen to switch systems more often when there is greater water contamination in their first birth. These estimates indirectly speak to whether any fertility effects we see are due to mother mobility.

The second birth-level regression model we employ is a within-mother analysis, which uses variation only across the births of the same mother. The regression is identical to the CWS fixed-effects model (equation 3) but mother fixed-effects are added, in addition to the current fixed-effects. Controlling for all time-invariant characteristics of mothers is a rigorous way to determine whether water quality variation affects birth outcomes, rather than being confounded with other mother characteristics.

As before, in addition to fitting mother fixed-effects models with birth outcomes as the dependent variable, we can also fit these with mother mobility (switching CWS) as the outcome. We estimate mobility effects within CWS and within mothers. The effects of contamination on mobility are identified off relating within-mother differences in contamination that mothers face across births to within-mother differences in whether the mother has switched CWS between births.[37]

The recent literature on difference-in-differences (DD) and two-way fixed effects (TWFE) estimators have made significant advances in a staggered treatment setting (Callaway et al., 2021; Callaway and SantAnna, 2021; De Chaisemartin and dHaultfoeuille, 2020; Goodman-Bacon, 2021). While we do employ a panel model that leverages within water system or within mother variation over time, we are not explicitly using a difference-in-differences model due to this being a non-standard setting with no clear event. Contamination can steadily increase over time, stay relatively constant, reduce abruptly following a change in technology or following a regulatory violation/remediation, and it is not clear that it follows a standard pattern commonly used in a DD

[36] An exception to using the same controls is that we use the overall share of plural births (twins or greater) rather than the shares of each plurality type.

[37] Note that to be included in mother FE regressions we need mothers to have three or more births, with at least two of them with contamination measures: for example, for a mother with exactly three births, we observe our switching indicator for the first two births only, and then the mother's fixed-effect absorbs another degree of freedom. The CWS FE model only requires two births.

model exploiting pre/post treatment (with continuous treatment). Here, there are also unlikely to be "never treated" units.

While the literature on TWFE DD and event studies with continuous treatment suggests this could still be an issue (Callaway et al., 2021), the literature does not at this time have a solution to employ. The primary issue that the continuous treatment in a TWFE design presents is a selection issue. The decomposition in Callaway et al. (2021) shows that the presence of this selection bias without an easy fix will require credibly arguing that the average treatment effects are equal across all dosage groups at the same dosage level. The identification assumption is strong– that there is no selection into the intensity of treatment. However, we are studying dosages of water contamination that are compliant with current regulations and does not trigger a violation (i.e., information to the consumer). The water systems do not have an incentive to manipulate concentrations except near the regulatory threshold. We show that mothers are not changing water systems in response, nor are there large changes in fertility. Finally, our "dose" is defined by a water quality index that is not observed by the consumer, the operator, or the policymaker, and it is unlikely that any systems or mothers are selecting on the level of treatment.

4 Results

Tables 3 and 4 display our main estimates of the overall effects of drinking water contamination in community water systems (CWS) on birth outcomes. Tables 5 and 6 present the results for the reproductive-specific water quality index. Table 7 displays our estimates of the effects of water contamination on mother mobility or overall CWS-level fertility. Throughout, we restrict to our analytic sample. In particular, we have removed all births that have a gestation period which overlaps with any MCL violation compliance period. Thus our estimates are of the effects of water contamination on birth outcomes, mobility, or fertility, for variation in water contamination which does not coincide with MCL violations.

4.1 Birth Outcomes

Fitting the model described by Equation 3, we find that the chance of low birth weight (LBW) and preterm birth (PTB) are both significantly positively related to contamination overall and chemical contamination specifically (Table 3). Reading across columns, we find that the addition of our extensive set of control variables does not change inference, despite raising adjusted R^2 substantially. Note, however, that point estimates are attenuated by 10 to 28% in the fully adjusted specifications (columns (5) and (6)) relative to the unadjusted specifications (columns (1) and (2)). Generally, effect estimates are larger in the unadjusted models. Reading across rows, we find that our results are more sensitive to whether we use a CWS fixed-effects (CWS FE) specification or a mother fixed-effects (MFE) specification (that also includes CWS FE). In general estimates are larger in the CWS FE specifications. All coefficients are estimated effects of a change from the 10th percentile to the 90th percentile in the Pennsylvanian distribution of exposure to CWS drinking water contamination.[38] Thus they are comparable across specifications. In both cases, effects are economically meaningful: based on columns (5) and (6), in the MFE specification, a 10-to-90th percentile move leads to about a 0.8 percentage point increase in low birth weight (LBW). From the base LBW rate for non-plural births of 6.7% (Table 2), this is about a 12% increase. Analogously for PTB, we find about a 17% increase.

In order to formally calculate the degree to which the addition of control variables affects our inference, we report "AETO's δ." By the acronym AETO we refer to the two papers Altonji et al. (2005) and Oster (2016); the latter work substantially builds on the former and provides us the Stata implementation we use.[39] The interpretation of δ is the minimum ratio of selection on unobservables to selection on observables that would be needed to make our estimated coefficients zero.[40] A ratio of at least one is a plausible rule for saying an effect estimate is "robust to controls."[41] In our case, AETO's δ tends to be large in the CWS FE models, but much smaller in the mother (and CWS) fixed-effects models (Table 3). This is partly by construction: the mom fixed-effects absorb many potential unobservables, making it unlikely that the inclusion of control variables will appreciably change the R^2 in the model. Indeed, consistent with this idea, when we compare for example, columns (1) and (3) in the mom fixed-effects specifications, the control variables do not substantially increase adjusted R^2. That said, most of our estimates have AETO's δ's greater than one. This also lends support for our earlier claim that there is not a clear indication of selection into a particular dose of treatment.

In Table 4, we report the estimated effects on more direct measures of intrauterine growth restriction: small for gestational age (SGA) and term birth weight (TBW). These estimates are consistent with our main findings. In other words, after applying at least some approaches to removing the effects of contamination on gestation length, we continue to find effects of contamination on birth outcomes, suggesting that effects on gestation length are not the only channels by which water contamination affects birth outcome, and intrauterine growth restriction (IUGR) is a possible

[38]For the distribution of the underlying result relative to MCL (RRMCL) measure, see Table 1. Section 2.3 and footnote 29 provide details for alternative scaling.

[39]Specifically we use Oster's code available from the Boston College Statistical Software Components (SSC) archive under the Stata command -psacalc-. We modify this code to work with the Stata command -reghdfe-.

[40]We use the recommendation of Oster (2016), based on an empirical analysis of randomized controlled trials published in prestigious economics journals, and assume a maximum R^2 (the R^2 we would obtain if all potential confounders were included) of 1.3 times the R^2 estimate from the model with controls.

[41]Note that a negative δ means that selection on unobservables would have to be the reverse of selection on observables to bring the effect to zero.

consequence of contamination.

While our main analysis focuses on indices of water quality that include a large number of contaminants, as discussed in Section 2.3.1 and Appendix Section A1.3, an alternative approach is to restrict to contaminants which are commonly sampled and for which we could find research literature suggesting deleterious health effects for infants. We ultimately chose five contaminants to include in a reproductive health specific water quality index. Estimates using this five-contaminant index to measure water quality are included in Tables 5 and 6 (the tables include estimates for our all index of contamination for comparison). Across the various specifications, we see slightly stronger effects for the reproductive-specific index. In particular, focusing on columns (5) and (6) and mother fixed-effects regressions, we find that a 10th-to-90th percentile movement in RRMCL for our five-contaminant index leads to a 1.3 percentage point increase in low birth weight (about 19% relative to the mean), a 1.8 percentage point increase in preterm birth (about 22% relative to the mean), 0.3 percentage point increase in small for gestational age (significant at only the 10% level), and a 12.59 gram decrease in term birth weight.

For interested readers, in Appendix Section A3 we estimate models studying the effects of MCL *violations* on health, following Currie et al. (2013). We find that 3% of our sample has a regulatory violation during gestation, compared to 8% in NJ in the prior study. We find similar effects in magnitude and statistical precision for any MCL violation, but not for chemical violations, which is in contrast to Currie et al. (2013).[42]

4.2 Avoidance via Migration

Table 7, Panels A and B, display the estimated effects of contamination on mother mobility. Mobility is defined for a given birth as an indicator for the mother being in a different CWS at the time of her next birth. Regressions in this table are similar to equation 3, except with mobility as the dependent variable. As before, we restrict to our analytic sample as described in Section 2. In these models, the independent variable of interest is CWS drinking water contamination, as measured during the gestation period of the current birth.

Mothers would exhibit avoidance behavior via migration if, when they experience higher levels of water contamination during this gestation period, they are more likely to be found in a different CWS at the time of their next birth. We find little evidence for this type of avoidance behavior, which is consistent with mothers not behaviorially responding to contamination levels that do not trigger MCL violations. In Table 2 about 37% of births have their mother in a different CWS at the next birth (this is consistent with statistics from New Jersey in Currie et al. (2013)), so the point estimates from Table 7 are also economically small.

[42]More details and discussion of our findings are in the Appendix given the focus of this paper on contamination compliant with regulatory standards.

While we have evaluated the degree of avoidance behavior through very costly migration decisions, this is not the only possible and probably not the most likely form of avoidance behavior. It is possible that very attentive mothers purchase bottled water instead of using tap water, or they filter their tap water, when contamination rises. This avoidance behavior is less likely in our setting where we study water quality that doesn't trigger a health-based violation unless their system previously experienced violations. However, mobility looms large as a potential avoidance behavior that could lead to bias in our estimated effects of contamination on birth outcomes, and especially fertility, discussed in the next section, so it is helpful to note little evidence for this particular behavior.

4.3 Fertility

Table 7, Panel C, displays the estimated effects of contamination on fertility rates from the regression described in equation 4. These regressions are weighted by CWS population served. CWS fertility rates are defined as monthly births per 100,000 people served.[43] Our findings are consistent with poor water quality reducing fertility. A typical birth is in a CWS with about 113 births per 100,000 served per month (Table 2). We find that changes in overall contamination from the 10th to the 90th percentile reduces this number by about 0.07 births per 100,000 per month, or approximately a 0.06% reduction in fertility. Effects for chemical contamination are similar. Comparing these effects to those found for lead contamination, Grossman and Slusky (2019) found that the Flint water crisis led to a 12% reduction in fertility, while Clay et al. (2021) found, for topsoil lead, that bringing a typical above-median county to the lead levels of a typical below-median county would lead to approximately an 11% increase in the birth rate. By contrast, we find that going from the 10th to the 90th percentile of water contamination in our sample has a much smaller, though still statistically significant, negative effect on fertility. Lead is known to induce miscarriage, whereas the more general contamination studied in our paper does not have strong evidence for causing pregnancy loss. Given that mobility effects appear to be limited, the effects on fertility are likely due to selection out of pregnancy, either through choice or through contamination leading to reduced fertility.

Our fertility regressions in Panel C of Table 7 use time-invariant CWS population served in the denominator of the dependent variable. As a robustness check, in Panel D of the same table we use log births as the outcome variable, thereby avoiding the use of the time-invariant population served

[43] Analogous to our approach at the individual birth level, we remove CWS-months for which an MCL violation occurred in the last 9 months, and we remove CWS-months above the 99th percentile of contamination (among all CWS-months in the data). However, the units of contamination used in this model are the same as in the previous regressions. Specifically, a one unit increase of the contamination measure reflects a change from the 10th to the 90th percentile in the overall index of drinking water contamination among births (rather than among CWS-months) after removing births above the 99th percentile.

variable. This approach is commonly used (Guldi, 2008; González, 2013; Ananat and Hungerman, 2012). Effects in Panel D are larger in magnitude, at about a 0.1% reduction in fertility, but they are of the same sign and statistical significance as our main fertility results in Panel C.

5 Sensitivity Analyses

In this section, we discuss the sensitivity of our findings to different subsamples. For all results presented thus far, we have restricted to non-plural births, and have included all mothers, including those who switch CWS. We show robustness to these sample selection choices in this section. All tables referenced in this section are included in Online Appendix Section A2.

5.1 Robustness to sample restrictions based on plurality and mobility

In the Online Appendix, we estimate our models with two alternative samples: first, all births (i.e., including plural births), and second, sub-setting to never-moving mothers. Including plural births does little to change our findings (Tables A9 and A10). Our findings are also robust to excluding all mothers with multiple births who we ever observe in different CWS across births (Tables A11 and A12). This is surprising since this sample restriction removes a quarter to a half of the births in our data, depending on specification; moreover, one component of our variation in water quality from the mother fixed-effects analyses was due to mothers switching water systems with different average contamination. The robustness of our results in this very restrictive sample provides additional support for our within-water system estimates.

5.2 Robustness to sample restrictions based on MCL violations

Recall that throughout our analysis, we excluded births exposed to MCL violations. In the Online Appendix, we also re-estimate our models with two additional alternative sample restrictions: we run the analysis keeping births that are exposed to MCL violations, and we run the analysis removing births with any samples exceeding the MCL threshold. The latter analysis is motivated by the complexity of the regulatory environment, in that in many cases violations are triggered only when a particular average of samples exceeds the regulatory threshold (e.g. TTHM and HHA5 use running averages), and a concern that the MCL violation data may be missing some violations. See the Online Appendix Section A2.3 and Section A2.4 for further details on these alternative samples. We also include models estimating the effects *of* MCL violations on birth outcomes in Online Appendix Section A3.

Our findings are robust to including births exposed to MCL violations (Online Appendix Tables A13 and A14). This may not be surprising since our sample size increases by only about 3 to

5% (depending on specification) by adding these births. More surprisingly, our findings are robust to removing births exposed to samples exceeding MCL thresholds (Online Appendix Tables A15 through A20). In conducting these further restrictions, we document in the Appendix Section A2.4 that four contaminants very often have single samples that exceed their respective MCL (Coliform, TTHM, HAA5, and lead), and this makes sense because none of these contaminants have violations driven by single-sample exceedences. Our findings are robust to several different ways to treating these four contaminants, which is an important consideration since if we drop births exposed to any samples with exceedences for these contaminants we lose about 85% of our data. In any case, this robustness check is intended to simulate a stricter definition of MCL violation, and reduces sample sizes by 25% to 85% depending on specification. That our results are robust to these stricter definitions of "compliant drinking water" suggests that there are indeed statistically significant effects of water contamination on birth outcomes, even when systems are in compliant with regulations or all sample concentrations are below the regulatory threshold.

5.3 Robustness to restricting to MFE sample using CWS FE

In the main specifications, using mother fixed-effects (MFE) rather than community water system fixed-effects (CWS FE) often appreciably and significantly changed the effect estimates (e.g., the differences between the CWS FE and MFE subpanels in Table 3). In fact, this is the main way in which our results are sensitive, since the addition of mother and birth control variables does little to change our estimates.

In the Online Appendix, we present CWS FE and MFE estimates sub-setting throughout to the sample used in the MFE specifications. Our results are robust in Tables A21 and A22. Thus, the differences between the CWS FE and MFE specifications are due to differences in the models we fit, not due to differences in the samples we employ.

5.4 Study Limitations

Our estimates may be affected by a few forms of exposure misclassification which is common in environmental health research. First, our index is created from the measured contamination during gestation periods. The regulations provide monitoring schedules that vary by size of system, season, and contaminant (among others). For some contaminants, systems are required to measure monthly, quarterly, annually, every 3 years, and in some cases every 6 years. Unfortunately, this temporal sparsity of the data leads to unobserved contamination levels during gestation for some contaminants that make up our index for any individual pregnancy. We chose not to impute contamination because we do not believe we have adequate information to predict contamination over time across systems. Despite this limitation, we believe using the sampling data provides more temporal frequency than the MCL violations data as they only report violations or not in each monitoring period (approx. quarterly) but when there is no violation, the researcher doesn't observe whether it is a true zero (measured) or just no violation in the monitoring period (which leads to censoring). Second, our estimates are dependent on the accuracy of the technology measuring the concentrations for CWS. In particular, in the Pennsylvania data, non-detect values are treated as zero, when in fact, the technology may have a minimum detection limit (MDL) such that we cannot be confident that the concentrations are in fact zero. In Appendix Section A1.4, we examine how replacing non-detects with the MDL changes the distribution of our index, and furthermore show that Pennsylvania MDL are typically lower than California's, suggesting Pennsylvania is a decent setting to conduct this study. Third, we rely on water system service areas to define exposure to CWS and assign water quality during gestation at the system-level to each mother residing in that water system. There could be exposure misclassification if the water ingested by an individual mom is more or less contaminated than what is measured prior to the water being released into the distribution system. Ideally, we would have distribution maps and be able to assign samples for sub-areas of the service area. Unfortunately, these distribution maps are not available and this is a data limitation.[44] Despite this limitation, we believe that this is analogous to what is used in air quality studies extrapolating monitor or satellite data (Currie and Walker, 2019) or water quality studies at the county or city level (Grossman and Slusky, 2019; fly, 2199). However, we are using system boundaries providing sub-county variation in water quality, which should improve upon existing approaches that use county-level variation (McDonald et al., 2022). Finally, in our analysis, we used a singular vintage of the water system boundaries. In the Appendix, we checked to see if there was any correlation between systems that changed over time and the water quality index. See Appendix Section A1.5 for details. We found that while there were some changes over time, they are minimally associated with our measure of water quality. We hold the boundaries constant over our study period to ensure that factors that might influence changing systems (e.g., population growth, population decline, water contamination, etc.) are not influencing our assignment of treatment.

6 Discussion

Policymakers have increasing begun using summary indices to regulate or improve communication of quality in various sectors for consumers. For example, Centers for Medicaid and Medicare Services (CMS) has promoted Hospital Compare, Nursing Home Compare, and Medicare Care Compare as mechanisms for improving health care quality. Most related to our context, the EPA has long provided the U.S. Air Quality Index (AQI; updates by metro area daily) and the AQI is now reported in AirAlerts communicating to the public when air quality could be risky for at-risk

[44]More details about this issue are discussed in Appendix Section A1.5.

individuals. State departments of education provide school report cards and provide summary measures of quality using "grades". Researchers have studied responses by consumers and suppliers (e.g., providers, schools) to use of these indices for communicating quality. For example, AQI and communicating poor air quality has been shown to change participation in outdoor activities (Zivin and Neidell, 2009; Wen et al., 2009) and following star-ratings for Nursing Home Compare, low scoring nursing homes lost demand and high scoring nursing homes experienced an increase in demand (Werner et al., 2016). With recent scrutiny on water quality, use of a water quality index to communicate more frequent water quality information to society could be beneficial for avoiding poor outcomes for the most sensitive groups.[45]

Figure 2 displays an excerpt of Philadelphia's 2017 Consumer Confidence Report (CCR), which was provided to consumers of Philadelphia water to inform them of water quality in the prior year. Sample result data in the report is displayed by contaminant, for example, the highest level of Barium reported in Philadelphia's water in 2017 was 0.044ppm. Except for reported regulatory violations, which are rare, consumers are unlikely to extract actionable information from current Consumer Confidence Reports. Motivated by our results, we propose that regulators consider providing a summative, index measure of water quality for CWS, to be displayed on annual CCRs or to be used more frequently as water alerts to community members.

We run a simple simulation to provide an example policy counterfactual, if systems were to adopt our overall index measure. Our example accountability policy takes our index measure, measured at the birth level, and aggregates to the CWS level by averaging the measure for all births (below the 99th percentile of contamination) since the start of 2010 through the end of 2014.[46] This gives us a "mean result relative to MCL," or "mean RRMCL" measure. We then split systems into letter grades, A through F, as follows. We begin by splitting up the CWS-level mean RRMCL measure into four quartiles, and assigning letter grades A through D to those quartiles. Then, if a system receives any MCL violation (or for lead and copper, any trigger of the action level) over the 2010 through 2014 period, we assign the system to grade F. Several large water systems with their letter grades assigned according to our method are displayed in Table 8.

In Table 9, we display results from a counterfactual simulation, where we assume that consumers in systems with letter grades C, D, or F engage in avoidance behavior or lobby their water systems for better water treatment, and these responses ultimately move their effective water quality to the mean quality for systems of letter grade A. The table displays back-of-the-envelope estimated effects on birth outcomes.[47] To put these numbers in context, there are about 150,000 singleton births

[45]This is the stated purpose of the EPA's Air Alerts: "The AQI informs the public about air quality in the area, tells who may be affected, and provides steps to take to reduce exposure when pollution levels are unhealthy."

[46]We use births below the 99th percentile to ensure consistency with the rest of the paper; naturally, the letter grades would be different were we to include these outliers.

[47]Consistent with our approach elsewhere in the paper, we remove births exposed to contamination above the 99th

in Pennsylvania each year, and about 10,005 are low birth weight (6.67 out of 100); the proposed policy could reduce this to about 9,645 low birth weight infants, thus, about 360 infants per year would avoid being low birth weight (3,960 over our 11-year period). Similarly, for preterm birth, there are about 12,195 such infants born in Pennsylvania each year; the policy could reduce this to 11,565, thus preventing 630 preterm births annually (6,930 over 11 years). This policy could save around $290 million in total costs for PTB and LBW associated hospitalizations and medical services before age 9.[48] Given the large human capital costs of poor birth outcomes for later life education outcomes, labor market outcomes, and life expectancy (Black et al., 2007; Almond et al., 2018), our estimates clearly have economic and policy significance.

Based on this simulation, we argue that providing a grade for water systems could be an improvement on the current information provided to consumers. We have to acknowledge, however, that while this provides more information for the consumer, it puts the burden of preventing exposure on the consumer and may disproportionately affect low-income families. Research on water systems in California, however, suggests that water systems may not be removing contaminants even after regulatory violations (Grooms, 2016). Given this imperfect enforcement, we do believe that adding a summary index (after further validation) to CCR or used in water quality alerts could further protect public health at the very least by informing consumers of water quality in their local systems. For example, the EPA's Air Alerts indicate levels of concern as well as which populations might be most at risk and should avoid exposure. The actual implementation of this information would need to take into account equity and environmental justice concerns, much like air quality alerts where low-income families may not have air conditioning, access to air filters or be able to avoid exposure due to employment that requires spending time outdoors.

The index that we used in this study is but one possible summary measure, and we encourage researchers and policymakers to consider alternatives. Our index, or an alternative water quality index, could be used to study other health outcomes and other populations that could be at risk of adverse outcomes from exposure to drinking water contamination (e.g., children, older adults). As research causally links various water quality indices that may be relevant for particular populations (such as our reproductive index described in Section 2.3.1) or how our overall index affects the health of other populations, "water alerts" could be targeted to specific populations at risk with enough research and validation to support such implementation.

We view our letter grading policy proposal as a way to consider the potential benefits of

percentile before conducting these counterfactual exercises. In this analysis, we kept births exposed to MCL violations, however.

[48]Russell et al. (2007) calculated that each instance of LBW costs $14,500, on average, due to increased hospitalization and related medical services use ages 0-9 years. For each instance of PTB, Behrman et al. (2007) calculated the hospital costs to be $33,284. Back-of-the-envelope estimates suggest a savings of $57.4 million for LBW and $230.7 million for avoided PTB. These estimates are likely overstated because PTB babies are also more likely to be LBW and therefore could double count savings.

improving water quality, and our study is not a formal cost-benefit analysis of safe drinking water policies. The cost-benefits literature on water contamination is relatively new, and a recent evaluation of the Clean Water Act suggested that its benefits did not exceed its costs (Keiser and Shapiro, 2019a). One approach to assess costs in our context might be to assume that households make bottled water purchases to replace tap water with higher quality water, but it is unclear whether bottled water is a sufficient substitute for quality tap water. In particular, the persistent and large negative house price effects of the Flint water crisis found in Christensen et al. (2023), despite three years of free bottled water provision and several programs to distribute free water filters and filter cartridges, suggests that private goods are imperfect substitutes for quality tap water.

Our results using a water quality index have two additional implications for water regulation. The scientific literature has suggested that certain co-occurring contaminants that are not regulated by SDWA and therefore are unmeasured in our data could be explaining some of the effects we observe (Bradley et al., 2018). Using this index could be helpful, given regulators can not monitor all contaminants in drinking water. Furthermore, as briefly discussed in the introduction, SDWA regulates contaminants on a contaminant-by-contaminant basis. Toxicology literature suggests that mixtures of contamination may be more problematic for human health (Gennings et al., 2018; Braun et al., 2016), and people are exposed to mixtures in real-world settings. Our approach of using a summary index could be a first step towards having drinking water policy take into account mixtures.

In this paper, we define compliance as not having a health-based violation as reported to the EPA. Recent work (Baker et al., 2022) has argued that while there is some evidence that violations go unreported to the national EPA, the majority (79%) are accurately reported.[49] Given that we condition on no health-based violations as reported to the EPA, we join other studies (Allaire et al., 2019) that suggest that improving this reporting could benefit communities health through communicating water quality accurately.[50] Furthermore, if violations do not result in improved compliance, as Grooms (2016) suggests, our findings could reflect this lack of enforcement resulting in poor birth outcomes.

We address the question of "missed" or unreported violations by crudely removing samples that are above the regulatory threshold for each contaminant (Appendix Section A2.4). The current regulations do not treat these thresholds as binding for all contaminants, and so our removal of concentrations above any regulatory threshold is much more stringent than the current regulations. We find similar results when we treat thresholds as binding for all contaminants in Appendix Tables A19 and A20. While this suggests that even when concentrations are below the regulatory

[49]We found 94% of violations were reported to EPA. See Appendix Section A2.4 for more details.

[50]Allaire et al. (2019) suggest using bottled water purchases for surveillance to identify emerging water quality problems that might go undetected or unreported.

threshold, water contamination could be impacting infant health, we lose substantial sample and caution that more research is needed to confirm these findings.

For analyses such as those reported in this paper, the lack of regular monitoring for certain contaminants that may be very important for reproductive health outcomes (e.g., lead, uranium) limited our ability to create a reproductive health-specific water quality index containing the full list of contaminants that the EPA has indicated could harm reproductive health or infants.[51] Frequent monitoring has costs and the EPA has already considered these costs when defining monitoring schedules for each contaminant. Still, our analysis has limitations (see Section 5.4 for discussion) due to the limited frequency of sampling, and this limitation applies to other analyses that may wish to inform regulators.

7 Conclusions

In this paper, we studied the effects of drinking water contamination on birth outcomes, using panel data on births within community water systems (CWS) for the state of Pennsylvania. In order to provide the most policy-relevant estimates possible, we restricted the sample to births that were not exposed to a water quality regulatory violation. We found that when we measure water contamination using either an omnibus measure of water quality or a measure removing microbial contaminants, birth outcomes respond negatively to water contamination, even in this no-violation-exposure sample. We find consistent evidence when using a reproductive-health-specific index and when we remove samples below the regulatory threshold (despite a substantial loss in sample size). In addition, we identify small negative effects on fertility. We find little evidence of mobility in response to contamination. This is consistent with our variation in water contamination not triggering significant avoidance behavior, which is intuitive since we have removed births exposed to regulatory violations that trigger public notification. Finally, we argue that current consumer confidence reports provided to consumers by CWS do not provide much information that can be easily understood by the consumer. Our index could be used to characterize the level of contamination in a simpler way and help consumers avoid exposure.

Although our findings are of short-run health consequences of poor water quality, even at levels that do not trigger regulatory violations, long-run effects are not a distant possibility. Our results call for further research on these important environmental and health policy issues. Future studies could repeat our analysis on different samples and fit models that critique and refine our index measures. In addition, previous research has shown that consumer confidence report policy changes can influence water system behavior (Bennear and Olmstead, 2008), while there is weaker evidence that violations can (Grooms, 2016); our findings motivate developing and adding omnibus water

[51]See Section 2.3.1 and Appendix Section A1.3 more details

quality measures to these consumer confidence reports, which could be similar to letter grading in K-12 education (e.g., Figlio and Lucas, 2004). Innovative accountability policies such as this may be able to improve water quality even for systems that are not in violation of regulatory standards, an outcome that our results suggest could improve public health.

References

Albouy-Llaty, Marion, Frédérike Limousi, Camille Carles, Antoine Dupuis, Sylvie Rabouan, and Virginie Migeot. 2016. "Association between exposure to endocrine disruptors in drinking water and preterm birth, taking neighborhood deprivation into account: a historic cohort study." *International journal of environmental research and public health*, 13(8): , p. 796.

Allaire, Maura, Taylor Mackay, Shuyan Zheng, and Upmanu Lall. 2019. "Detecting community response to water quality violations using bottled water sales." *Proceedings of the National Academy of Sciences*, 116(42): 20917–20922.

Allaire, Maura, Haowei Wu, and Upmanu Lall. 2018. "National trends in drinking water quality violations." *Proceedings of the National Academy of Sciences of the United States of America*.

Almberg, Kirsten S, Mary E Turyk, Rachael M Jones, Kristin Rankin, Sally Freels, and Leslie T Stayner. 2018. "Atrazine contamination of drinking water and adverse birth outcomes in community water systems with elevated atrazine in Ohio, 2006–2008." *International journal of environmental research and public health*, 15(9): , p. 1889.

Almond, Douglas, Kenneth Y. Chay, and David S. Lee. 2005. "The Costs of Low Birth Weight." *The Quarterly Journal of Economics*, 120(3): 1031–1083.

Almond, Douglas, Janet Currie, and Valentina Duque. 2018. "Childhood Circumstances and Adult Outcomes: Act II." *Journal of Economic Literature*, 56(4): 1360–1446.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling." *The Journal of Human Resources*, 40(4): 791–821.

Ananat, Elizabeth Oltmans, and Daniel M Hungerman. 2012. "The power of the pill for the next generation: Oral contraception's effects on fertility, abortion, and maternal and child characteristics." *Review of Economics and Statistics*, 94(1): 37–51.

Aschengrau, Ann, Lisa G Gallagher, Michael Winter, Lindsey Butler, M Patricia Fabian, and Veronica M Vieira. 2018. "Modeled exposure to tetrachloroethylene-contaminated drinking water and the occurrence of birth defects: a case-control study from Massachusetts and Rhode Island." *Environmental Health*, 17(1): 1–11.

Aschengrau, Ann, Michael R Winter, Lisa G Gallagher, Veronica M Vieira, Lindsey J Butler, M Patricia Fabian, Jenny L Carwile, Amelia K Wesselink, Shruthi Mahalingaiah, Patricia A Janulewicz et al. 2020. "Reproductive and developmental health effects of prenatal exposure to tetrachloroethylene-contaminated drinking water." *Environmental Science: Processes & Impacts*, 22(3): 555–566.

Baker, Jonathan, Lori Bennear, and Sheila Olmstead. 2022. "Does Information Disclosure Reduce Drinking Water Violations in the United States?" *Journal of the Association of Environmental and Resource Economists*.

Baker, Jonathan, Lori Bennear, and Sheila Olmstead. 2023. "Does information disclosure reduce drinking water violations in the united states?" *Journal of the Association of Environmental and Resource Economists*, 10(3): 787–818.

Beatty, Timothy K. M., and Jay P. Shimshack. 2014. "Air pollution and children's respiratory health: A cohort analysis." *Journal of Environmental Economics and Management*, 67(1): 39–57.

Behrman, Richard E, Adrienne Stith Butler et al. 2007. "Preterm birth: causes, consequences, and prevention."

Bennear, Lori S., Katrina K. Jessoe, and Sheila M. Olmstead. 2009. "Sampling Out: Regulatory Avoidance and the Total Coliform Rule." *Environmental Science & Technology*, 43(14): 5176–5182.

Bennear, Lori S., and Sheila M. Olmstead. 2008. "The impacts of the "right to know": Information disclosure and the violation of drinking water standards." *Journal of Environmental Economics and Management*, 56 117–130.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2007. "From the cradle to the labor market? The effect of birth weight on adult outcomes." *The Quarterly Journal of Economics*, 122(1): 409–439.

Bove, Frank J., Mark C. Fulcomer, Judith B. Klotz, Jorge Esmart, Ellen M. Dufficy, and Jonathan E. Savrin. 1995. "Public Drinking Water Contamination and Birth Outcomes." *American Journal of Epidemiology*, 141(9): 850–862.

Bozack, Anne K, Andres Cardenas, Quazi Quamruzzaman, Mahmuder Rahman, Golam Mostofa, David C Christiani, and Molly L Kile. 2018. "DNA methylation in cord blood as mediator of the association between prenatal arsenic exposure and gestational age." *Epigenetics*, 13(9): 923–940.

Bradley, Paul M, Dana W Kolpin, Kristin M Romanok, Kelly L Smalling, Michael J Focazio, Juliane B Brown, Mary C Cardon, Kurt D Carpenter, Steven R Corsi, Laura A DeCicco et al. 2018. "Reconnaissance of mixed organic and inorganic chemicals in private and public supply tapwaters at selected residential and workplace sites in the United States." *Environmental science & technology*, 52(23): 13972–13985.

Braun, Joseph M, Chris Gennings, Russ Hauser, and Thomas F Webster. 2016. "What can epidemiological studies tell us about the impact of chemical mixtures on human health?" *Environmental health perspectives*, 124(1): A6–A9.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro HC SantAnna. 2021. "Difference-in-differences with a continuous treatment." *arXiv preprint arXiv:2107.02637*.

Callaway, Brantly, and Pedro HC SantAnna. 2021. "Difference-in-differences with multiple time periods." *Journal of Econometrics*, 225(2): 200–230.

Christensen, Peter, David A. Keiser, and Gabriel E. Lade. 2023. "Economic Effects of Environmental Crises: Evidence from Flint, Michigan." *American Economic Journal: Economic Policy*, 15(1): 196–232.

Clay, Karen, Margarita Portnykh, and Edson Severnini. 2021. "Toxic Truth: Lead and Fertility." *Journal of the Association of Environmental and Resource Economists*, 8(5): 975–1012, Publisher: The University of Chicago Press.

Coffman, Vanessa R, Anja Søndergaard Jensen, Betina B Trabjerg, Carsten Bøcker Pedersen, Birgitte Hansen, Torben Sigsgaard, Jørn Olsen, Jörg Schullehner, Marie Pedersen, and Leslie T Stayner. 2022. "Prenatal exposure to nitrate from drinking water and the risk of preterm birth: A Danish nationwide cohort study." *Environmental Epidemiology*, 6(5): , p. e223.

Correia, Sergio. 2015. "Singletons, Cluster-Robust Standard Errors and Fixed Effects: A Bad Mix." Unpublished working paper.

Currie, Janet, Lucas Davis, Michael Greenstone, and Reed Walker. 2015. "Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings." *American Economic Review*, 105(2): 678–709.

Currie, Janet, and Enrico Moretti. 2007. "Biology as Destiny? Short- and Long-Run Determinants of Intergenerational Transmission of Birth Weight." *Journal of Labor Economics*, 25(2): 231–264.

Currie, Janet, and Reed Walker. 2011. "Traffic Congestion and Infant Health: Evidence from E-ZPass." *American Economic Journals: Applied Economics*, 3(1): 65–90.

Currie, Janet, and Reed Walker. 2019. "What do economists have to say about the Clean Air Act 50 years after the establishment of the Environmental Protection Agency?" *Journal of Economic Perspectives*, 33(4): 3–26.

Currie, Janet, Joshua Graff Zivin, Katherine Meckel, Matthew Neidell, and Wolfram Schlenker. 2013. "Something in the water: contaminated drinking water and infant health." *Canadian Journal of Economics*, 46(3): 719–810.

Danagoulian, Shooshan, Daniel Grossman, and David Slusky. 2022. "Health care following environmental disasters: evidence from Flint." *Journal of Policy Analysis and Management*, 41(4): 1060–1089.

Danagoulian, Shooshan, and Derek Jenkins. 2021. "Rolling back the gains: Maternal stress undermines pregnancy health after Flint's water switch." *Health Economics*, 30(3): 564–584.

Davalos, Angel D., Thomas J. Luben, Amy H. Herring, and Jason D. Sacks. 2017. "Current approaches used in epidemiologic studies to examine short-term multipollutant air pollution exposures." *Annals of Epidemiology*, 27(2): 145–153.e1.

Dave, Dhaval M, and Muzhe Yang. 2022. "Lead in drinking water and birth outcomes: A tale of two water treatment plants." *Journal of Health Economics*, 84, p. 102644.

De Chaisemartin, Clment, and Xavier dHaultfoeuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review*, 110(9): 2964–2996.

Ebdrup, Ninna Hinchely, Jörg Schullehner, Ulla Breth Knudsen, Zeyan Liew, Anne Marie Ladehoff Thomsen, Julie Lyngsø, Bjørn Bay, Linn Håkonsen Arendt, Pernille Jul Clemmensen, Torben Sigsgaard et al. 2022. "Drinking water nitrate and risk of pregnancy loss: a nationwide cohort study." *Environmental Health*, 21(1): 1–13.

Ferguson, Kelly K, Thomas F McElrath, and John D Meeker. 2014. "Environmental phthalate exposure and preterm birth." *JAMA pediatrics*, 168(1): 61–67.

Ferguson, Kelly K, Emma M Rosen, Zaira Rosario, Zlatan Feric, Antonia M Calafat, Thomas F McElrath, Carmen Vélez Vega, José F Cordero, Akram Alshawabkeh, and John D Meeker. 2019. "Environmental phthalate exposure and preterm birth in the PROTECT birth cohort." *Environment international*, 132, p. 105099.

Figlio, David N., and Maurice E. Lucas. 2004. "What's in a Grade? School Report Cards and the Housing Market." *American Economic Review*, 94(3): 591–604.

.

Forand, Steven P, Elizabeth L Lewis-Michl, and Marta I Gomez. 2012. "Adverse birth outcomes and maternal exposure to trichloroethylene and tetrachloroethylene through soil vapor intrusion in New York State." *Environmental health perspectives*, 120(4): 616–621.

Gennings, Chris, Huan Shu, Christina Rudén, Mattias Öberg, Christian Lindh, Hannu Kiviranta, and Carl-Gustaf Bornehag. 2018. "Incorporating regulatory guideline values in analysis of epidemiology data." *Environment international*, 120 535–543.

Gilliom, Robert J., and Pixie A. Hamilton. 2006. "Pesticides in the Nation's Streams and Ground Water, 1992âĂŞ2001âĂŤA Summary."Technical report, US Geological Survey.

González, Libertad. 2013. "The effect of a universal child benefit on conceptions, abortions, and early maternal labor supply." *American Economic Journal: Economic Policy*, 5(3): 160–188.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*, 225(2): 254–277.

Govarts, Eva, Nina Iszatt, Tomas Trnovec, Marijke de Cock, Merete EggesbÃÿ, Lubica Palkovicova Murinova, Margot van de Bor, MÃšnica Guxens, CÃľcile Chevrier, Gudrun Koppen, Marja Lamoree, Irva Hertz-Picciotto, Maria-Jose Lopez-Espinosa, Aitana Lertxundi, Joan O. Grimalt, Maties Torrent, Fernando GoÃśi-Irigoyen, Roel Vermeulen, Juliette Legler, and Greet Schoeters. 2018. "Prenatal exposure to endocrine disrupting chemicals and risk of being born small for gestational age: Pooled analysis of seven European birth cohorts." *Environment International*, 115 267–278.

Grooms, Katherine K. 2016. "Does Water Quality Improve When a Safe Drinking Water Act Violation Is Issued? A Study of the Effectiveness of the SDWA in California." *BE J. Econ. Anal. Policy*, 16(1): 1–23.

Grossman, Daniel S, and David JG Slusky. 2019. "The impact of the Flint water crisis on fertility." *Demography*, 56(6): 2005–2031.

Grossman, Daniel, and David Slusky. 2017. "The Effect of an Increase in Lead in the Water System on Fertility and Birth Outcomes: The Case of Flint, Michigan." Available at http://www2.ku.edu/~kuwpaper/2017Papers/201703.pdf, accessed 2018-09-02.

Guldi, Melanie. 2008. "Fertility effects of abortion and birth control pill access for minors." *Demography*, 45 817–827.

Hill, Elaine L, and Lala Ma. 2022. "Drinking water, fracking, and infant health." *Journal of Health Economics*, 82, p. 102595.

Horton, Bethany Jablonski, Thomas J Luben, Amy H Herring, David A Savitz, Philip C Singer, Howard S Weinberg, and Katherine E Hartmann. 2011. "The effect of water disinfection by-products on pregnancy outcomes in two southeastern US communities." *Journal of occupational and environmental medicine/American College of Occupational and Environmental Medicine*, 53(10): , p. 1172.

Keiser, David A, and Joseph S Shapiro. 2019a. "Consequences of the Clean Water Act and the demand for water quality." *The Quarterly Journal of Economics*, 134(1): 349–396.

Keiser, David A, and Joseph S Shapiro. 2019b. "US Water Pollution Regulation over the Past Half Century: Burning Waters to Crystal Springs?" *Journal of Economic Perspectives*, 33(4): 51–75.

Kile, Molly L., Andres Cardenas, Ema Rodrigues, Maitreyi Mazumdar, Christine Dobson, Mostofa Golam, Quazi Quamruzzaman, Mahmudar Rahman, and David C. Christiani. 2016. "Estimating Effects of Arsenic Exposure During Pregnancy on Perinatal Outcomes in a Bangladeshi Cohort." *Epidemiology (Cambridge, Mass.)*, 27(2): 173–181.

Knittel, Christopher R., Douglas L. Miller, and Nicholas J. Sanders. 2016. "Caution, Drivers! Children Present: Traffic, Pollution, and Infant Health." *Review of Economics & Statistics*, 98(2): 350–366.

Manassaram, Deana M., Lorraine C. Backer, and Deborah M. Moll. 2006. "A Review of Nitrates in Drinking Water: Maternal Exposure and Adverse Reproductive and Developmental Outcomes." *Environmental Health Perspectives*, 114(3): 320–327.

Marcus, Michelle. 2019. "Going Beneath the Surface: Petroleum Pollution, Regulation, and Health." *Working Paper*.

Marcus, Michelle. 2022. "Testing the Water: Drinking Water Quality, Public Notification, and Child Outcomes." *The Review of Economics and Statistics*, 104(6): 1289–1303.

Mashau, Funanani, Esper Jacobeth Ncube, and Kuku Voyi. 2018. "Drinking water disinfection by-products exposure and health effects on pregnancy outcomes: a systematic review." *Journal of water and health*, 16(2): 181–196.

McDonald, Yolanda J, Kayla M Anderson, Mariah D Caballero, Ke Jack Ding, Douglas H Fisher, Caroline P Morkel, and Elaine L Hill. 2022. "A systematic review of geospatial representation of United States community water systems." *AWWA Water Science*, 4(1): , p. e1266.

McDonald, Yolanda J, and Nicole E Jones. 2018. "Drinking water violations and environmental justice in the United States, 2011–2015." *American Journal of Public Health*, 108(10): 1401–1407.

Migeot, V, M Albouy-Llaty, C Carles, F Limousi, S Strezlec, A Dupuis, and S Rabouan. 2013. "Drinking-water exposure to a mixture of nitrate and low-dose atrazine metabolites and small-for-gestational age (SGA) babies: a historic cohort study." *Environmental research*, 122 58–64.

Murphy, Eileen A, Gloria B Post, Brian T Buckley, Robert L Lippincott, and Mark G Robson. 2012. "Future challenges to protecting public health from drinking-water contaminants." *Annual review of public health*, 33 209–224.

Myers, Sharon L, Danelle T Lobdell, Zhiyi Liu, Yajuan Xia, Haixia Ren, Yuxing Li, Richard K Kwok, Judy L Mumford, and Pauline Mendola. 2010. "Maternal drinking water arsenic exposure and perinatal outcomes in inner Mongolia, China." *Journal of Epidemiology & Community Health*, 64(4): 325–329.

Oakes, Michelle, Lisa Baxter, and Thomas C. Long. 2014. "Evaluating the application of multipollutant exposure metrics in air pollution health studies." *Environment International*, 69 90–99.

Oreopoulos, Philip, Mark Stabile, Randy Walld, and Leslie L. Roos. 2008. "Short-, Medium-, and Long-Term Consequences of Poor Infant Health An Analysis Using Siblings and Twins." *The Journal of Human Resources*, 43(1): 88–138.

Oster, Emily. 2016. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics*, 0(0): 1–18.

Porpora, Maria Grazia, Ilaria Piacenti, Sara Scaramuzzino, Luisa Masciullo, Francesco Rech, and Pierluigi Benedetti Panici. 2019. "Environmental contaminants exposure and preterm birth: a systematic review." *Toxics*, 7(1): , p. 11.

Rahman, Mohammad L, Linda Valeri, Molly L Kile, Maitreyi Mazumdar, Golam Mostofa, Qazi Qamruzzaman, Mahmudur Rahman, Andrea Baccarelli, Liming Liang, Russ Hauser et al. 2017. "Investigating causal relation between prenatal arsenic exposure and birthweight: Are smaller infants more susceptible?" *Environment international*, 108 32–40.

Rinsky, Jessica L, Claudia Hopenhayn, Vijay Golla, Steve Browning, and Heather M Bush. 2012. "Atrazine exposure in public drinking water and preterm birth." *Public health reports*, 127(1): 72–80.

Rohr, Jason R., and Krista A. McCoy. 2010. "A Qualitative Meta-Analysis Reveals Consistent Effects of Atrazine on Freshwater Fish and Amphibians." *Environmental Health Perspectives*, 118(1): 20–32.

Royer, Heather. 2009. "Separated at Girth: US Twin Estimates of the Effects of Birth Weight." *American Economic Journal: Applied Economics*, 1(1): 49–85.

Russell, Rebecca B, Nancy S Green, Claudia A Steiner, Susan Meikle, Jennifer L Howse, Karalee Poschman, Todd Dias, Lisa Potetz, Michael J Davidoff, Karla Damus et al. 2007. "Cost of hospitalization for preterm and low birth weight infants in the United States." *Pediatrics*, 120(1): e1–e9.

Saha, Kuntal K, Annette Engström, Jena Derakhshani Hamadani, Fahmida Tofail, Kathleen M Rasmussen, and Marie Vahter. 2012. "Pre-and postnatal arsenic exposure and body size to 2 years of age: a cohort study in rural Bangladesh." *Environmental health perspectives*, 120(8): 1208–1214.

Sanders, Alison P, Tania A Desrosiers, Joshua L Warren, Amy H Herring, Dianne Enright, Andrew F Olshan, Robert E Meyer, and Rebecca C Fry. 2014. "Association between arsenic, cadmium, manganese, and lead levels in private wells and birth defects prevalence in North Carolina: a semi-ecologic study." *BMC Public Health*, 14.

Savitz, David A, Philip C Singer, Amy H Herring, Katherine E Hartmann, Howard S Weinberg, and Christina Makarushka. 2006. "Exposure to drinking water disinfection by-products and pregnancy loss." *American Journal of Epidemiology*, 164(11): 1043–1051.

Schlenker, Wolfram, and Michael J. Roberts. 2009. "Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change." *Proceedings of the National Academy of Sciences*, 106(37): 15594–15598.

Sherris, Allison R, Michael Baiocchi, Scott Fendorf, Stephen P Luby, Wei Yang, and Gary M Shaw. 2021. "Nitrate in drinking water during pregnancy and spontaneous preterm birth: a retrospective within-mother analysis in California." *Environmental health perspectives*, 129(5): , p. 057001.

Snowden, Jonathan M., Colleen E. Reid, and Ira B. Tager. 2015. "Framing Air Pollution Epidemiology in Terms of Population Interventions, with Applications to Multipollutant Modeling." *Epidemiology*, 26(2): 271–279.

Sonnenfeld, N, I Hertz-Picciotto, and WE Kaye. 2001. "Tetrachloroethylene in drinking water and birth outcomes at the US Marine Corps Base at Camp Lejeune, North Carolina." *American journal of epidemiology*, 154(10): 902–908.

Stayner, Leslie Thomas, Kirsten Almberg, Rachael Jones, Judith Graber, Marie Pedersen, and Mary Turyk. 2017. "Atrazine and nitrate in drinking water and the risk of preterm delivery and low birth weight in four Midwestern states." *Environmental research*, 152 294–303.

US Environmental Protection Agency. 2006. “IDL-MDL-PQL "What the "L" is going on? What does all this alphabet soup really mean?.” Available at https://www.epa.gov/sites/production/files/2015-06/documents/whatthel.pdf, accessed 2018-08-05.

US Environmental Protection Agency. 2016a. “Development of Estimated Quantitation Levels for the Third Six-Year Review of National Primary Drinking Water Regulations (Chemical Phase Rules).” p. 55, Available at https://www.epa.gov/sites/production/files/2016-12/documents/810r16002.pdf, accessed 2018-09-02.

US Environmental Protection Agency. 2016b. “EPA Protocol for the Third Review of Existing National Primary Drinking Water Regulations.” p. 43.

US Environmental Protection Agency. 2018. “National Primary Drinking Water Regulations.” Available at https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations, accessed 2018-09-02.

US Environmental Protection Agency. 2019a. “Population Served by Community Water Systems with No Reported Violations of Health-Based Standards.” Available at https://cfpub.epa.gov/roe/indicator.cfm?i=45, accessed 1-14-2020.

US Environmental Protection Agency. 2019b. “Title XIV of The Public Health Service Act: Safety of Public Water Systems (Safe Drinking Water Act).” Available at https://www.epa.gov/sdwa/title-xiv-public-health-service-act-safety-public-water-systems-safe-drinking-water-act-0, accessed 2022-12-13.

US EPA, OEI. 2013. “Toxics Release Inventory (TRI) Program.” January.

Wen, Xiao-Jun, Lina Balluz, and Ali Mokdad. 2009. “Association between media alerts of air quality index and change of outdoor activity among adult asthma in six states, BRFSS, 2005.” *Journal of community health*, 34(1): 40–46.

Werner, Rachel M, R Tamara Konetzka, and Daniel Polsky. 2016. “Changes in consumer demand following public reporting of summary quality ratings: an evaluation in nursing homes.” *Health services research*, 51 1291–1309.

Wigle, Donald T., Tye E. Arbuckle, Michelle C. Turner, Annie BÃľrubÃľ, Qiuying Yang, Shiliang Liu, and Daniel Krewski. 2008. “Epidemiologic Evidence of Relationships Between Reproductive and Child Health Outcomes and Environmental Chemical Contaminants.” *Journal of Toxicology and Environmental Health, Part B*, 11 373–517.

Yang, Muzhe, and Shin-Yi Chou. 2018. “The impact of environmental regulation on fetal health: Evidence from the shutdown of a coal-fired power plant located upwind of New Jersey.” *Journal of Environmental Economics and Management*, 90 269–293.

Young, Holly A, Korine N Kolivras, Leigh-Anne H Krometis, Cristina E Marcillo, and Julia M Gohlke. 2022. “Examining the association between safe drinking water act violations and adverse birth outcomes in Virginia.” *Environmental Research*, p. 114977.

Zivin, Joshua Graff, and Matthew Neidell. 2009. "Days of haze: Environmental information disclosure and intertemporal avoidance behavior." *Journal of Environmental Economics and Management*, 58(2): 119–128.

Zivin, Joshua Graff, Matthew Neidell, and Wolfram Schlenker. 2011. "Water Quality Violations and Avoidance Behavior: Evidence from Bottled Water Consumption." *American Economic Review*, 101(3): 448–453.

8 Tables and Figures

Figure 1: Map of Pennsylvania and 1,768 CWS Service Areas

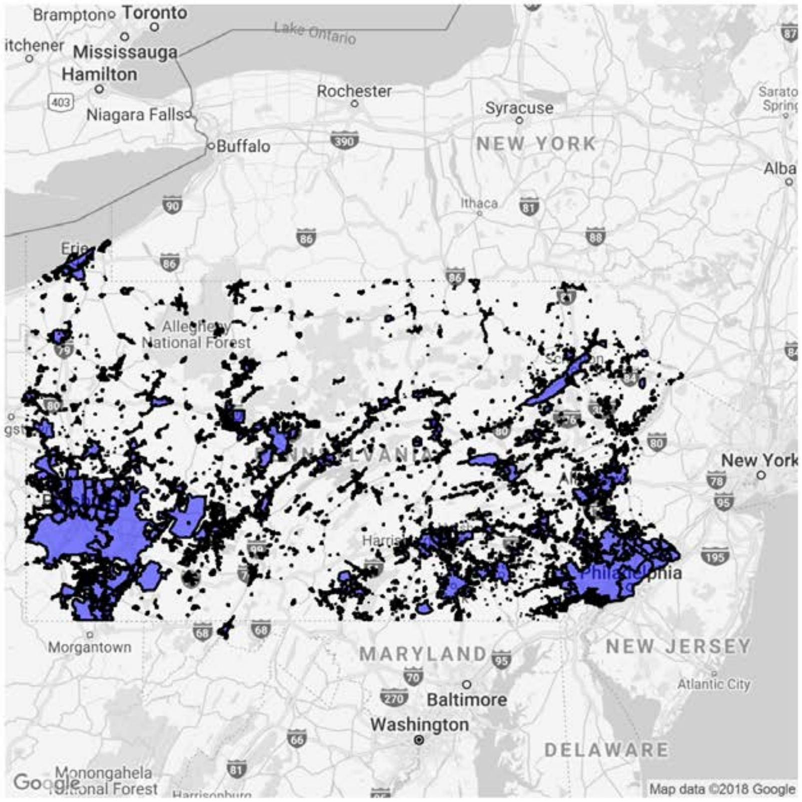

Table 1: Distribution of drinking water contamination experienced by Pennsylvanian births during gestation.

Contaminant	(1) Mean	(2) SD	(3) 10th	(4) 25th	(5) 50th	(6) 75th	(7) 90th	(8) Max	(9) Births
Panel A: Trimmed sample (births used in analysis)									
All	0.08	0.08	0.02	0.03	0.05	0.08	0.17	0.50	1,216,132
Chemical	0.07	0.08	0.02	0.03	0.04	0.07	0.15	0.87	1,214,709
Repro. Index	0.08	0.09	0.00	0.02	0.06	0.11	0.17	2.80	1,147,893
Panel B: Raw sample (including births not used)									
All	0.09	2.24	0.02	0.03	0.05	0.09	0.18	1425.00	1,228,531
Chemical	0.09	2.24	0.02	0.03	0.04	0.07	0.17	1425.00	1,227,035
Repro. Index	0.08	0.59	0.00	0.02	0.06	0.11	0.18	193.80	1,151,025
Panel C: Trimmed sample, after removing mother fixed-effects									
All	-0.00	0.05	-0.05	-0.02	-0.00	0.01	0.05	0.37	700,951
Chemical	-0.00	0.05	-0.05	-0.01	-0.00	0.01	0.04	0.54	699,656
Repro. Index	-0.00	0.05	-0.05	-0.02	0.00	0.02	0.05	1.26	645,353

Notes: Table displays the average sample result relative to MCL (RRMCL) over gestational periods for births in Pennsylvania, 2003-2014. Throughout, the table restricts to non-plural births not exposed to MCL violations. Panel A trims the sample of births at the 99th percentile (separately for each contaminant group). This is the sample we use throughout the paper. Panel B shows the raw statistics for comparison. Panel C shows the distribution after removing mother fixed-effects, which is relevant given our causal design. "All" and "Chemical" are our indexes described in Section 2.3; "Repro. Index" is our five-contaminant index described in Section 2.3.1.

Table 2: Descriptive Statistics of Births in Pennsylvania, 2003-2014

Characteristic	(1) Non-Plural No MCL	(2) Analytic Sample	(3) Tails of Contamination Distribution ≤ 10th	(4) Tails of Contamination Distribution ≥ 90th	(5) Mom FE Sample
Sample Sizes					
No. of births	1,241,656	1,216,132	122,856	110,523	700,815
No. of moms	832,674	820,173	112,209	103,069	304,934
No. of CWSs	1,537	1,459	1,302	885	1,343
Outcomes					
Low birth weight	0.067	0.067	0.057	0.079	0.065
Preterm birth	0.082	0.082	0.074	0.102	0.081
Small for gestational age	0.095	0.096	0.083	0.095	0.091
Term birth weight (g)	3,396	3,396	3,423	3,389	3,402
	(468.9)	(468.8)	(465.0)	(467.5)	(467.2)
CWS births per 100k per month	113.4	112.7	141.0	109.6	112.1
	(107.4)	(103.3)	(223.4)	(108.1)	(98.59)
Other characteristics					
Mom age (years)	27.87	27.87	28.01	27.80	27.69
	(6.054)	(6.057)	(5.854)	(6.023)	(5.858)
Mom Black	0.210	0.213	0.074	0.185	0.220
Mom Hispanic	0.068	0.069	0.038	0.060	0.074
Mom white, not Hispanic	0.722	0.718	0.889	0.757	0.713
HS or less	0.401	0.401	0.377	0.402	0.406
Mom smokes	0.224	0.223	0.262	0.238	0.216
Mom married	0.569	0.568	0.642	0.581	0.578
WIC/Medicaid	0.477	0.478	0.422	0.459	0.482
Different CWS next birth	0.369	0.368	0.456	0.356	0.368
Mean RRCML, all contaminants	0.095	0.078	0.014	0.285	0.078
	(2.237)	(0.078)	(0.006)	(0.082)	(0.077)
Mean RRCML, chemical only	0.085	0.069	0.013	0.281	0.069
	(2.238)	(0.080)	(0.006)	(0.112)	(0.079)

Notes: Sample consists of births matched to a public water system. Means (for indicator variables, proportions) are displayed, with standard deviations in parentheses when appropriate. Column (1) provides statistics for non-plural births without MCLs. Column (2) restricts to our analytic sample, keeping only births with an overall measure of contamination and for which that measure is below the 99th percentile. Columns (3) and (4) describe births in the tails of the overall contamination distribution (among births in column (2)). Column (5) describes births in our mother fixed-effects specifications.

Table 3: Effects of drinking water contamination on low birth weight and preterm birth

	(1)	(2)	(3)	(4)	(5)	(6)
	No controls		Adding mother ctrls		Adding temp & TRI ctrls	
	All	Chemical	All	Chemical	All	Chemical
Panel A: low birth weight (LBW)						
CWS FE						
10th to 90th RRMCL	0.01356***	0.01456***	0.01249***	0.01349***	0.01059***	0.01050***
	(0.00302)	(0.00133)	(0.00285)	(0.00124)	(0.00240)	(0.00140)
Observations	1,216,132	1,214,688	1,216,132	1,214,688	1,216,132	1,214,688
Adj R^2	0.0050	0.0052	0.0456	0.0458	0.1058	0.1059
AETO's δ			32.59	34.50	10.78	7.82
Mom & CWS FE						
10th to 90th RRMCL	0.00957***	0.01048***	0.00932***	0.01003***	0.00819***	0.00774***
	(0.00245)	(0.00120)	(0.00232)	(0.00115)	(0.00198)	(0.00133)
Observations	700,815	699,546	700,815	699,546	700,815	699,546
Adj R^2	0.1506	0.1508	0.1689	0.1691	0.2103	0.2104
AETO's δ			2.29	1.43	1.13	0.55
Panel B: preterm birth (PTB)						
CWS FE						
10th to 90th RRMCL	0.01997***	0.02134***	0.01866***	0.02001***	0.01626***	0.01623***
	(0.00405)	(0.00178)	(0.00375)	(0.00164)	(0.00306)	(0.00175)
Observations	1,216,132	1,214,688	1,216,132	1,214,688	1,216,132	1,214,688
Adj R^2	0.0036	0.0040	0.0467	0.0471	0.1316	0.1318
AETO's δ			40.16	41.35	13.38	9.72
Mom & CWS FE						
10th to 90th RRMCL	0.01571***	0.01716***	0.01533***	0.01658***	0.01405***	0.01377***
	(0.00363)	(0.00166)	(0.00345)	(0.00160)	(0.00276)	(0.00170)
Observations	700,815	699,546	700,815	699,546	700,815	699,546
Adj R^2	0.1486	0.1489	0.1663	0.1665	0.2252	0.2253
AETO's δ			2.44	1.79	2.00	0.99

Notes: standard errors in parentheses; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors are twoway clustered on public water system and mother. Each cell is a separate regression. Observations (number of births) and adjusted R^2s (for the full model, i.e. including the fixed-effects) are reported. In each regression, the sample is restricted to births in our analytic sample for whom we observe at least one water quality sample of the given contaminant group during gestation in the public water system of residence at the time of birth. Control variables used vary over super-columns; independent variable, i.e. the contaminant group studied, vary across columns; the outcomes varies over panels; the specification used (either public water system fixed-effects or mom fixed-effects) varies over sub-panels; temp & TRI ctrls = weather and toxics release inventory (air pollution) controls.

Table 4: Effects of contamination on small for gestational age and term birth weight

	(1)	(2)	(3)	(4)	(5)	(6)
	No controls		Adding mother ctrls		Adding temp & TRI ctrls	
	All	Chemical	All	Chemical	All	Chemical
Panel A: small for gestational age (SGA)						
CWS FE						
10th to 90th RRMCL	0.00341***	0.00268***	0.00282***	0.00217***	0.00258***	0.00183***
	(0.00077)	(0.00064)	(0.00074)	(0.00061)	(0.00079)	(0.00062)
Observations	1,214,205	1,212,766	1,214,205	1,212,766	1,214,205	1,212,766
Adj R^2	0.0054	0.0054	0.0381	0.0381	0.0390	0.0390
AETO's δ			13.15	11.59	8.30	5.86
Mom & CWS FE						
10th to 90th RRMCL	0.00279***	0.00179**	0.00276***	0.00165**	0.00290***	0.00173**
	(0.00100)	(0.00083)	(0.00096)	(0.00081)	(0.00099)	(0.00083)
Observations	698,761	697,487	698,761	697,487	698,761	697,487
Adj R^2	0.1386	0.1393	0.1542	0.1548	0.1544	0.1550
AETO's δ			6.36	0.65	-1.90	1.26
Panel B: term birth weight (TBW)						
CWS FE						
10th to 90th RRMCL	-10.61***	-10.04***	-9.75***	-9.44***	-9.54***	-8.77***
	(1.39)	(1.25)	(1.55)	(1.21)	(1.69)	(1.32)
Observations	1,116,570	1,115,266	1,116,570	1,115,266	1,116,570	1,115,266
Adj R^2	0.0169	0.0169	0.0943	0.0944	0.0987	0.0988
AETO's δ			30.56	42.24	22.73	17.70
Mom & CWS FE						
10th to 90th RRMCL	-9.05***	-8.82***	-9.15***	-8.58***	-9.33***	-8.25***
	(1.73)	(1.34)	(1.84)	(1.37)	(1.87)	(1.46)
Observations	612,960	611,893	612,960	611,893	612,960	611,893
Adj R^2	0.3437	0.3437	0.3827	0.3828	0.3848	0.3849
AETO's δ			-10.33	3.87	-5.48	1.47

Notes: see notes to Table 3.

Table 5: Effects of drinking water contamination on low birth weight and preterm birth, using five-contaminant reproductive index.

	(1)	(2)	(3)	(4)	(5)	(6)
	No controls		Adding mother ctrls		Adding temp & TRI ctrls	
	All	5 Contam	All	5 Contam	All	5 Contam
Panel A: low birth weight						
CWS FEs						
10th to 90th RRMCL	0.01356***	0.01936***	0.01249***	0.01833***	0.01059***	0.01406**
	(0.00302)	(0.00639)	(0.00285)	(0.00580)	(0.00240)	(0.00576)
Observations	1,216,132	1,139,430	1,216,132	1,139,430	1,216,132	1,139,430
Adj R^2	0.0050	0.0049	0.0456	0.0452	0.1058	0.1049
Mom & CWS FEs						
10th to 90th RRMCL	0.00957***	0.01709***	0.00932***	0.01637***	0.00819***	0.01282***
	(0.00245)	(0.00405)	(0.00232)	(0.00381)	(0.00198)	(0.00411)
Observations	700,815	637,615	700,815	637,615	700,815	637,615
Adj R^2	0.1506	0.1507	0.1689	0.1686	0.2103	0.2094
Panel B: preterm birth						
CWS FEs						
10th to 90th RRMCL	0.01997***	0.02743***	0.01866***	0.02621***	0.01626***	0.02096***
	(0.00405)	(0.00863)	(0.00375)	(0.00798)	(0.00306)	(0.00788)
Observations	1,216,132	1,139,430	1,216,132	1,139,430	1,216,132	1,139,430
Adj R^2	0.0036	0.0034	0.0467	0.0460	0.1316	0.1306
Mom & CWS FEs						
10th to 90th RRMCL	0.01571***	0.02370***	0.01533***	0.02279***	0.01405***	0.01834***
	(0.00363)	(0.00525)	(0.00345)	(0.00497)	(0.00276)	(0.00543)
Observations	700,815	637,615	700,815	637,615	700,815	637,615
Adj R^2	0.1486	0.1487	0.1663	0.1659	0.2252	0.2247

Notes: see notes to Table 3. Results for the overall index included for comparison, which are the same as the "All" columns in Table 3.

Table 6: Effects of drinking water contamination on small for gestational age and term birth weight, using five-contaminant reproductive index.

	(1)	(2)	(3)	(4)	(5)	(6)
	No controls		Adding mother ctrls		Adding temp & TRI ctrls	
	All	5 Contam	All	5 Contam	All	5 Contam
Panel A: small for gestational age						
CWS FEs						
10th to 90th RRMCL	0.00341***	0.00511***	0.00282***	0.00470***	0.00258***	0.00390***
	(0.00077)	(0.00143)	(0.00074)	(0.00125)	(0.00079)	(0.00138)
Observations	1,214,205	1,137,738	1,214,205	1,137,738	1,214,205	1,137,738
Adj R^2	0.0054	0.0053	0.0381	0.0380	0.0390	0.0390
Mom & CWS FEs						
10th to 90th RRMCL	0.00279***	0.00349**	0.00276***	0.00350**	0.00290***	0.00317*
	(0.00100)	(0.00164)	(0.00096)	(0.00167)	(0.00099)	(0.00168)
Observations	698,761	635,832	698,761	635,832	698,761	635,832
Adj R^2	0.1386	0.1405	0.1542	0.1559	0.1544	0.1561
Panel B: term birth weight						
CWS FEs						
10th to 90th RRMCL	-10.61***	-13.86***	-9.75***	-13.92***	-9.54***	-12.58***
	(1.39)	(3.94)	(1.55)	(3.57)	(1.69)	(3.91)
Observations	1,116,570	1,047,149	1,116,570	1,047,149	1,116,570	1,047,149
Adj R^2	0.0169	0.0167	0.0943	0.0945	0.0987	0.0989
Mom & CWS FEs						
10th to 90th RRMCL	-9.05***	-13.74***	-9.15***	-13.86***	-9.33***	-12.59***
	(1.73)	(2.79)	(1.84)	(2.82)	(1.87)	(2.99)
Observations	612,960	558,355	612,960	558,355	612,960	558,355
Adj R^2	0.3437	0.3483	0.3827	0.3867	0.3848	0.3887

Notes: see notes to Table 3. Results for the overall index included for comparison, which are the same as the “All” columns in Table 4.

Table 7: Effects of drinking water contamination on mobility and fertility.

	(1)	(2)	(3)	(4)	(5)	(6)
	No controls		Adding mother controls		Adding weather controls	
	All	Chemical	All	Chemical	All	Chemical
Panel A: Mobility − CWS FE						
10th to 90th RRMCL	-0.00108	-0.00014	-0.00165	-0.00068	-0.00069	-0.00100
	(0.00189)	(0.00149)	(0.00178)	(0.00151)	(0.00152)	(0.00152)
Observations	413,135	412,785	413,135	412,785	413,135	412,785
Adj R^2	0.1271	0.1271	0.1470	0.1470	0.1472	0.1472
Panel B: Mobility − Mom FEs						
10th to 90th RRMCL	0.00255	0.00348	0.00165	0.00297	0.00298	0.00280
	(0.00315)	(0.00260)	(0.00309)	(0.00251)	(0.00252)	(0.00257)
Observations	172,651	172,498	172,651	172,498	172,651	172,498
Adj R^2	0.2494	0.2486	0.2694	0.2689	0.2696	0.2691
Panel C: Fertility (birth rate per 100,000 people served) − CWS FE						
10th to 90th RRMCL	-0.07271**	-0.08445**	-0.06937**	-0.07647**	-0.07472**	-0.07040**
	(0.03491)	(0.03351)	(0.02934)	(0.03040)	(0.02939)	(0.02799)
Observations	59,763	59,406	59,763	59,406	59,763	59,406
Adj R^2	0.7397	0.7393	0.7420	0.7417	0.7422	0.7418
Panel D: Fertility (log number of births) − CWS FEs						
10th to 90th RRMCL	-0.00114***	-0.00095***	-0.00114***	-0.00089***	-0.00122***	-0.00093***
	(0.00041)	(0.00033)	(0.00036)	(0.00031)	(0.00037)	(0.00031)
Observations	33,122	32,912	33,122	32,912	33,122	32,912
Adj R^2	0.9924	0.9925	0.9925	0.9926	0.9925	0.9926

Notes: standard errors in parentheses; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors are twoway clustered on public water system and mother in panels A and B, and clustered on public water system in panel C. Observations (number of births in panel A and B; public water system months in panel C) and adjusted R^2s (inclusive of fixed-effects) are reported. Each cell is a separate regression. In panels A and B, the sample is our analytic sample; in panel A, mothers must have two or more births in the data (to measure mobility), with at least one having a measure of contamination, to be included. Panel B requires mothers to have three or more births in our sample due to the addition of mom fixed-effects. In both panels A and B, the dependent variable (measured for each birth) is whether the mother is in a different public water system at the time of the next birth (being in no public water system in the next birth is considered being in a different water system for this purpose). Observations in panel C are public water system quarters, and the dependent variable is the number of births per 100,000 people served by the public water system. Regressions in panel C are weighted by public water system population served. Regressions in panel D are the same as panel C, except the outcome variable is the natural log of the number of births (community water system quarters with zero births are dropped). Control variables used vary over super-columns; independent variables, i.e. the contaminant group studied, vary across columns.

Figure 2: Philadelphia Water Consumer Confidence Report Summary Results, 2017

www.phila.gov/water • @PhillyH2O

2017 Drinking Water Quality Results

Listed on pages 16 – 18 are our Drinking Water Quality Results for 2017. All results are better than the recommended federal levels designed to protect public health. By reporting these results in the tables below, we are meeting a requirement of the EPA. Please see the glossary on page 19 for definitions of abbreviations used in the tables. Some contaminants may pose a health risk at certain levels to people with special health concerns. Others are used as indicators for treatment plant performance. For more information, please visit our website at www.phila.gov/water or call us at 215.685.6300.

LEAD AND COPPER - *Tested at Customers' Taps - Testing is done every 3 years. Most recent tests were done in 2017.*

	EPA's Action Level - for a representative sampling of customer homes	Ideal Goal (EPA's MCLG)	90% of PWD customers' homes were less than	Number of homes considered to have elevated levels	Violation	Source
Lead	90% of homes must test less than 15 ppb	0 ppb	2.0 ppb	3 out of 89	No	Corrosion of household plumbing; Erosion of natural deposits
Copper	90% of homes must test less than 1.3 ppm	1.3 ppm	0.23 ppm	1 out of 89	No	Corrosion of household plumbing; Erosion of natural deposits; Leaching from wood preservatives

CRYPTOSPORIDIUM - *Tested at Source Water to Water Treatment Plants Prior to Treatment*

Treatment Technique Requirement	Baxter WTP One Year Range	Belmont WTP One Year Range	Queen Lane WTP One Year Range	Source
Total Number of Samples Collected	6	6	6	Naturally present in the environment.
Number of Cryptosporidium Detected	15	2	6	
	0.250 count/L	0.033 count/L	0.100 count/L	

Cryptosporidium is a microbial pathogen found in surface water throughout the U.S. Although filtration removes Cryptosporidium, the most commonly-used filtration methods cannot guarantee 100 percent removal. Our monitoring indicates the presence of these organisms in our source water. Current test methods do not allow us to determine if the organisms are dead or if they are capable of causing disease.

BACTERIA IN TAP WATER - *Tested throughout the Distribution System. Over 460 samples collected throughout the City every month.*

	Highest Level Allowed (EPA's MCL)	Ideal Goal (EPA's MCLG)	Highest Monthly % or Yearly Total of Positive Samples	Monthly Range (% or #)	Violation	Source
Total Coliform	5% of monthly samples are positive*	0	0.75%	0 – 0.75%	No	Naturally present in the environment
Fecal Coliform or E. coli		0	1	0 – 1	No	Human or animal fecal waste

Every sample that is positive for total coliforms must also be analyzed for E. coli. If a system has two consecutive total coliform positive samples, and one is also positive for E. coli then the system has an MCL violation. There were no Level 1 and Level 2 assessments required under Revised Total Coliform Rule in 2017.

INORGANIC CHEMICALS (IOC) - *PWD monitors for IOC more often than required by EPA.*

Chemical	Highest Level Allowed (EPA's MCL)	Ideal Goal (EPA's MCLG)	Highest Result	Range of Test Results for the Year	Violation	Source
Antimony	6 ppb	6 ppb	0.4 ppb	0 - 0.4 ppb	No	Discharge from petroleum refineries; fire retardants; ceramics; electronics; solder
Barium	2 ppm	2 ppm	0.044 ppm	0.027 – 0.044 ppm	No	Discharges of drilling wastes; Discharge from metals refineries; Erosion of natural deposits
Chromium	100 ppb	100 ppb	1 ppb	0 – 1 ppb	No	Discharge from steel and pulp mills; Erosion of natural deposits
Fluoride	2 ppm*	2 ppm*	0.72 ppm	0.68 – 0.72 ppm	No	Erosion of natural deposits; Water additive which promotes strong teeth; Discharge from fertilizer and aluminum factories

Table 8: Largest systems by water quality letter grades based on our index measure

Water System Name	(1) Population Served	(2) Mean RRMCL	(3) Letter Grade
Ambler Boro Water Dept	20,000	0.02	A
Lansford Coaldale Jt Water Aut	9,300	0.02	A
Paw Bangor District	9,008	0.02	A
Erie City Water Authority	180,000	0.03	B
York Water Co	159,623	0.04	B
North Penn Water Authority	82,822	0.03	B
Pittsburgh Water & Sewer Auth	250,000	0.06	C
West View Boro Muni Auth	200,000	0.05	C
Westmd Mun Auth-Sweeney Plant	140,000	0.06	C
Philadelphia Water Department	1,600,000	0.07	D
Pa Amer Water Co-Pittsburgh	507,675	0.08	D
City Of Lancaster	120,000	0.09	D
Aqua Pa Main System	784,939	0.06	F
Easton Area Water System	93,400	0.05	F
Capital Region Water	66,540	0.06	F

Notes: Letter grades assigned based on average result relative to MCL for all contaminants available since 2010 (reported in the mean RRMCL column): the cleanest 25% public water systems are assigned grade “A”, the next 25% are assigned grade “B”, and so forth.

Table 9: Possible effects of an information provision intervention - assigning letter grades to systems on consumer confidence reports.

	(1)	(2)	(3)
	A or B	C, D, or F	Overall
Panel A: Low birth weight			
Actual rates	5.73	6.90	6.67
After policy change	5.73	6.60	6.43
Panel B: Pre-term birth			
Actual rates	7.28	8.33	8.13
After policy change	7.28	7.81	7.71
Number of births	242,388	1,014,850	1,257,238

Notes: We assume that people in systems assigned to letter grades C, D, or F engage in avoidance behavior and obtain water quality at the average for letter grade A systems. Estimates give an idea of the magnitude of public health gains that might be obtained from adopting a consumer confidence report simplification policy.