I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12115

Something in the Pipe: Flint Water Crisis and Health at Birth

Rui Wang
Xi Chen
Xun Li

JANUARY 2019

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12115

Something in the Pipe: Flint Water Crisis and Health at Birth

Rui Wang
Tulane University

Xi Chen
Yale University and IZA

Xun Li
Wuhan University

JANUARY 2019

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Something in the Pipe: Flint Water Crisis and Health at Birth

Flint changed its public water source in 2014, causing severe water contamination. We estimate the effect of in utero exposure to polluted water on health at birth using the recent Flint water crisis as a natural experiment. Matching vital statistics birth records with various sources of data, we employ a Difference-in-Differences (DID) approach as well as a Synthetic Control Method (SCM) to identify its causal impact on key birth outcomes. Our results suggest that the crisis modestly increased the rate of low birth weight (LBW) by 1.1-1.8 percentage points but had little effect on length of gestation or prematurity. The effects are larger for black or less educated mothers. Children born to disadvantaged mothers demonstrated 1.2-2.0 percentage points (or 10.4-17.4 percent) and 0.2- 0.6 percentage points (or 9.5-28.6 percent) rise in LBW and VLBW, respectively. We find little evidence that the Crisis increased fetal death, suggesting that the scarring effect in utero may dominate the channel of mortality selection. These results survive a rich set of placebo and falsification tests. Finally, our results lend support to three mechanisms at work linking water contamination and birth outcomes, i.e. biological effect, maternal stress, and avoidance actions.

JEL Classification: I14, I18, Q53, Q58

Keywords: water pollution, lead exposure, Flint infants, low birth weight

Corresponding author:
Xi Chen
Department of Health Policy and Management
Yale School of Public Health
Yale University
60 College Street
New Haven, CT 06520-8034
USA
E-mail: xi.chen@yale.edu

1. Introduction

Drinking polluted water has long been a top environmental concern for Americans and has presented large threats to health. According to the U.S. EPA, a fifth of the U.S. population have been exposed to unsafe water more than once during the past decade. There are nearly 3,000 recently recorded areas with peak level lead poisoning rates at least double those in Flint. More than 1,100 of these communities had a rate of elevated blood tests approximately four times higher than those in Flint (Centers for Disease Control and Prevention, 2012 & 2016; Pell and Schneyer, 2016). However, compared to a burgeoning economics literature on air pollution and child health (e.g. Currie and Neidell, 2004; Currie, Neidell, and Schmeider, 2009), the impact of contaminated drinking water on child health has received much less attention (Currie et al. 2013). More research is required to review current policies and advise further legislations for safer tap water.

Environmental and epidemiological studies on lead in the water find that maternal lead exposure is associated with worse fetal growth and birth outcomes, such as preterm birth (Zhu et al., 2010) and fetal death (Edwards, 2014), and maternal ingestion of lead crossing the placenta provides a potential pathway to lead poisoning of the fetus (Taylor, Golding, and Emond, 2015). However, little economic research establishes a causal relationship. One exception is Grossman and Slusky (2017) who find lead increases fetal death. Moreover, other pollutants in the drinking water, such as trihalomethanes (THMs) and coliform bacteria, as well as outbreaks of Legionnaire's Disease, are also associated with health at birth.[1] Since the common changes in color, taste, and odor indicate simultaneous changes in other components of the local drinking water, attributing the main adverse birth effects to lead only may result in an overestimation of the lead effect.

This paper estimates the effects of *in utero* exposure to the Crisis on health at birth using the universe of birth records for the city of Flint from 2008 to 2015. We use variations in water quality, caused by a switch of the drinking water source from Lake Huron to the Flint River, as a natural experiment to identify the effect of drinking water pollution. The city of Flint changed its main drinking water source at the end of April 2014 from Detroit-supplied water to the considerably more corrosive Flint River (see the timeline in Figure A1), resulting in major water contamination[2] and a rise in the complaints and reporting of changes in the color, taste, and odor of the water. We leverage a change in the water supply in Flint but not in the rest of Michigan or other cities in the

[1] The byproducts of chlorinated disinfectants, such as THMs, have been found associated with SGA, low birth weight, and spontaneous abortions (Gallagher et al., 1998; Waller et al., 1998).

[2] In August and September 2014, coliform bacteria were detected in the water, representing a violation of the Safe Drinking Water Act. To eliminate the bacteria, additional chlorine was added, but this caused another violation of the same Act with respect to the total level of trihalomethanes (THMs). In October 2014, Flint's General Motors plant stopped using Flint tap water due to the corrosion of engine parts caused by the high levels of chlorine. To reduce the THM levels in the water, ferric chloride was added, making the water 19 times more corrosive than the Detroit-supplied water. This increased level of corrosivity facilitated the leaching of lead into drinking water for the roughly 40% of Flint homes supplied by lead piping. In six of the nine city wards, the water in 20% to 32% of homes had levels of lead above 15 ppb, a concentration that triggers remedial action under the Safe Drinking Water Act's Lead and Copper Rule. The 90th percentile was 25 ppb, and in some homes the lead levels exceeded 1000 ppb (Bellinger 2016). Figure A2 shows elevated blood lead levels among local tested Children.

United States. Our two empirical strategies enable us to compare birth outcomes across treatment and control groups who are exposed to differing levels of drinking water contaminants while in utero. First, using a standard difference-in-differences approach, we compare newborns in Flint before and after the Crisis to those in comparison cities in Michigan that followed similar trends in fertility and birth outcomes with the exception of the Crisis. Second, we employ the synthetic control method and utilize a rich set of observables to best match Flint with a large set of candidate cities in Michigan and the rest of the U.S. (i.e. the counterfactual Flint). Our results also survive a large set of placebo tests and robustness checks.

Our results suggest modest effects of drinking water contamination on all children. Specifically, there is an increased rate of low birth weight (hereafter LBW) by 1.1-1.8 percentage points or 9.5-15.5 percent following the water change in Flint compared to other control cities. Moreover, we find larger and more significant effects on the probability of LBW and very low birth weight (hereafter VLBW) among infants born to black mothers or less educated mothers compared to those born to white, college-educated mothers. Heterogeneity also exists by the timing and duration of the gestational exposure. Larger effects on LBW and VLBW are found for the infants exposed in each trimester. There is no observable effect on sex ratios, therefore the scarring effect may dominate the mortality selection effect. We also find little evidence on length of gestational week or likelihood of prematurity, suggesting that LBW and VLBW are mainly driven by being born too small rather than too soon.

We further explore three potential mechanisms: the biological effect, maternal stress (Goodnough and Atkinson, 2016; Duncan et al., 2017) and avoidance behavior (e.g. bottled water consumption, or fertility choices). The gradual development of the Crisis enables us to partially distinguish the roles played by each of the three channels. Our mechanism tests suggest that biological effects, maternal stress, and avoidance behavior all matter and play different roles in different stages of the water contamination incident. We find no notable changes in the number of city-level births or marital status of Flint mothers after the Crisis.

Our identified effects may underestimate the biological effect of the Flint water contamination. First, our sample included all mothers living in Flint. Since we have no direct information on their blood lead level or other pollutants during pregnancy, our estimated effect should be interpreted as the intent-to-treat effect, as not every mother in Flint was actually exposed to contaminated water. Second, our datasets have no direct information on avoidance actions taken by local residents. Therefore, not accounting for avoidance behavior may result in the identified effects being a mixture of biological and avoidance effects.[3] Third, the latent health effect of the water switch may manifest later in life, such as being in poor health, low educational attainment, poor labor market

[3] For example, if women avoided pregnancy due to their concern over water pollution, the affected infants in our sample would be fewer than the potentially affected infants if there were no pregnancy avoidance. Similarly, if pregnant women drink bottle water instead of tap water, the effect we found would be smaller than the effect without avoidance behaviors.

performance in adulthood, increased behavioral problems and criminal activities (see e.g., Almond, Currie, and Duque, 2017; Aizer and Currie, 2017).[4]

Our study distinguishes from Grossman and Slusky (2017) in several aspects. First, we identify the effect for survived children, specifically their LBW and preterm, while Grossman and Slusky (2017) focus on those being selected out via reduced fertility or fetal mortality. Second, we examine the overall effect as well as heterogeneous effects by maternal socioeconomic status (hereafter SES) as maternal disadvantage may worsen health inequality at birth. Third, we construct synthetic Flint using a much larger set of candidate cities across the U.S. to fit Flint with the synthetic Flint. Fourth, we attempt to exploit various channels through which the Crisis may affect birth outcomes, including the biological effect of maternal exposure to lead that Grossman and Slusky (2017) explore in their work.

Our paper adds to the broad literature testing fetal origin hypothesis and provides new evidence on inequality of health at birth. We make contributions to key methodological issues relevant to the study of a broad range of fetal and infant health effects. First, women who are exposed to pollutants differ in observable ways from those who are not, and they may also differ in unobservable ways. These differences must be accounted for, or they will bias the estimated effects of exposure. We solve this important issue via two state-of-the-art designs with appropriate assumptions to form a counterfactual Flint as close as possible to the real Flint in absence of the Crisis. Third, among studies on the relationship between the Crisis and birth outcomes, we exclude the cities with health-based violations to the Safe Drinking Water Act (SDWA) to construct a cleaner comparison group for impact evaluation. Our findings may also help formulate public policy to improve safe water supply, or more generally public good provision, key to child development and narrowing health inequality at birth. In the past decades, accumulating knowledge about determinants of infant health and how to apply this knowledge to policy practice that protects infants have generated tremendous benefits (Aizer and Currie, 2014). Therefore, the widening wealth gap in the parents' generation has not transmitted to widening the gap of child health. Rather, we have observed significant convergence of health at birth or during the early stage of life, which are attributable to effective public programs and public goods provision.

The paper proceeds as follows. The next section lays out our empirical strategies. Section 3 presents the data. Section 4 reports the results. Section 5 concludes.

2. Empirical Strategy

In this section, we use both a difference-in-differences (DID) method and a synthetic control method (SCM), to identify the effects of the Crisis on birth outcomes.

[4] While Michigan expanded Medicaid over this time period, we find no evidence of increase in the proportion of births covered by Medicaid. The use of prenatal care during this time period decreases. Therefore, Medicaid coverage is unlikely to attenuate our results.

2.1. Difference-in-Differences (DID)

The DID method is applied to a sample of newborns, comparing the probability of LBW between infants in Flint and those in control cities before and after the Crisis. The main specification is

$$y_{ict} = \beta_0 + \gamma * I(F)_i * T_{it} + \beta_1 * I(F)_i + \beta_2 * T_{it} + \beta_3 * X_{ict} + city_c + \vartheta_t + \varphi_t + \omega_{ct} + \varepsilon_{ict} \quad (1)$$

where y_{ict} denotes birth outcomes for infant i born in city c at time t. y_{ict} represents birth weight, LBW, VLBW, gestation length, and prematurity. $I(F)_i$ is a binary variable, equal to 1 for Flint infants and 0 otherwise. T_{it} is a time dummy taking the value of 1 if the infant i was born after the Crisis (April of 2014)[5], and 0 otherwise. γ is the variable of interest, the DID estimator of the effect of exposure to polluted water *in utero*. X_{ict} is a vector of covariates, including maternal characteristics and child gender. The main specification also includes city fixed effects ($city_c$), year fixed effects (φ_t), and month fixed effects (ϑ_t) to adjust for city heterogeneity and capture time trend. ω_{ct} further adjusts for city-specific linear time trend. Finally, ε_{ict} represents an unobserved disturbance, clustered at the city level.[6] When the outcomes are binary variables to indicate LBW, VLBW, or prematurity, equation (1) estimates linear probability models, and the DID estimator (γ) is interpreted as the percentage point change in the probability of LBW, VLBW, or premature births.

The identification assumption underlying Equation (1) is that the time trends in birth outcomes should be similar between the Flint infants and control city infants in the absence of the Crisis. Therefore, it is crucial to find a comparable control group of cities holding parallel trends in the birth outcomes with Flint before the Crisis. Intuitively, Michigan cities would be the ideal comparison cities. They have similar demographics, economy, history, and weather with Flint. Some of them even used the same water supply as Flint before the Crisis. Figure A3 shows similar pre-trends in the average birth weight and the percentage of LBW births between the control cities and Flint, though the trend for the group of control cities is relatively smoother after taking average. We also formally check the pre-treatment trends by slightly modifying equation (1) using a family of intersections of month dummies and $I(F)_i$ instead of $I(F)_i * T_{it}$. We cannot reject the hypotheses that the pre-treatment trend for all of our interested birth outcomes are the same between control cities and Flint at conventional levels of statistical significance.

2.2. Synthetic Control Methods (SCM)

The DID approach may introduce substantial ambiguity about how comparison groups are chosen since researchers select comparison groups based on subjective measures of affinity between the treated group and the

[5] Flint switched to use Flint River water on April 21st of 2014. We consider May of 2014 as the first treatment month.

[6] Since there are only 7 clusters in the DID setting, we use a wild cluster bootstrap proposed by Cameron, Gelbach, and Miller (2008) to adjust for standard errors.

untreated group (Abadie et al., 2010). Also, compared to DID, SCM can substantially reduce potential endogeneity problems caused by time varying unobservables and avoid the over-rejection bias of the standard t-test (Abadie et al., 2010; Adhikari et al., 2016). To confirm that the DID results are not an artifact of the choice of control cities and to ensure that the estimates are not biased by unobserved factors, we further use the SCM to find the "best" counterfactual version of Flint (referred to as "synthetic Flint") from untreated cities.

By assigning different weights to maternal characteristics, average number of birth, and the average pre-treatment birth outcomes in some pre-treatment periods of potential comparison cities, the SCM constructs a synthetic Flint that optimally resembles the characteristics of Flint in the pre-treatment period and simulates what birth outcomes Flint infants would have if they had not undergone the Crisis. To avoid potential downward bias, cities in "donor pool" are limited to those that did not have any health-based drinking water violations of SDWA during the study period.[7] Following Abadie et al. (2003, 2010) and McClelland and Gault (2017), we conduct a series of placebo tests by iteratively applying the SCM to every other city in the donor pool. According to the distribution of the ratios of post/pre-crisis Root Mean Square Prediction Error (post/pre RMSPE) for Flint and cities in the donor pool generated by placebo tests, we calculate a permutation p-value to assess the significance of the difference between Flint and Synthetic Flint. See Appendix B for a detailed description of SCM.

3. Data

This study relies on three sources of data: the Centers for Disease Control and Prevention (CDC) vital statistics natality records (birth certificates), the American Community Survey, and drinking water violations across cities in the United States from 2008 to 2015.[8] Birth certificates include a record for every birth comprised of health at birth, gestational age, and maternal characteristics (including race, education, age, prenatal care, and health). Information on the mother's resident city recorded on the birth certificates enables us to match with other datasets. Our analysis focuses on singleton birth records.[9] We obtain information on city-level population size from the American Community Survey.

The Safe Drinking Water Information system (SDWIS) of the Environmental Protection Agency (EPA) reports all types of drinking water violations for both active and inactive water systems across the U.S.[10] Violations are categorized into two types: health-based violations, when water is not treated properly or when the level of contaminants exceeds safety standards; non-health-based violations, when failing to adhere to mandatory guidelines such as collecting samples or reporting requirements. We carefully choose control cities without

[7] We observed over 150 major cities across the U.S. with health-based drinking water violations of SDWA from 2008 to 2015.
[8] Flint switched back to Detroit water in October 2015, symbolizing government intervened the crisis. To exclude the effects of various government actions, we drop data after September 2015.
[9] Twin births take up about 4% of the sample.
[10] For cities in which SDWIS does not provide related water system information, we first find all the water systems in their respective counties from SDWIS, and then identify the matched water system by via Google search.

health-based drinking water violations between 2008 and 2015 to eliminate potential biases. Births without geographic identifiers are dropped from our sample, as we could not match their residences to the water systems and water quality data.[11]

Our DID design chooses surrounding control cities from Michigan as they are probably the most comparable to Flint. There are 8 cities (including Flint) in Michigan with a population of more than 100,000 in 1990. The city Livonia was further dropped due to the absence of birth records in its post treatment period. The final sample includes 7 cities with 167,757 birth records (see in Table A1 comparisons between Flint and control cities pre- and post- the Crisis).[12] Our treatment group includes all Flint singleton births from January of 2008 to September of 2015, with 12,788 and 2,637 births in the pre- and post- treatment periods, respectively. Given that 40% of homes in Flint installed lead pipes, our sample includes both mothers living in areas with lead pipes and those with no lead pipes. This sample allows us to evaluate the combined effect of the Crisis on birth outcomes, including biological effect through direct exposure and indirect channels such as psychological shocks or microorganisms. The geographic information on living areas enables us to partially distinguish the effect of water contamination from that of stress. In our SCM analysis, we aggregate the individual birth data to the city / half-year level. Figure 1 maps the locations of Flint, 6 control cities in the DID analysis, and 162 cities in the donor pool of the SCM analysis. Table 1 presents summary statistics of the observable characteristics. Compared to other cities, Flint has worse birth outcomes including lower birth weight, shorter length of gestation, and higher proportion of infants born with LBW and prematurity. With respect to mothers' characteristics, Flint has fewer of the following: mothers older than 35, white mothers, and mothers with a bachelor's degree or higher. Moreover, Flint has more teenage mothers, black mothers, less educated mothers, single mothers, diabetic mothers, mothers with chronic hypertension, and mothers with pregnancy-associated hypertension.

4. Results

4.1. Main Results

Table A1 compares summary statistics between Flint and control cities in DID analysis before and after the Crisis. Column 7 suggests that Flint infants were more likely to be of LBW than infants in the control cities after the Crisis. Potential confounders include educational attainment and health status of mothers, and prenatal care in Flint. The significant changes in these factors alert us to consider mother's unobserved characteristics. We check the correlation between the Crisis and mother's characteristics by regressing maternal characteristics on the indicator of the Crisis, city, year and month fixed effects both with and without linear city-specific trends.

[11] Births without geographic identifiers are either from the small cities or from towns, which in average have better birth outcomes and demographics than those in big cities.

[12] The 7 cities are Ann Arbor, Detroit, Flint, Grand Rapids, Lansing, Sterling Heights, and Warren.

Table A2 shows no association between the Crisis and mother's characteristics once we control for the city-specific trends, underpinning the appropriateness of our model.

Table 2 presents the main results of the Crisis on birth outcomes. The left and right panels show results using the DID and the SCM methods, respectively. Column 1 displays the most parsimonious specification without any covariates, and the Crisis appeared to only affect length of gestation and prematurity. Controlling for mother and infant characteristics, Column 2 and Column3 both show that the Crisis significantly reduced birth weight and increased the rate of LBW, VLBW, and prematurity. However, all the effects become insignificant when we further include city-specific linear trends in Column 4, except the smaller and marginally significant effect on LBW. Given 11.4 percent LBW for Flint infants before the Crisis and the estimate in Column 4, our findings point to a 9.64 percent increase in LBW.

Panel B in Table 2 reports the results estimated by SCM, which are in line with column (4) in Panel A. The effect on the probability of LBW is only marginally significant and similar in size. As the SCM allows for time varying unobservable, the consistent findings suggest that our DID model with city-specific trends is appropriate. Our specification complements Grossman and Slusky (2017) and Abouk and Adams (2017) by further adjusting for unobserved time trend. Without controlling for city-specific trends, our results are similar to Abouk and Adams (2017), who find 71 grams reduction in birth weight and a 26% increase in LBW among white mothers. However, like Grossman and Slusky (2017), our identified effect on birth weight declines with the trend.

The visualized results for Panel B in Table 2 are in Figure 2, which displays birth outcomes in Flint and synthetic Flint (generated from a donor pool of 162 cities in the U.S.) from 2008 to 2015. Flint and the synthetic Flint in Figure 2a had a similar trend in average birth weight prior to the Crisis but diverged afterwards. Average birth weight in Flint sharply declined within the first year after the Crisis, but bounced back at the end of 2015. Figure 2b shows the synthetic Flint tracked the trajectory of LBW rate in Flint very well before the Crisis but diverged afterwards, with the former being around 11.8% and the latter increasing to 13%, suggesting that the Crisis led to a higher rate of LBW. In Figure 2c, 2d and Figure 2e, the trends for probability of VLBW, length of gestation, and prematurity showed less observable difference between Flint and the synthetic Flint after the Crisis.

4.2.Placebo Tests

To alleviate the concern that the effect on rate of LBW may be confounded by other factors, we conduct several placebo tests using the DID. First, columns 1-5 of Table 3 use the specification in column (4) of Table 2 with pre-treatment data only, assigning falsified cut-offs, i.e. May 2009, May 2010, May 2011, May 2012, and May 2013, as the starting months of the Crisis, respectively. Table 3 shows consistently insignificant coefficients, validating no effect of placebo shocks. This test also reassures us the parallel trend between Flint and control cities before the Crisis. Second, a set of placebo tests in Table 4 check if any control cities may drive the estimated effects by chance. We separately assign each control city as a pseudo-treatment city (i.e. Ann Arbor, Detroit, Grand Rapids, Lansing, Sterling Heights and Warren) and the remaining cities (including Flint) as pseudo-controls. None of the regressions show significant and adverse effect of exposure to the Crisis on the probability of LBW, further reassuring us that our results should be causal.

Next, we present placebo tests using the SCM. Figure 3 draws 162 solid gray lines representing the difference in the probability of LBW between each city in the donor pool and its respective synthetic version. The black line denotes the effect when Flint is assigned to be the treatment city. The dashed red vertical line denotes the start of the Crisis. Figure 3a shows that the effect size for Flint is larger than most of the placebo treatment effects after the Crisis. As Panel A contains cities that are poor fits prior to the Crisis, in Panel B we only keep better matched cities (i.e. those with pre RMSPE less than twice of the Flint's pre RMSPE). Without poorly fitted cities, the Flint line is saliently on the top of all other gray lines in the post-treatment period. Checking into the distribution of the ratio of post/pre RMSPE in the placebo tests, 154 out of 162 placebo cities have smaller ratios, indicating that 94.5 percent placebo treatment effects are smaller in size than the real treatment effect.

The consistent results using DID and SCM suggest that the Crisis significantly increased the probability of LBW, and this increase was not correlated with prematurity. This finding suggests that the Crisis mainly affected infants around the LBW cut-off than across the whole distribution. Figure A4 provides visual evidence for this finding. Compared to control cities in the DID and cities in the donor pool in the SCM, the distribution of birth weight in Flint shifted to the left after the Crisis, especially around the LBW cut-off, suggesting a higher density of infants with LBW.

4.3.Heterogeneous Effects

The full sample analysis may hide potentially large effect for more vulnerable population. Inequality of health at birth can be enlarged by environmental exposures. Disadvantaged mothers may have poorer health behavior, less access to medical care, and are more likely to be exposed to harmful environments (Aizer and Currie, 2014). For example, Figure A5 maps the residential segregation and the lead service line connections in Flint. Most black mothers and other disadvantaged mothers live in the areas served by lead pipes.[13] In contrast, advantaged mothers are more likely to live far away from the area with lead polluted water. Meanwhile, a higher proportion of children with elevated blood lead levels live in SES disadvantaged neighborhoods in Flint (Hanna-Attisha et al. 2016).

Since the natality data do not include maternal income, we rely on maternal race and education as a proxy for maternal SES. Disadvantaged mothers are defined as black or non-college educated mothers, while advantage

[13] Most of non-white people live in the west and north of Flint, and downtown Flint. These areas are all in the zones with lead service line connections. The non-white population here mainly refers to African Americans, who represent 95% population of non-white in Flint.

mothers are all non-black mothers with college degrees. Table A3 shows differences between disadvantaged and advantaged mothers. It is notable that marriage rate of advantaged mothers is four times that of disadvantaged mothers. Children born to both groups of mothers demonstrated worse birth outcomes after the Crisis. Although advantaged mothers generally have better birth outcomes than disadvantaged mothers, the declines in birth outcomes are more salient for advantaged mothers than for disadvantaged mothers. Therefore, it is important to adjust for confounders, such as changes in the portfolio of mothers who gave birth after the Crisis.

Table 5 shows the respective estimates for disadvantaged mothers and advantaged mothers. Panel A shows results using the specifications in columns (3) and (4) of Table 2. Without controlling for city-specific trends in columns (1) and (3), birth weight decreased and the probability of prematurity increased for both disadvantaged and advantaged mothers. However, these effects disappeared after accounting for city-specific trends in columns (2) and (4). In contrast, the effects on the rates of LBW and VLBW were consistently significant for disadvantaged mothers. The rate of LBW increased by 1.2 percentage point (or 10.43 percent) and the rate of VLBW increased by 0.2 percentage point (or 9.52 percent) among Flint infants born to disadvantaged mothers. There is no significant effect on LBW or VLBW for advantaged mothers.

Panel B in Table 5 shows that results using the SCM are largely consistent with those using the DID, suggesting that heterogeneous effects did exist between disadvantaged and advantaged mothers. Specifically, there is no significant effect on any birth outcomes for advantaged mothers. LBW and VLBW for disadvantaged mothers increased by 2.0 and 0.6 percentage points, respectively. On the other hand, the results are somewhat different between the two approaches: For birth weight and length of gestation, the SCM estimates show significant effects for disadvantaged mothers, while the DID estimates are insignificant. The effect on the length of gestation (0.003 week) is economically insignificant, considering average length of gestation being 38.2 weeks.

Figure 4 and Figure A6 display the results separately for disadvantaged and advantaged mothers estimated by the SCM. In Figure 4, the synthetic Flint tracks the trajectory of Flint very closely before the Crisis for each birth outcome but deviated afterwards. Specifically, birth weight and length of gestation among Flint newborns declined right after the Crisis but returned to the level of synthetic Flint at the end of 2015. The rates of LBW, VLBW and prematurity in Flint increased sharply right after the Crisis, while the trends for such birth outcomes in the synthetic Flint were relatively smooth. Note that the effects are more salient for LBW and VLBW than prematurity. In general, smaller effects are found for advantaged mothers in Figure A6.

In the placebo tests, Figure A7 displays real gaps in the rate of LBW and placebo gaps. Figure A7a shows that the estimated gap for disadvantaged mothers after the Crisis is consistently larger than the gaps for the cities in the donor pool. The probability of obtaining a post/pre RMSPE ratio as large as Flint's is 5/163=0.031, indicating its significance at the 5 percent level. In contrast, Figure A7b shows that the gap for advantaged mothers was not enlarged right after the Crisis. While it did increase dramatically in part of the post treatment period, the gap approached zero afterwards. Figure A8 displays similar pattern in the gaps for the rate of VLBW.

Finally, we re-estimate Equation 1 separately for infants exposed to different time windows in utero. As shown in columns 1 and 2 in Table 6, the Crisis did not affect birth outcomes if the exposure only overlapped in the 3rd trimester. Again, there was no significant effect on birth outcomes for infants exposed in the second/third trimesters. However, the Crisis affected birth outcomes, especially the rates of LBW and VLBW, for those exposed in all three trimesters. Column 6 shows the rates of LBW and VLBW increased by 1.1 and 0.4 percentage points, respectively, slightly larger than those for the full sample.

4.4. Further Discussion

One concern is that the Crisis may cause mothers to move out of Flint for safe water, thus the Flint infants we observe were born to mothers who were unable to move out. Our results may therefore be overestimated if the mothers staying in Flint tended to be less educated and poorer, or underestimated if mothers who stayed tend to be more educated and richer. We address this concern by checking the changes in the portfolios of all local residents as well as females aged 15 to 49 in Flint. Figure A9 shows Flint and 6 control Michigan cities shared the same trend in total population and female population aged 15-49 before the Crisis. Flint experienced an increase in females aged 15-49 in 2014 followed by an immediate decline in 2015, while control cities had a relatively smoother trend. We formally check the relationship between the Crisis and the size of females aged 15-49 using the specification in Equation (1). Table A4 shows that the Crisis did not significantly reduce females aged 15-49 in Flint, suggesting that migration for those in child-bearing age should not be a concern.

Another concern is that if the Crisis results in miscarriage or fetal death (via mortality selection), we may underestimate the negative health effects because our sample would be made up of relatively healthier infants. Given that the CDC has not yet released the direct fetal death data after 2015 when we conduct this study, and it is difficult to record the incidence of miscarriage, we follow other studies such as Sanders and Stoecker (2015) to use male-to-female ratio at birth to indirectly test fetal death. The premise of using male-to-female ratio is that males *in utero* are often more vulnerable to negative health shocks than females. As such, the Crisis tends to result in more male fetal losses, and therefore reduces the probability of male births and the male-to-female ratio. We use both DID and SCM methods to estimate the effect on sex ratios at birth for the full sample, infants born to disadvantaged mothers, and infants born to advantaged mothers, respectively. Panel A of Table A5 shows no significant effect using the DID. Panel B of Table A5 using the SCM finds no significant decline in sex ratios. Figure 5 displays the visualized results. These tests mitigate the concern that the Crisis might be strong enough to downward bias our estimates through mortality selection. Overall, our findings point to the scarring effect as the driving force of worse birth outcomes we observe.

We conduct event studies to disentangle three potential mechanisms through which the Crisis could affect birth outcomes, including biological effects, maternal stress, and avoidance behavior. Figure 6 shows that the rate of LBW raised among Flint infants after the Crisis, especially for those born to disadvantaged mothers. As the advantaged mothers were more likely to live away from the lead pipes and take avoidance actions, it is possible that they were less stressed about water pollution. In contrast, disadvantaged mothers were more likely to live in areas with lead pollution and feel stressful when they were aware of the problem and less likely to avoid pollution.

Figure 6d focuses on the post-Crisis period in Figure 6b to show dynamic effects on LBW for disadvantaged mothers. We implement separate DID analyses to subsamples of infants born during each of the three consecutive months after the Crisis (i.e. from May of 2014 to September of 2015). Figure 6d plots the estimates in the post-treatment periods. There was little effect in the two months immediately following the Crisis, probably because the public was largely unaware of the water issue and therefore had low stress (as suggested by the Google search data in Figure A11). However, from July 2014 to September 2014, the impact on the rate of LBW significantly increased. This cumulative effect is likely attributable to the biological effect. The impact declined from September 2014 to January 2015, which may be explained by the sharply rising bottled water consumption (as suggested by its local sales data in Figure A10) that enabled at least some local residents to avoid exposure and therefore partially offset the biological effect. The green vertical line marks the starting point when the public became aware of the Crisis. From January 2015 to April 2015, the impact of the Crisis had an upward trend, though not precisely identified. One possible explanation is that public stress intensified during this period after learning about the hazardous nature of water pollution, which was indicated by the sharply rising Google search volume (Figure A11). The declining effect from April 2015 to September 2015 may correspond again to the sharply increased avoidance behavior (Figure A10) and lower stress (Figure A11) during this period.

5. Conclusions

Matching vital statistics on the universe of all birth records in Flint with various sources of data, we provide one of the few causal evidences on the short-term impact of drinking water contamination on birth outcomes. Our main results suggest that contaminated water modestly increased the rate of LBW by 1.1-1.8%, reduced birth weight by 6-32 grams, but had little effect on length of gestation or prematurity. Effects are larger for black and low educated mothers. After the Crisis, these disadvantaged mothers demonstrated 1.2-2.0 and 0.2-0.6 percentage points increase in LBW and VLBW, respectively. We find some evidence that longer exposure *in utero* corresponds to larger effects, while there is no evidence that the Crisis increased fetal death, suggesting that the scarring effect *in utero* dominates.

Future studies may improve in a few aspects. First, to better understand the physiological mechanisms, they may distinguish the effect due to a surge in lead from simultaneous changes in other pollutants. To achieve this, high quality data on the exact levels of multiple dominant pollutants, preferably in the blood, is necessary; Second, it is worthwhile to monitor longer term impact of the Crisis on health and human development; Third, to better understand potential psychological mechanisms, data that measure maternal stress during this Crisis could be helpful; Fourth, detailed avoidance behavior data is useful to distinguish biological effects from avoidance actions.

Funding Acknowledgement

Financial support from the James Tobin Research Fund at Yale Economics Department, Yale Macmillan Center faculty research award (2017-2019), the U.S. PEPPER Center Scholar Award (P30AG021342), National Science Foundation in China (71602149) and two NIH/NIA grants (K01AG053408; R03AG048920) are acknowledged. The views expressed herein and any remaining errors are the authors' and do not represent any official agency. None of the authors have potential conflicts of interests that could bias this work.

References

Abouk, R. and Adams, S. (2017). Bans on electronic cigarette sales to minors and smoking among high school students. Journal of Health Economics, 54, 17-24.

Abadie, A., Diamond, A. and Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490): 493-505.

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59(2):495–510.

Abadie, A. and Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. The *American Economic Review*, 93(1):113–132.

Adhikari, B., R. Duval, B. Hu, and P. Loungani. (2016). Can Reform Waves Turn the Tide? Some Case Studies Using the Synthetic Control Method. International Monetary Fund working paper.

Ahmad, S. A., Barua, S., Khan, M. H., Faruquee, M. H., Jalil, A., and Hadi, S. A., et al. (2001). Arsenic in Drinking Water and Pregnancy Outcomes. *Environmental Health Perspectives*, 109(6): 629-631.

Aizer, A., and Currie, J. (2017). Lead and juvenile delinquency: new evidence from linked birth, school and juvenile detention records. Nber Working Papers.

Aizer, A. and Currie, J. (2014). The intergenerational transmission of inequality: maternal disadvantage and health at birth. *Science*. 344(6186): 856-61.

Almond, D., J. Currie, and V. Duque. (2017). Childhood Circumstances and Adult Outcomes: Act II. *NBER Working Paper No. 23017*.

Bellinger, D. (2016). Lead Contamination in Flint — An Abject Failure to Protect Public Health. New *England Journal of Medicine*, 374:1101-1103.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3), 414-427.

Centers for Disease Control and Prevention. (2012). Lead in Drinking Water and Human Blood Lead Levels in the United States. *Morbidity and Mortality Weekly Report*, August 10.

Centers for Disease Control and Prevention (2016). Investigation: Blood Lead Levels Higher after Switch to Flint River water https://www.cdc.gov/media/releases/2016/p0624-water-lead.html

Christensen, P., David A. Keiser, and Gabriel D. Lade. (2017). The Economic Effects of Environmental Crises: Evidence from Flint, Michigan. NBER Eggtime, 2017.

Currie, Janet and Matthew Neidell. (2004). Air Pollution and Infant Health: What Can We Learn From California's Recent Experience? *Quarterly Journal of Economics,* 120 (3):1003-1030.

Currie, Janet, Matthew Neidell, and Johannes F. Schmieder. (2009). Air Polution and Infant Health: Lessons from New Jersey. *Journal of Health Economics*, 28:688-703.

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):791–810.

Duncan, Brian, Hani Mansour, and Daniel I. Rees. (2017). It's just a Game: the Super Bowl and Low Birth Weight. *Journal of Human Resource,* Vol.52, No.4 :946-978.

Edwards, M. (2014). Fetal Death and Reduced Birth Rates Associated with Exposure to Lead-Contaminated Drinking Water. *Environ. Sci. Technol.*, 2014, 48 (1): 739–746.

Fung, A., Mary Graham, and David Well. (2007). Full Disclosure: the Perils and Promise of Transparency. Cambridge University Press.

Gallagher, M. D., Nuckols, J. R., Stallones, L., and Savitz, D. A. (1998). Exposure to trihalomethanes and adverse pregnancy outcomes*. Epidemiology*, 9(5): 484-489.

Goodnough, A. and Atkinson, S. A potent side effect of the Flint water crisis: mental health problems. The New York Times. (2016). http://www.nytimes.com/2016/05/01/us/flint-michigan-water-crisis-mentalhealth.html. Accessed 16 March 2018.

Grossman, D. and Slusky, D. (2017). The Effect of an Increase in Lead in the Water System on Fertility and Birth Outcomes: The Case of Flint, Michigan, No 201703, WORKING PAPERS SERIES IN THEORETICAL AND APPLIED ECONOMICS, University of Kansas, Department of Economics.

Hanna-Attisha, M., Lachance, J., Sadler, R. C., and Champney Schnepp, A. (2016). Elevated Blood Lead Levels in Children Associated with the Flint Drinking Water Crisis: A Spatial Analysis of Risk and Public Health Response. *American Journal of Public Health*, 106(2), 283-290.

McClelland, R. and G. Sarah. (2017). The Synthetic Control Method as a Tool to Understand State Policy. *State and Local Finance Initiative*, Urban Institute.

Pell, M.B. and J. Schneyer. (2016). The thousands of U.S. locales where lead poisoning is worse than in Flint. *The Reuters*. December 19.

Sanders, H, J., and Charles Stoecker. (2015). Where Have All the Young Men Gone? Using Sex Ratios to Measure Fetal Death Rates. *Journal of Health Economics*, 41:30-45.

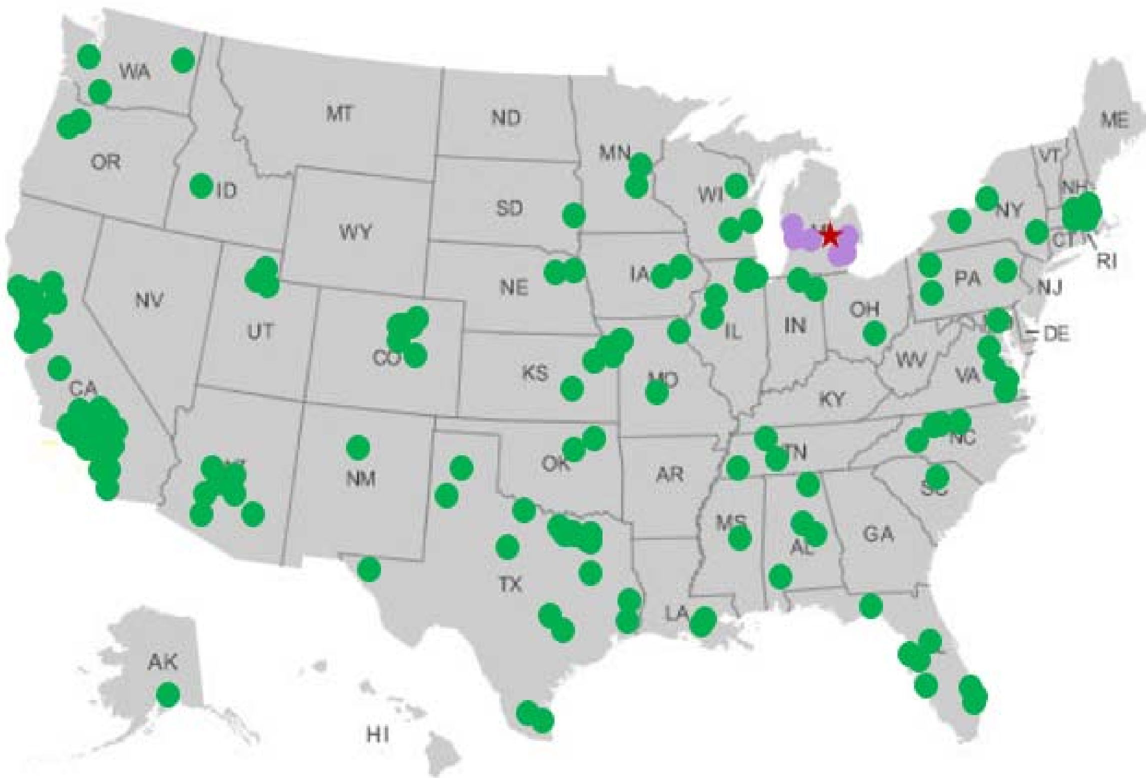
Taylor, CM, J Golding, and AM Emond. (2015). Adverse Effects of Maternal Lead Levels on Birth Outcomes in the ALSPAC Study: A Prospective Birth Cohort Study. *BJOG* 122(3): 322-328.

United States Environmental Protection Agency. (2015). High Lead Levels in Flint, Michigan - Interim Report. http://flintwaterstudy.org/wp-content/uploads/2015/11/Miguels-Memo.pdf Retrieved on Nov.10, 2016.

Waller K, Swan SH, DeLorenze G, and Hopkins B. (1998). Trihalomethanes in Drinking Water and Spontaneous Abortion. *Epidemiolog*y, 9:134-140.

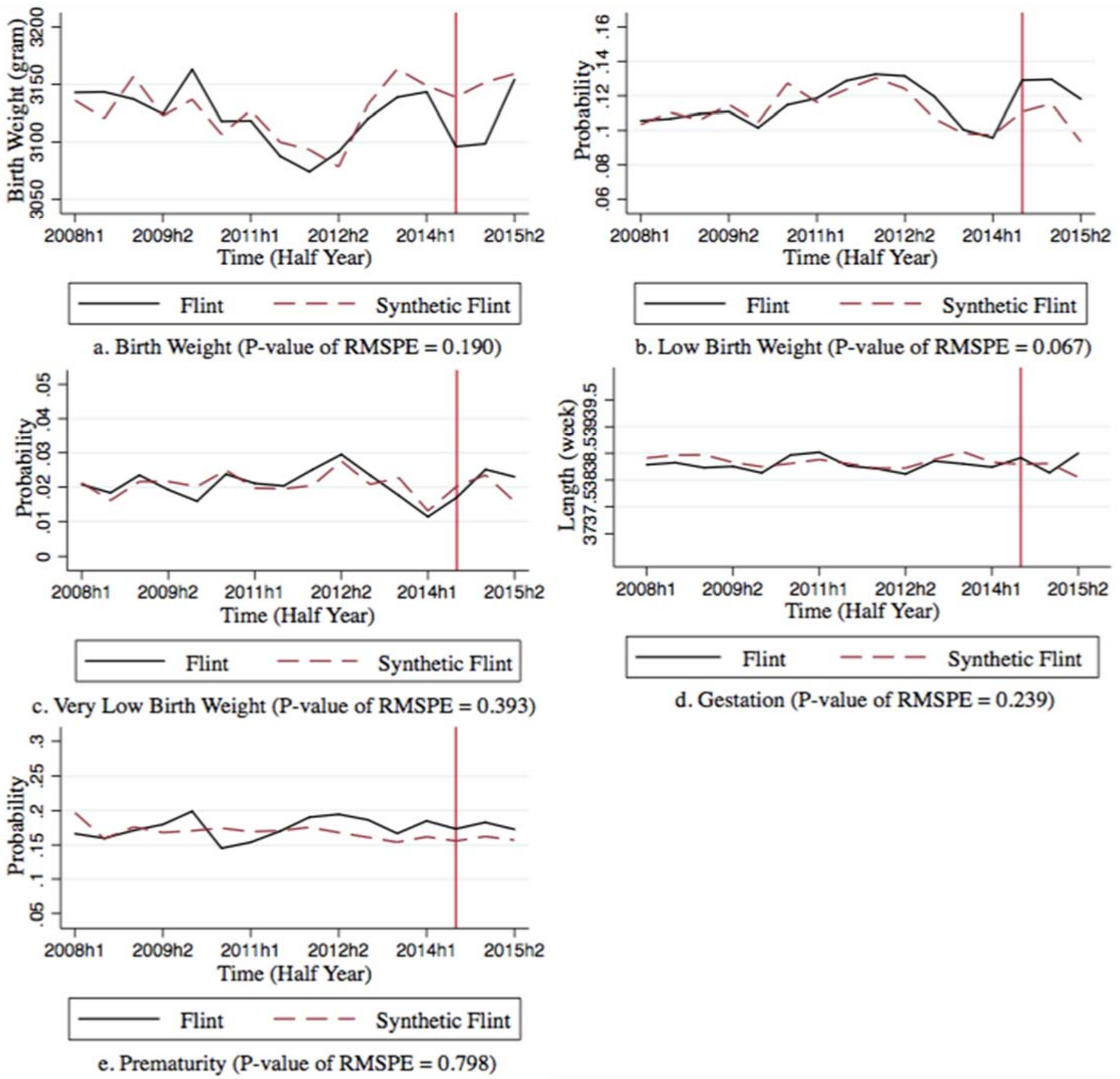
Zhu, M., Fitzgerald, E.F., Gelberg, K.H., Lin, S., and Druschel, C.M. (2010). Maternal Low level Lead Exposure and Fetal Growth. *Environ. Health Perspective*,118: 1471–1475.

Figure 1: Spatial Distribution of Selected Cities in SCM



Note: Flint is a red star, 6 Michigan cities used in DID are in purple, and 162 cities in donor pool are in green.

Figure 2: Trends in Birth Outcomes: Flint vs. Synthetic Flint



a. Birth Weight (P-value of RMSPE = 0.190)

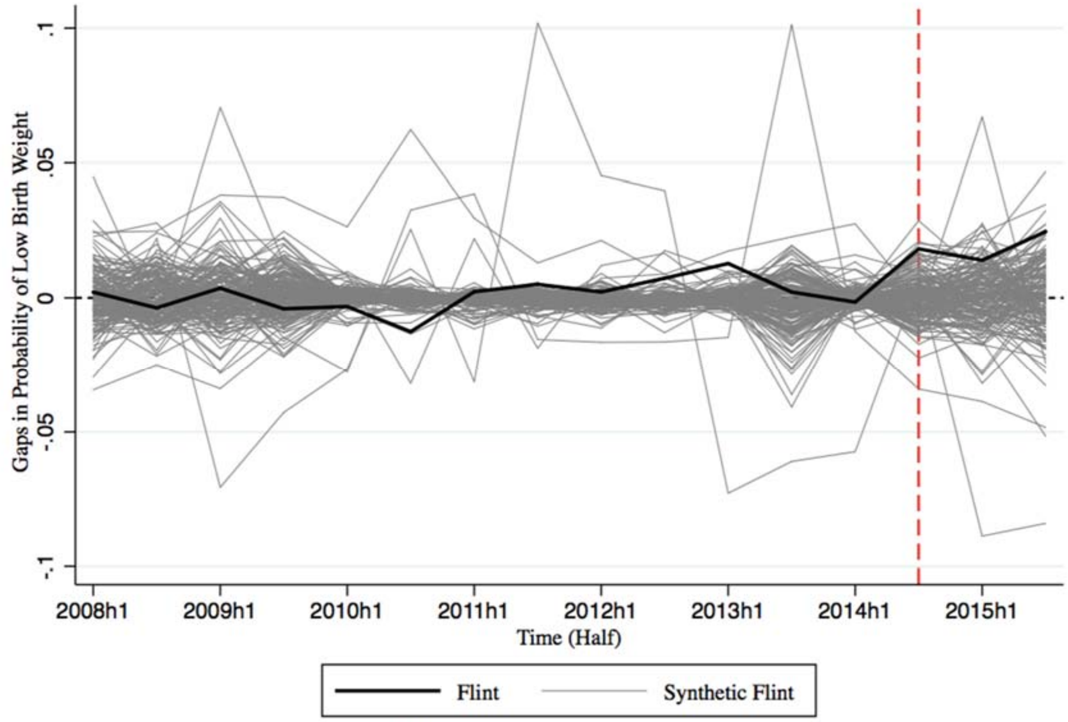
b. Low Birth Weight (P-value of RMSPE = 0.067)

c. Very Low Birth Weight (P-value of RMSPE = 0.393)

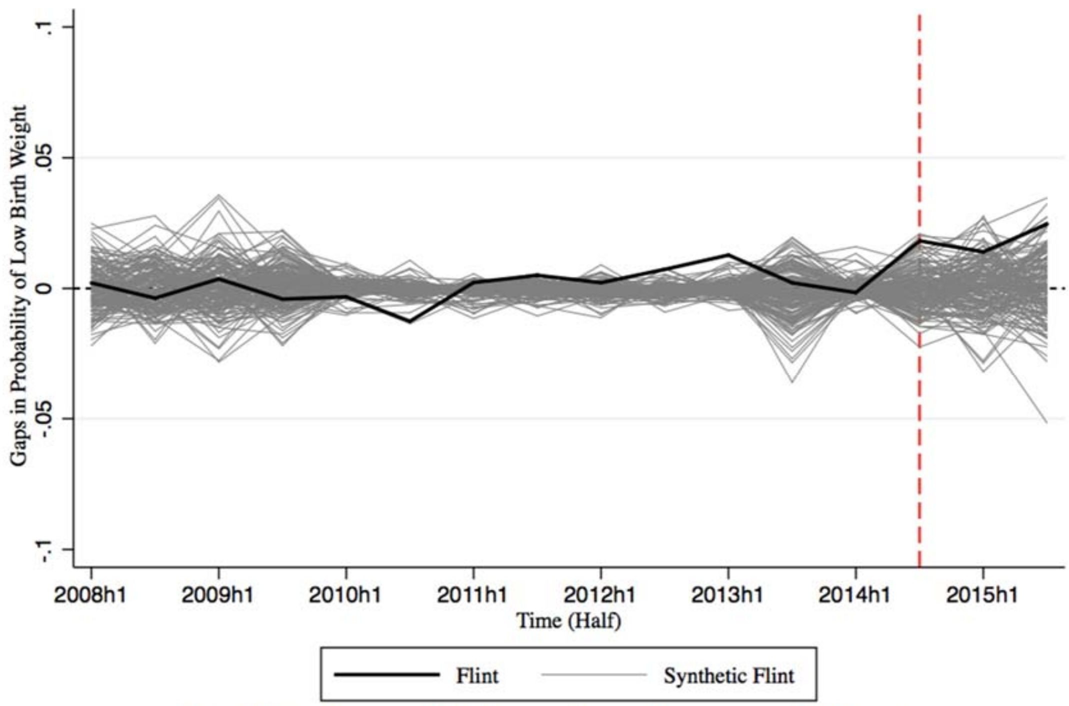
d. Gestation (P-value of RMSPE = 0.239)

e. Prematurity (P-value of RMSPE = 0.798)

Figure 3: Placebo Gaps for the Probability of LBW

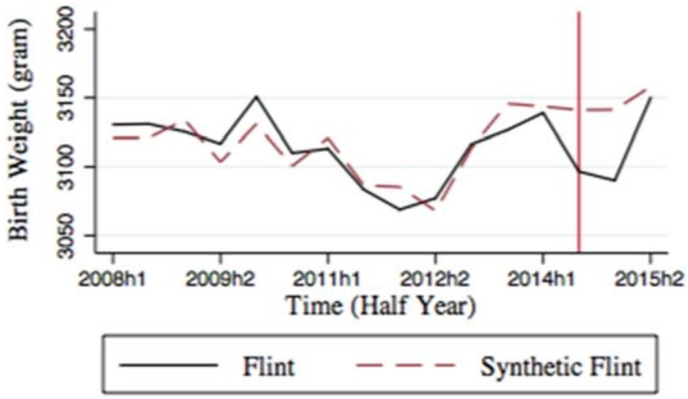


a. Probability of Low Birth Weight Gaps in Flint and Placebo Gaps in All 162 Control Cities

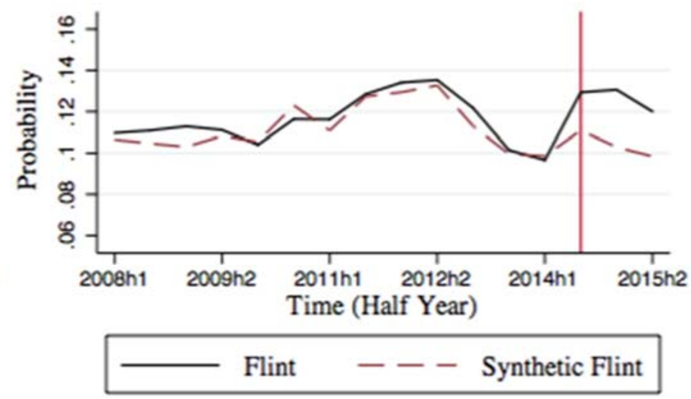


b. Probability of Low Birth Weight Gaps in Flint and Placebo Gaps in 154 Control Cities
(discarding cities with pre-RMSPE more than two times as high as Flint's)

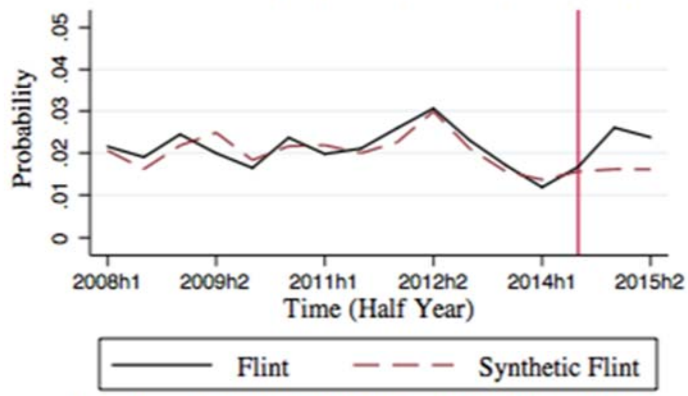
Figure 4: Trends in Birth Outcomes for Disadvantaged Mothers



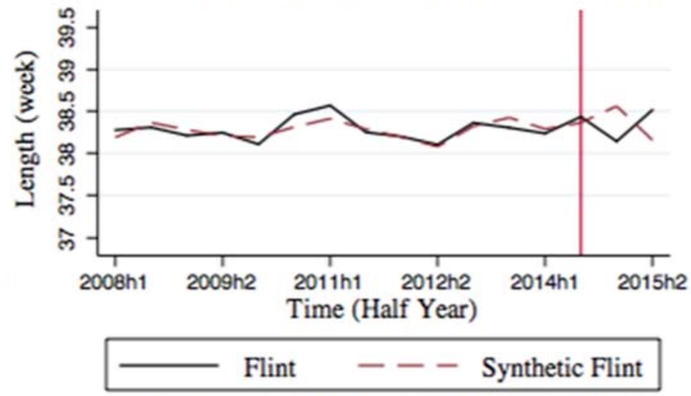
a. Birth Weight (P-value of RMSPE = 0.098)



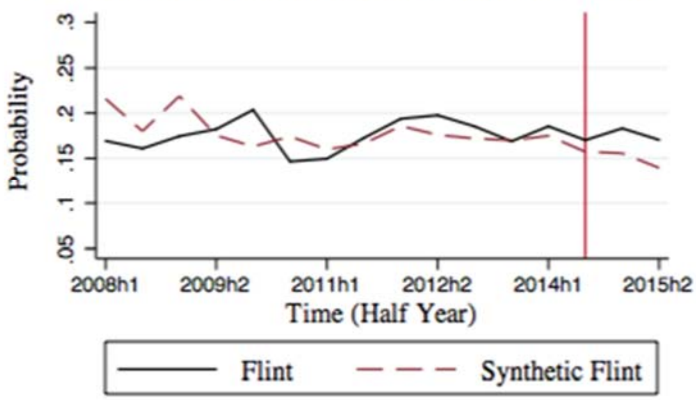
b. Low Birth Weight (P-value of RMSPE = 0.031)



c. Very Low Birth Weight(P-value of RMSPE = 0.092)

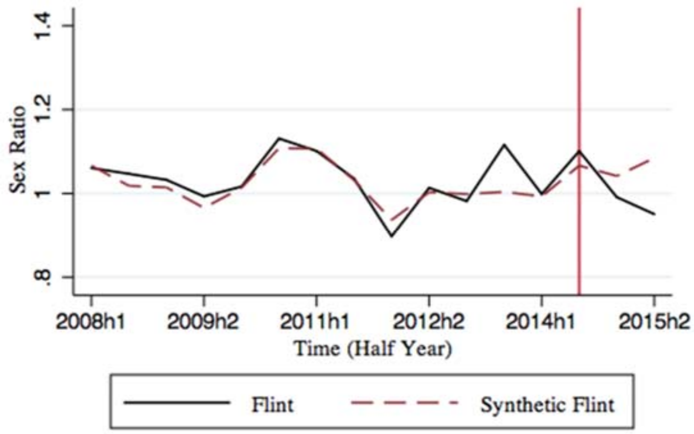


d. Gestation (P-value of RMSPE = 0.018)

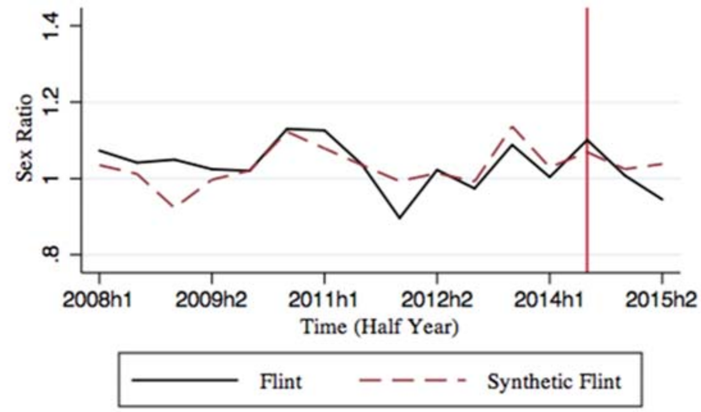


e. Prematurity (P-value of RMSPE = 0.583)

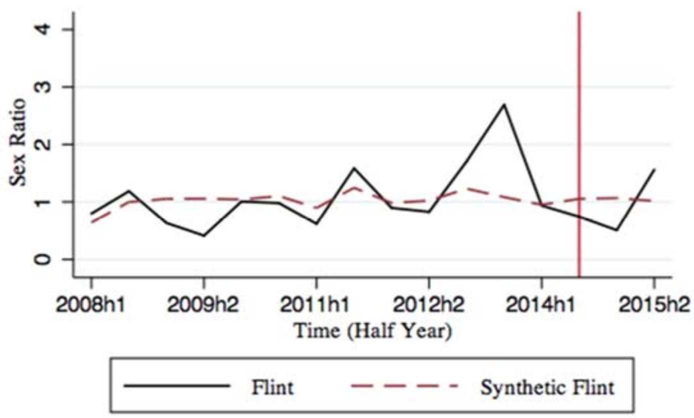
Figure 5: Trends in Sex Ratio (Male to Female)



a. Male to Female Sex Ratio (P-value of RMSPE = 0.276)

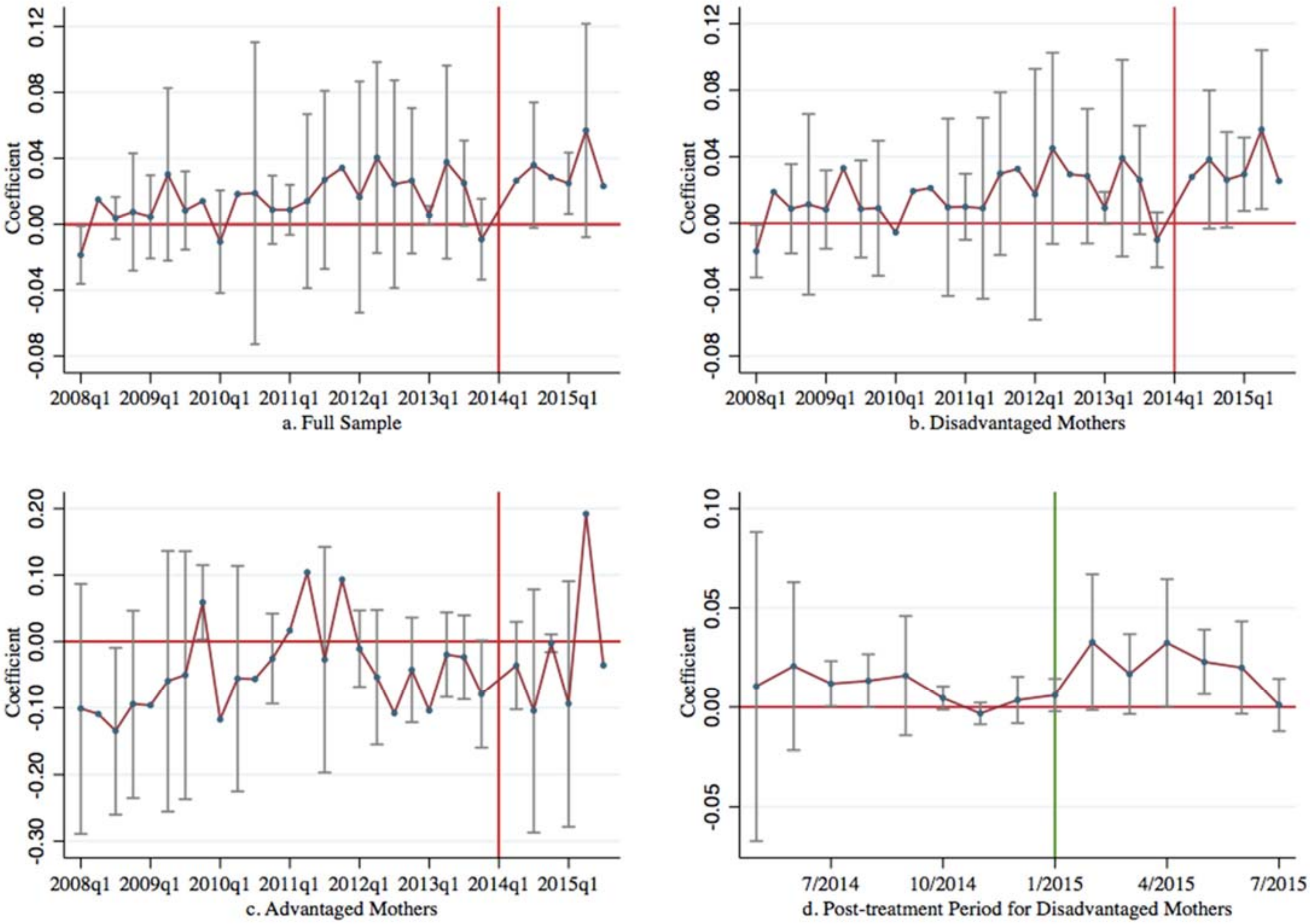


b. Male to Female Sex Ratio with Disadvantaged Mothers (P-value of RMSPE = 0.558)



c. Male to Female Sex Ratio with Advantaged Mothers (P-value of RMSPE = 0.459)

Figure 6: Dynamic Effects of Flint Water Crisis on The Probability of LBW



Notes: Each dot in the figure represents a point estimate with a 95% confidence interval. In figure d, the label of x-axis indicates the first month of each three-month period starting from May of 2014. The green line shows the time when a large proportion of Flint residents started to pay attention to the lead in drinking water.

Table 1: Summary Statistics for Samples

	Flint	6 Cities in Michigan	162 Cities in the United States
Birth Outcome			
Birth weight	3121.291	3207.782	3273.386
	(603.262)	(609.228)	(565.3924)
% LBW	0.116	0.092	0.068
	(0.320)	(0.288)	(0.252)
% VLBW	0.021	0.019	0.021
	(0.143)	(0.134)	(0.004)
Length of gestation	38.300	38.571	38.697
	(3.029)	(2.778)	(2.403)
% Prematurity	0.174	0.131	0.104
	(0.379)	(0.338)	(0.305)
% Male	0.508	0.511	0.512
	(0.500)	(0.500)	(0.500)
Mother Characteristics			
% Younger than 18	0.057	0.041	0.030
	(0.231)	(0.197)	(0.171)
% Aged between 18 and 35	0.877	0.845	0.822
	(0.329)	(0.362)	(0.383)
% Older than 35	0.067	0.114	0.148
	(0.249)	(0.318)	(0.355)
% White	0.402	0.424	0.690
	(0.490)	(0.494)	(0.463)
% Black	0.587	0.524	0.208
	(0.492)	(0.499)	(0.406)
% Less than high school	0.279	0.248	0.208
	(0.449)	(0.432)	(0.406)
% High school	0.324	0.315	0.264
	(0.468)	(0.464)	(0.441)
% Some college	0.343	0.262	0.261
	(0.475)	(0.440)	(0.439)
% Bachelor's degree	0.041	0.108	0.169
	(0.197)	(0.310)	(0.375)
% Graduate school	0.013	0.068	0.097
	(0.112)	(0.252)	(0.296)
% Married mother	0.210	0.389	0.541
	(0.407)	(0.488)	(0.498)
Previous termination	0.293	0.299	0.238
	(0.455)	(0.458)	(0.426)
Prenatal care	2.943	2.999	2.982
	(1.698)	(1.639)	(1.598)
Weight gain	32.085	30.059	30.109
	(16.762)	(15.788)	(14.520)
Diabetes	0.065	0.046	0.053
	(0.247)	(0.210)	(0.224)
Chronic hypertension	0.018	0.016	0.013
	(0.133)	(0.126)	(0.113)
Pregnancy associated hypertension	0.043	0.037	0.042
	(0.203)	(0.189)	(0.202)
Observation			
Total Births	15,425	152,332	5,476,852

Table 2: Main Results on Birth Outcomes

	Panel A: Difference in Differences Design				Panel B: Synthetic Control Methods	
	(1)	(2)	(3)	(4)		(5)
			Birth Weight (gram)			
Flint and Post	-16.001	-34.381**	-34.241**	-6.479	Treatment Effect	-32.030
	(20.767)	(14.658)	(16.417)	(5.788)	P-value of RMSPE	0.190
Observations	167,757	167,757	167,757	167,757		
			Probability of LBW			
Flint and Post	0.014	0.021**	0.021**	0.011*	Treatment Effect	0.018*
	(0.015)	(0.010)	(0.009)	(0.006)	P-value of RMSPE	0.067
Observations	167,757	167,757	167,757	167,757		
			Probability of VLBW			
Flint and Post	0.002	0.004***	0.004***	0.002	Treatment Effect	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	P-value of RMSPE	0.393
Observations	167,757	167,757	167,757	167,757		
			Length of Gestation (week)			
Flint and Post	-0.200**	0.034	0.030	0.079	Treatment Effect	0.202
	(0.092)	(0.084)	(0.086)	(0.068)	P-value of RMSPE	0.239
Observations	167,656	167,656	167,656	167,656		
			Probability of Prematurity			
Flint and Post	0.008***	0.012***	0.012***	-0.011	Treatment Effect	0.013
	(0.003)	(0.004)	(0.004)	(0.015)	P-value of RMSPE	0.798
Observations	167,656	167,656	167,656	167,656		
Covariates	No	Yes	Yes	Yes		
City FEs	No	No	Yes	Yes		
Year FEs	No	No	Yes	Yes		
Month FEs	No	No	Yes	Yes		
City-Month trends	No	No	No	Yes		

Notes: Values shown in column 1 to 4 are the coefficients of OLS regressions of birth weight, the binary variable of being the low-birth-weight birth, the very low-birth-weight birth, the length of gestation, and the binary variable of being premature birth on the regressands. Standard errors are clustered at city level by using wild bootstrap method with imposing null hypothesis of 0. Treatment infants are those with mothers living in Flint. Control infants are those with mothers living in the six control cities in Michigan. The sample uses data from 2008 to 2015. Covariates include mother's age (<18, 18-34, 35+), mother's race (white, black, others), mother's education (<12, 12, some college, college, college+, missing), mother married, diabetic status of mother, diabetic status of mother missing, mother with chronic hypertension, mother with chronic hypertension missing, mother with gestational hypertension, mother with gestational hypertension missing, mother with previous pregnancy termination, mother with previous termination missing, mother's weight gain, mother's weight gain missing, the week of mother's first prenatal care, the week of mother's first prenatal care missing, child male, and unemployment rate of the city. Column 5 shows the results estimated by Synthetic Control Methods. P-value of RMSPE indicates the chances of obtaining the ratio of post root mean squared prediction error (RMSPE) to the pre RMSPE for Flint.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 3: Placebo Tests by Applying Pseudo Shocks at Different Times Before the Flint Water Crisis

Regressand	May of 2009 Mean=0.107 Std.Dev.=0.209	May of 2010 Mean= 0.106 Std.Dev.=0.308	May of 2011 Mean=0.109 Std.Dev.=0.312	May of 2012 Mean=0.112 Std.Dev.=0.316	May of 2013 Mean=0.115 Std.Dev.=0.319
Flint and Post	0.005	0.016	0.011	-0.005	-0.019
	(0.008)	(0.015)	(0.233)	(0.008)	(0.077)
Covariates	Yes	Yes	Yes	Yes	Yes
City FEs	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes
Month FEs	Yes	Yes	Yes	Yes	Yes
City-Month trends	Yes	Yes	Yes	Yes	Yes
Observations	138392	138392	138392	138392	138392
R^2	0.047	0.047	0.047	0.047	0.047

Notes: Values shown are the coefficients of OLS regressions of a binary variable for having LBWon the regressands. Standard errors are clustered at city level by using wild bootstrap method with imposing null hypothesis of 0. Treatment infants are those with mothers living in the Flint. Control infants are those with mothers living in the Six control cities. Columns 1, 2, 3, 4, and 5 use May of 2009, May of 2010, May of 2011, May of 2012, and May of 2013 as the first months of the fake Flint water crises, respectively. All columns use samples before May of 2015. Specifically, Column 1 uses samples from January of 2008 to April of 2009 and from February of 2010 to April of 2014. Column 2 uses samples from January of 2008 to April of 2010 and from February of 2011 to April of 2014. Column 3 uses samples from January of 2008 to April of 2011 and from February of 2012 to April of 2014. Column 4 uses samples from January of 2008 to April of 2012 and from February of 2013 to April of 2014. Column 5 uses samples from January of 2008 to April of 2013 and from February of 2014 to April of 2014. Covariates are the same as those in Table 2.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Table 4: Placebo Tests by Applying Pseudo Shocks in Six Control Cites in MI

	Ann Arbor	Detroit	Grand Rapids	Lansing	Sterling Heights	Warren
Regressand	Mean=0.050 Std.Dev.=0.218	Mean=0.114 Std.Dev.=0.318	Mean= 0.069 Std.Dev.=0.253	Mean= 0.075 Std.Dev.=0.263	Mean=0.060 Std.Dev.=0.237	Mean=0.075 Std.Dev.=0.263
Flint and Post	-0.016 **	0.005	-0.010	-0.003	-0.006	0.010
	(0.008)	(0.004)	(0.014)	(0.007)	(0.007)	(0.037)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
City FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Month FEs	Yes	Yes	Yes	Yes	Yes	Yes
City-Month trends	Yes	Yes	Yes	Yes	Yes	Yes
Observations	167757	167757	167757	167757	167757	167757
R^2	0.046	0.046	0.046	0.046	0.046	0.046

Notes: Values shown are the coefficients of OLS regressions of a binary variable for having LBW on the regressands with robust standard errors. Treatment infants are those with mothers living in the city with a fake shock starting from May of 2014. Control infants are those with mothers living in the six largest cities in Michigan. Column 1, 2, 3, 4, 5, and 6 treat Ann Arbor, Detroit, Grand Rapids, Lansing, Sterling Heights, and Warren as the fake treatment city, respectively. The sample uses data from January of 2008 to December of 2015. Covariates are the same as those in Table2.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table 5: Main Results on Birth Outcomes: Disadvantaged Mothers vs. Advantaged Mothers

	Panel A: Difference in Differences Design					Panel B: Synthetic Control Methods	
	Disadvantaged Mothers		Advantaged Mothers			Disadvantaged Mothers	Advantaged Mother
	(1)	(2)	(3)	(4)		(5)	(6)
				Birth Weight (gram)			
Flint and Post	-27.891**	-1.293	-185.178**	-104.670	Treatment Effect	-35.717*	-141.931
	(11.891)	(3.188)	(73.707)	(1854.87)	P-value of RMSPE	0.098	0.508
Observations	144,485	144,485	22,681	22,681			
				Probability of LBW			
Flint and Post	0.020**	0.012**	0.026	-0.021	Treatment Effect	0.020**	0.039
	(0.009)	(0.006)	(0.021)	(0.024)	P-value of RMSPE	0.031	0.344
Observations	144,485	144,485	22,681	22,681			
				Probability of VLBW			
Flint and Post	0.003***	0.002***	0.003	-0.012	Treatment Effect	0.006*	-0.001
	(0.001)	(0.001)	(0.002)	(0.112)	P-value of RMSPE	0.092	0.623
Observations	144,485	144,485	22,681	22,681			
				Length of Gestation (week)			
Flint and Post	0.063	0.094	-0.535*	-0.136	Treatment Effect	-0.003**	-0.732
	(0.084)	(0.077)	(0.279)	(0.230)	P-value of RMSPE	0.018	0.156
Observations	144,389	144,389	22,678	22,678			
				Probability of Prematurity			
Flint and Post	0.007***	-0.015	0.107*	0.063	Treatment Effect	0.027	0.141
	(0.003)	(0.020)	(0.056)	(0.060)	P-value of RMSPE	0.583	0.131
Observations	144,389	144,389	22,678	22,678			
Covariates	Yes	Yes	Yes	Yes			
City FEs	Yes	Yes	Yes	Yes			
Year FEs	Yes	Yes	Yes	Yes			
Month FEs	Yes	Yes	Yes	Yes			
City-Month trends	No	Yes	No	Yes			

Notes: Values in column 1 to 4 in Panel A are the coefficients of OLS regressions of birth weight, the binary variable of being the low-birth-weight birth, the length of gestation, and the binary variable of being premature birth on the regressands. Column 1 and 3 are estimated by the specification (3) in Table 2. Column 2 and 4 are estimated by the specification (4) in Table 2. Standard errors are clustered at city level by using wild bootstrap method with imposing null hypothesis of 0. Panel B shows the results estimated by Synthetic Control Methods. P-value of RMSPE indicates the chances of obtaining the ratio of post root mean squared prediction error (RMSPE) to the pre RMSPE for Flint.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 6: Results on Birth Outcomes by the Timing of Potential Gestational Exposure

	3rd Trimester		2nd and 3rd Trimesters		All Three Trimesters	
	(1)	(2)	(3)	(4)	(5)	(6)
			Birth Weight (gram)			
Flint and Post	-31.439	-6.355	-36.102**	-10.007	-31.338*	-1.405
	(27.504)	(26.574)	(14.198)	(8.462)	(15.919)	(11.072)
Observations	143541	143541	148782	148782	158714	158714
			Probability of LBW			
Flint and Post	0.015	0.007	0.012*	0.004	0.021**	0.011**
	(0.010)	(0.007)	(0.006)	(0.002)	(0.009)	(0.005)
Observations	143541	143541	148782	148782	158714	158714
			Probability of VLBW			
Flint and Post	-0.000	-0.002	-0.001	-0.002	0.005***	0.004**
	(0.017)	(0.069)	(0.040)	(0.004)	(0.002)	(0.002)
Observations	143541	143541	148782	148782	158714	158714
			Length of Gestation (week)			
Flint and Post	0.331	0.365	0.248*	0.287	-0.046	-0.005
	(0.408)	(0.480)	(0.146)	(0.999)	(0.074)	(0.084)
Observations	143458	143458	148699	148699	158631	158631
			Probability of Prematurity			
Flint and Post	-0.025	-0.044	-0.002	-0.022	0.015*	-0.008
	(0.030)	(0.047)	(0.088)	(0.130)	(0.007)	(0.009)
Observations	143458	143458	148699	148699	158631	158631
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
City FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Month FEs	Yes	Yes	Yes	Yes	Yes	Yes
City-Month trends	No	Yes	No	Yes	No	Yes

Notes: We use the date of birth and the length of gestation to calculate the date of conception. The Flint infants who were exposed to the crisis only in their last trimester were born before September of 2014. The Flint infants who exposed to the crisis not in their first trimester but in other two trimesters were born before December of 2014. The Flint infants who have gestational exposure to the crisis in each trimester were born after September of 2014 and have at least 22 weeks of exposure.

Online Appendix

A. Figures and Tables

Figure A1: Timeline of the Flint Water Crisis

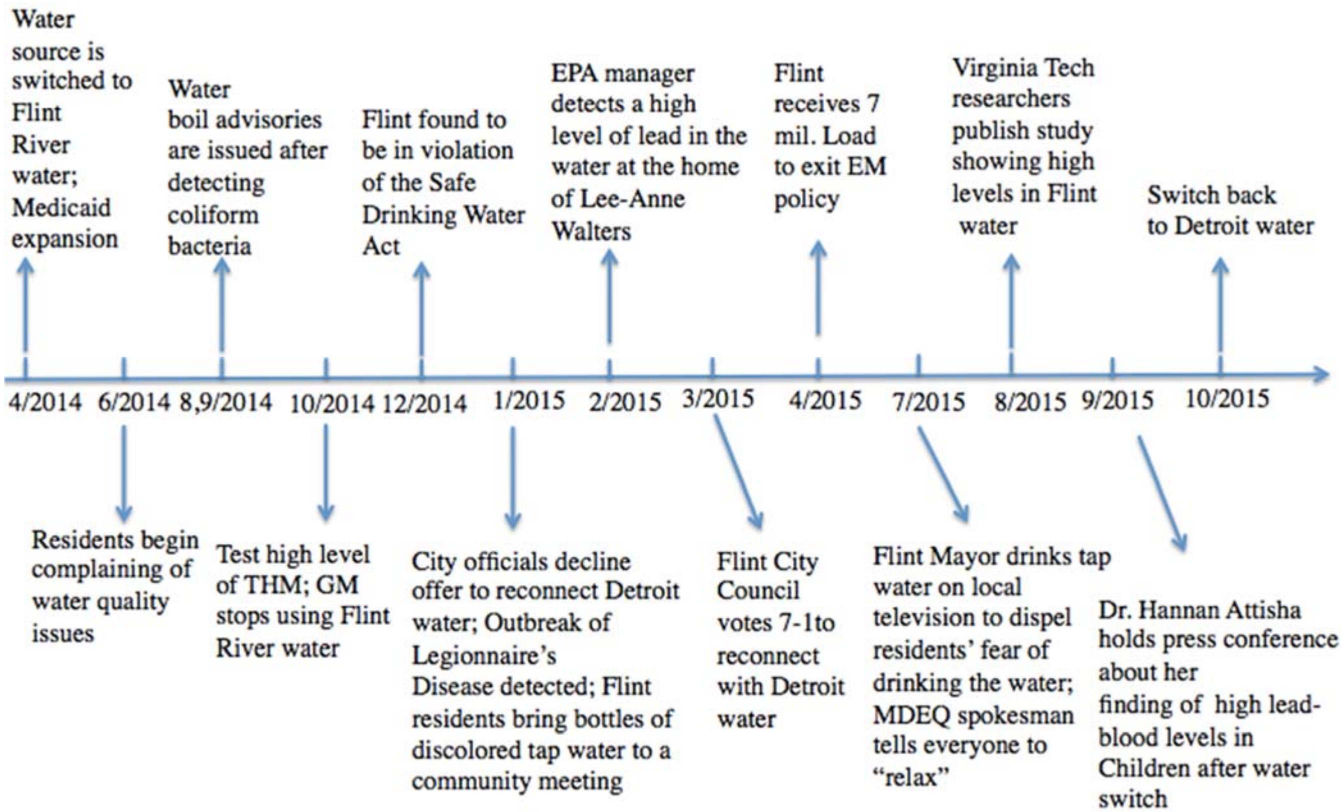
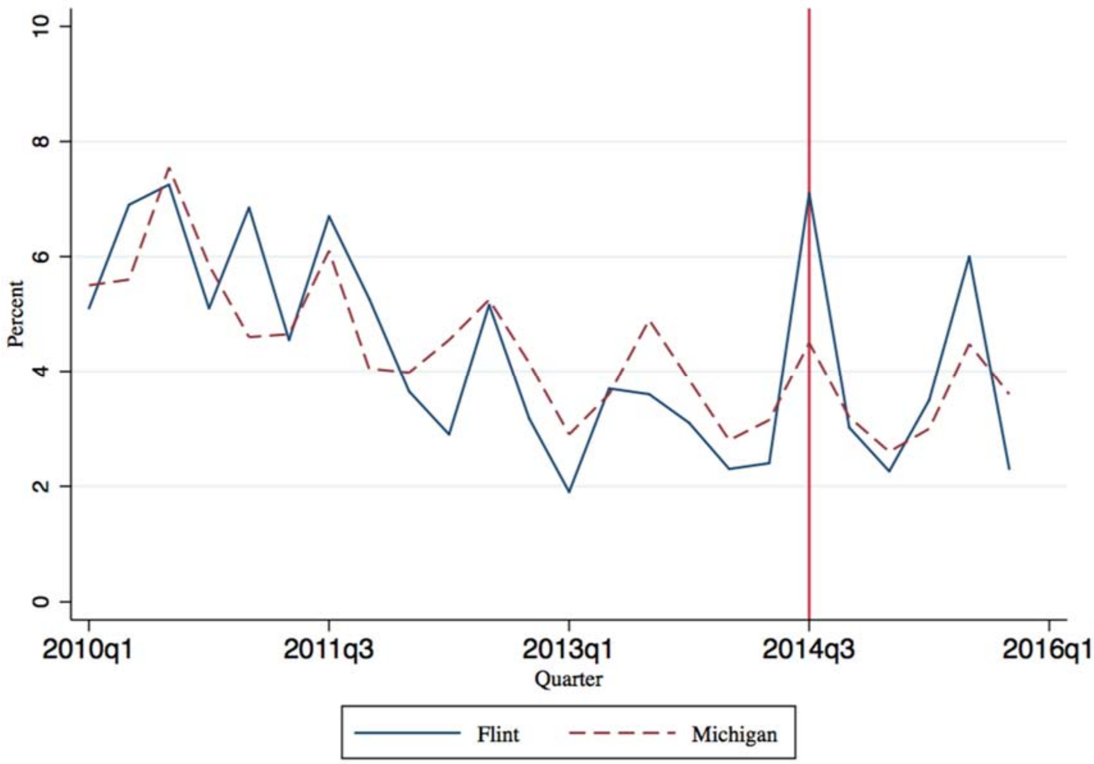
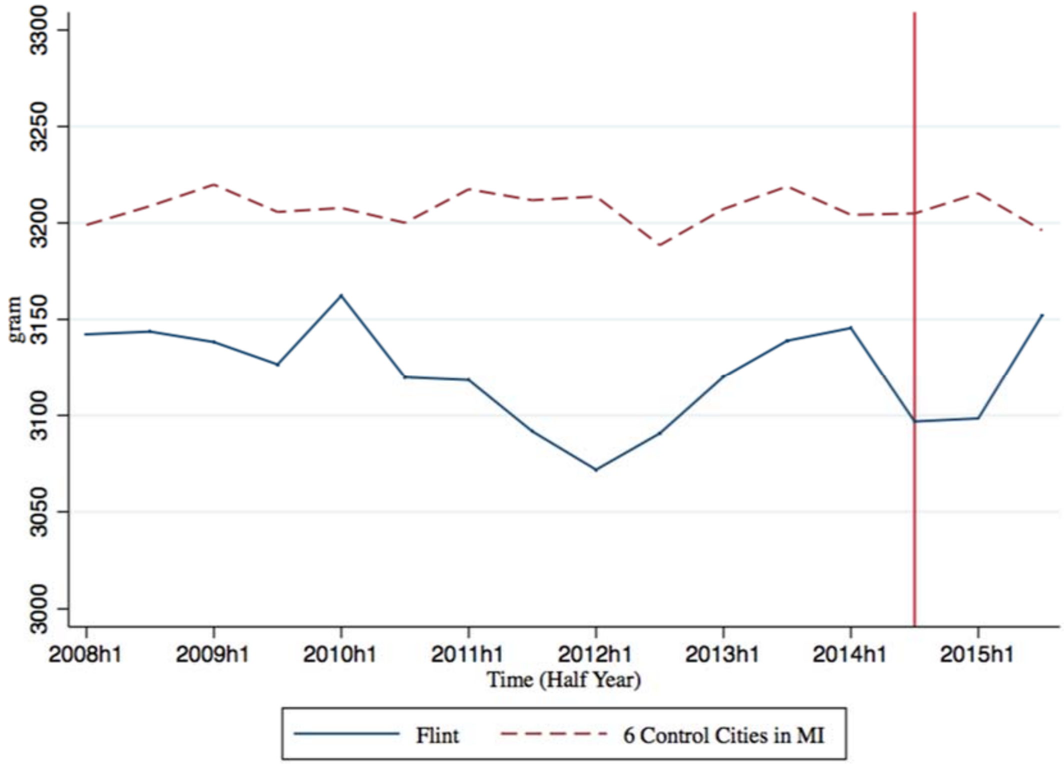


Figure A2: Elevated Blood Lead Levels (5 ug/dl) Among
Tested Children < 6 yrs Old by Quarter

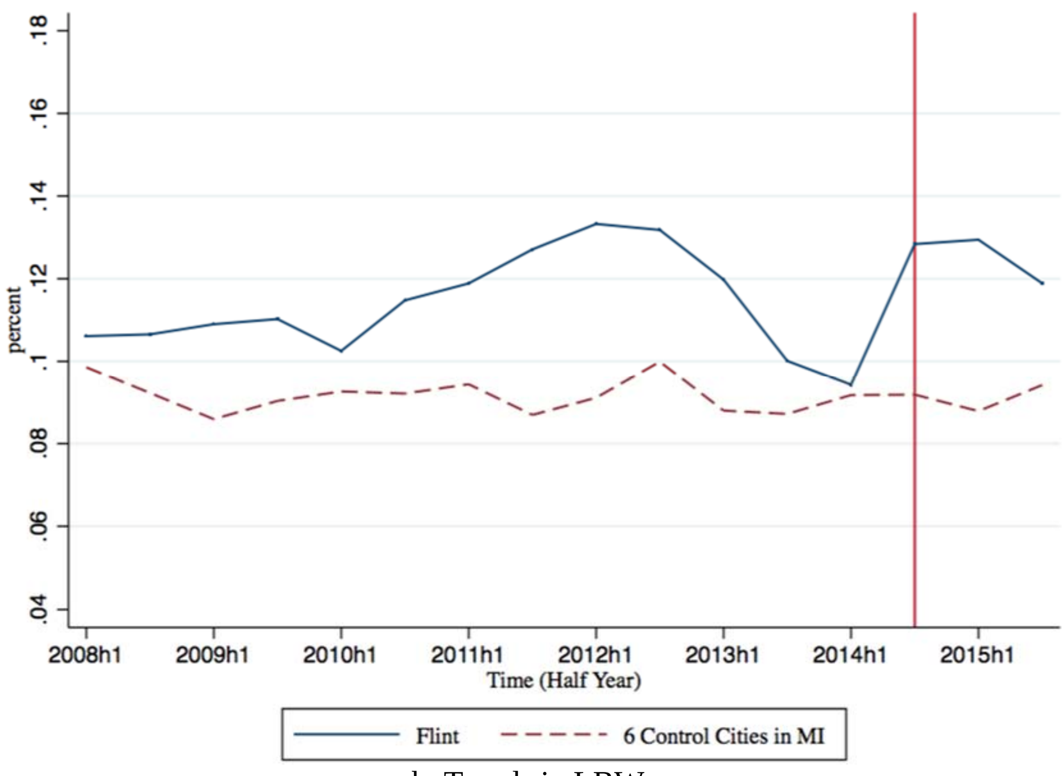


Source: Michigan Department of Environmental Quality (MDEQ). Michigan Department of Health and Human Services (MDHHS) Blood Lead Level Test Results for Flint Zip Codes 48501-48507, Genesee County, and the State of Michigan Summary as of 22 January 2016. Available online: http://www.michigan.gov/documents/flintwater/Flint_Blood_Testing_Report_12Jan22_512221_7.pdf (accessed on 5 February 2016).

Figure A3: Trends in Birth Outcomes: Flint vs. 6 Cities in Michigan (MI)



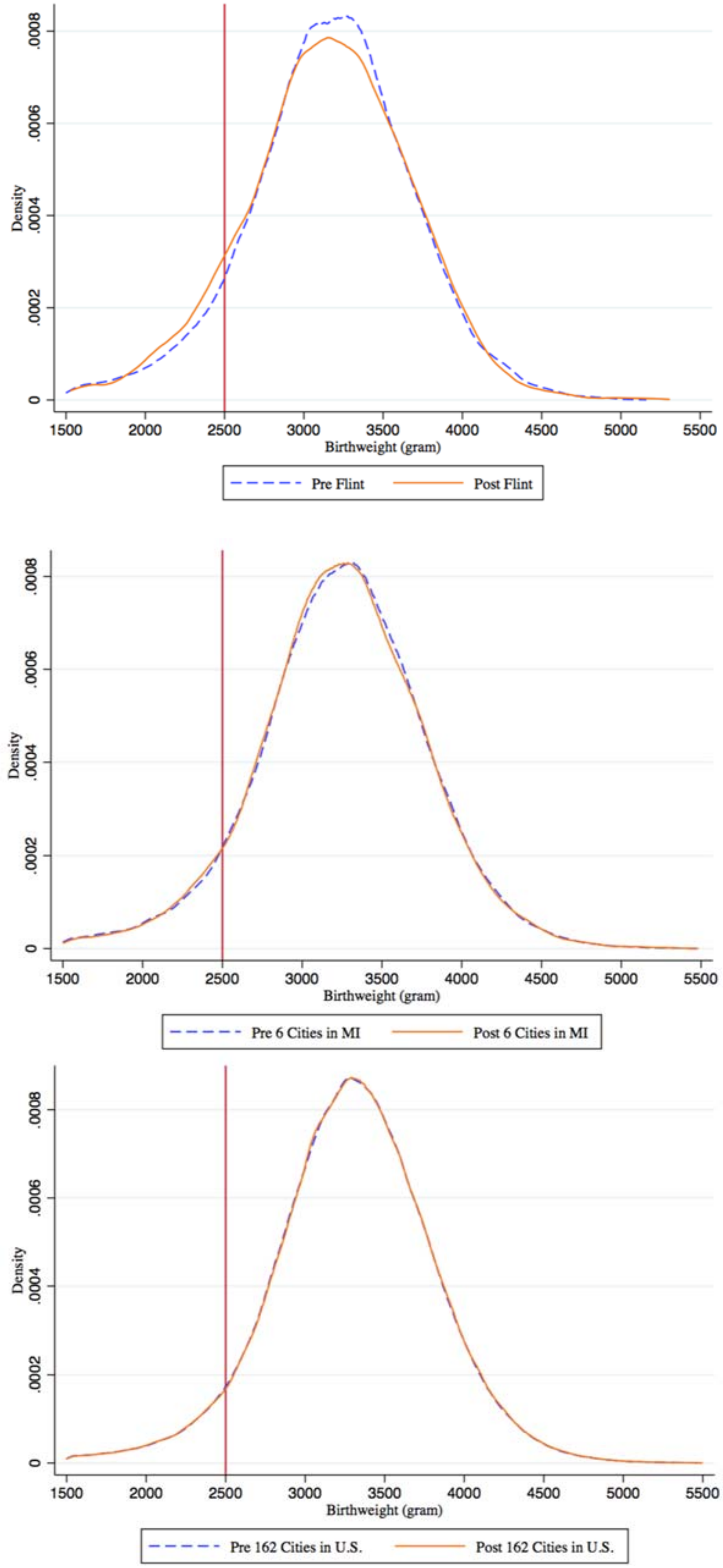
a. Trends in Birth Weight



b. Trends in LBW

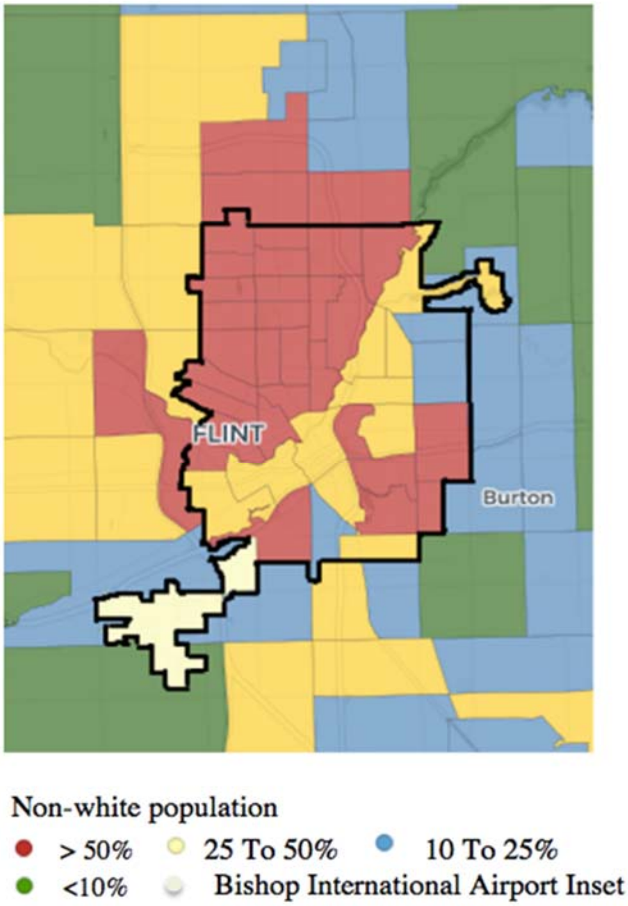
Notes: The vertical red line refers to the second half of 2014 from May to December when the crisis started. 2014h1 refers to January to April of 2014.

Figure A4: The Distributions of Birth Weight Before and After the Flint Water Crisis

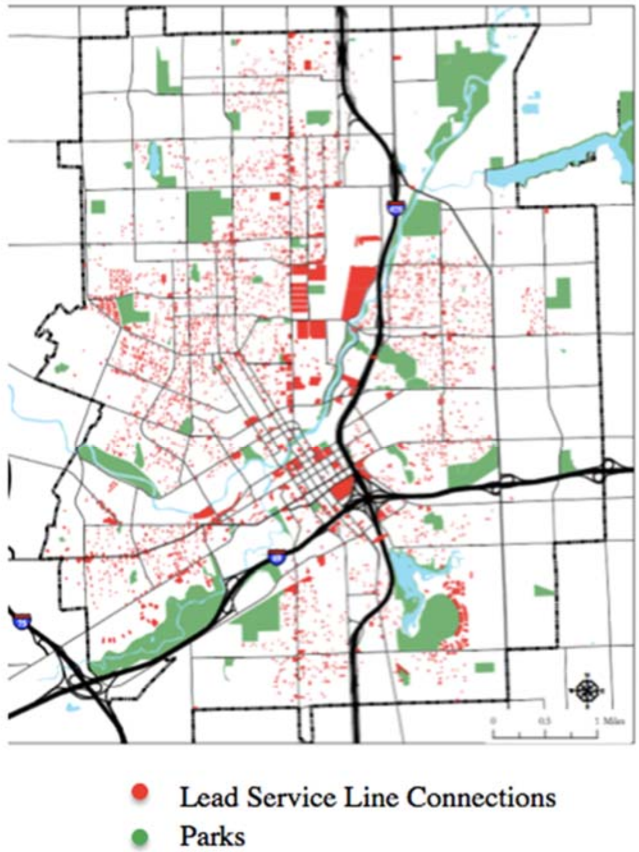


Notes: 6 cities are the control cities in Michigan used in difference-in-differences method. 162 cities in U.S. are the cities in the donor pool for the synthetic method. The vertical red lines are where birth weight equals 2500 grams.

Figure A5: Residence with Lead Pipes in Flint



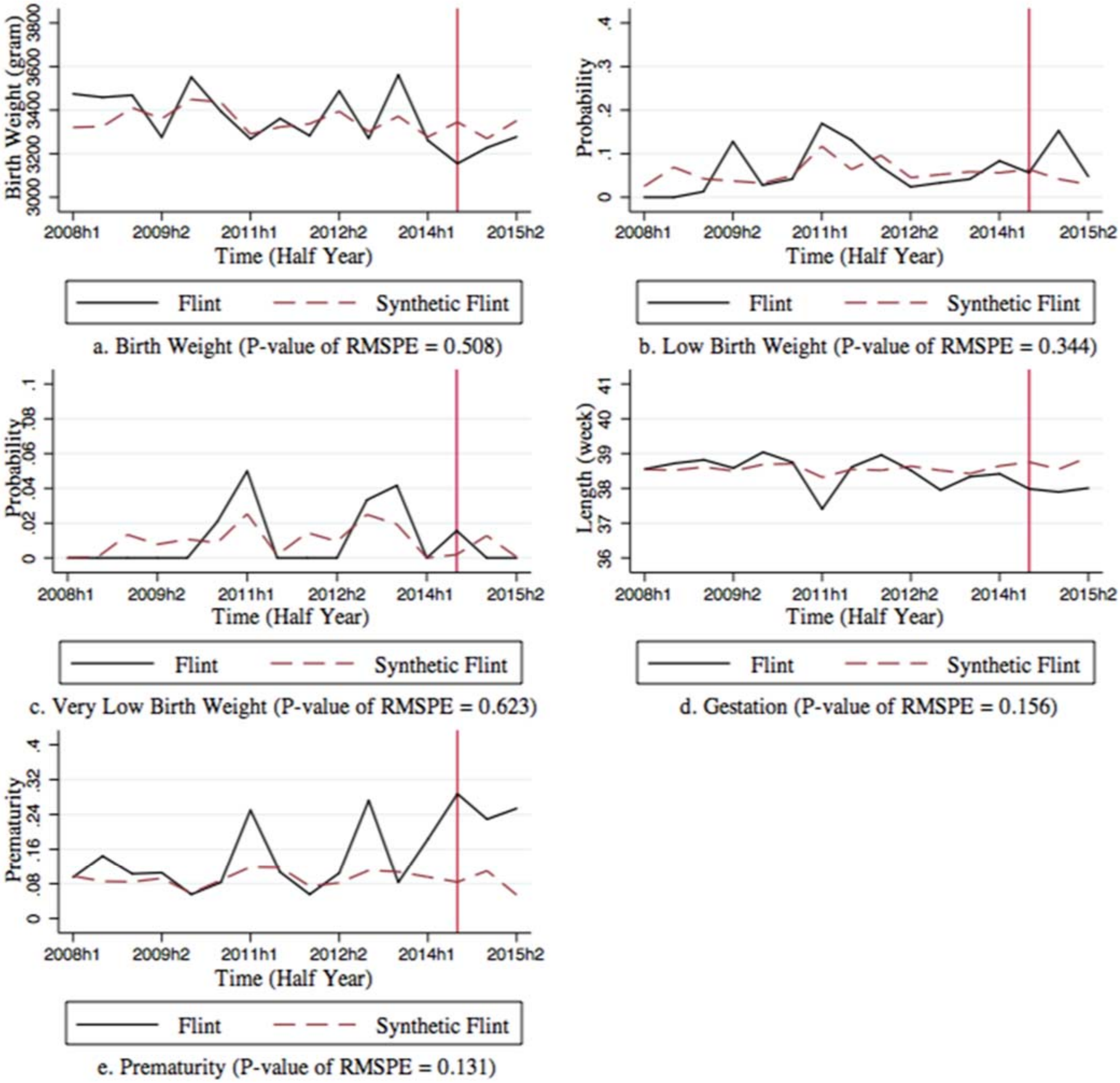
a. Residential Segregation in Flint



b. Map of Lead Water Pipes in Flint

Source: a. Flint, then and now. Bridge. News and analysis from The Center for Michigan. https://www.bridgemi.com/public-sector/flint-then-and-now.
b. Geographic Information Systems Center (GISC), University of Michigan-Flint.

Figure A6: Trends in Birth Outcomes for Advantaged Mothers



a. Birth Weight (P-value of RMSPE = 0.508)

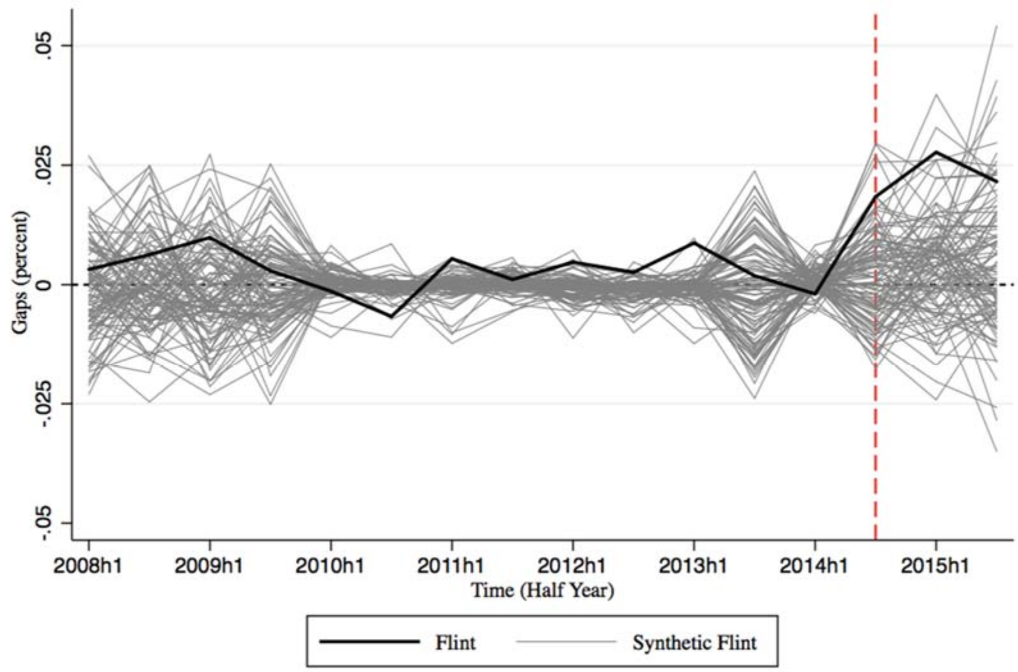
b. Low Birth Weight (P-value of RMSPE = 0.344)

c. Very Low Birth Weight (P-value of RMSPE = 0.623)

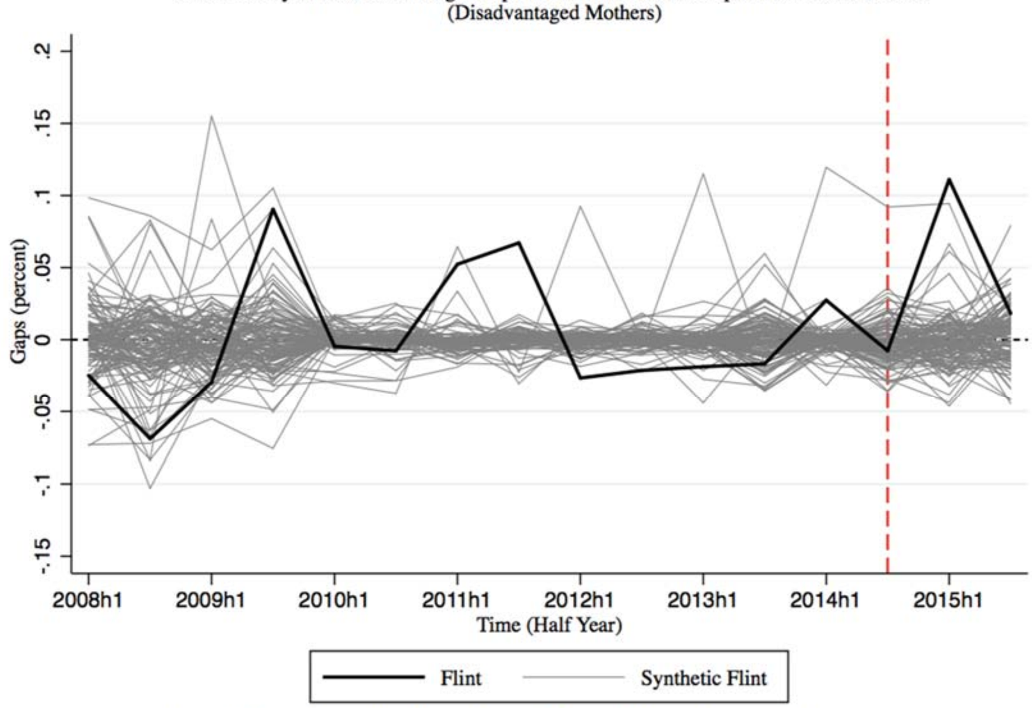
d. Gestation (P-value of RMSPE = 0.156)

e. Prematurity (P-value of RMSPE = 0.131)

Figure A7: Percentage of LBW Gaps in Flint Infants and Placebo Gaps in Control Cities
(Discarding cities with pre-RMSPE more than two times as high as Flint's)

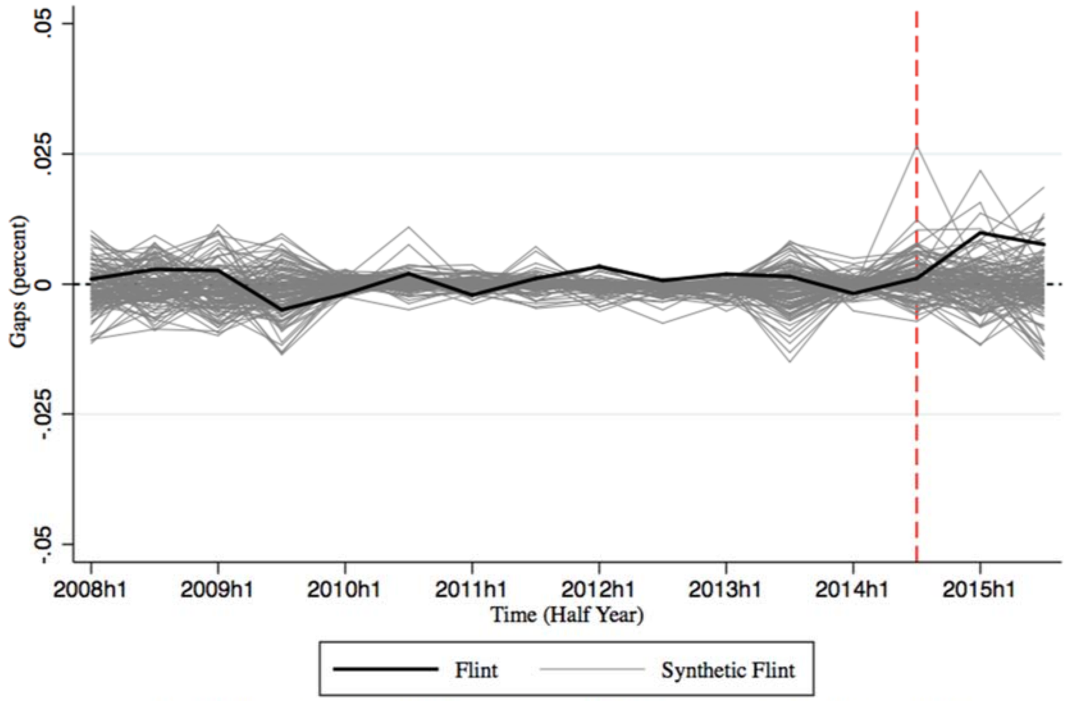


a. Probability of Low Birth Weight Gaps in Flint and Placebo Gaps in 101 Control Cities
(Disadvantaged Mothers)

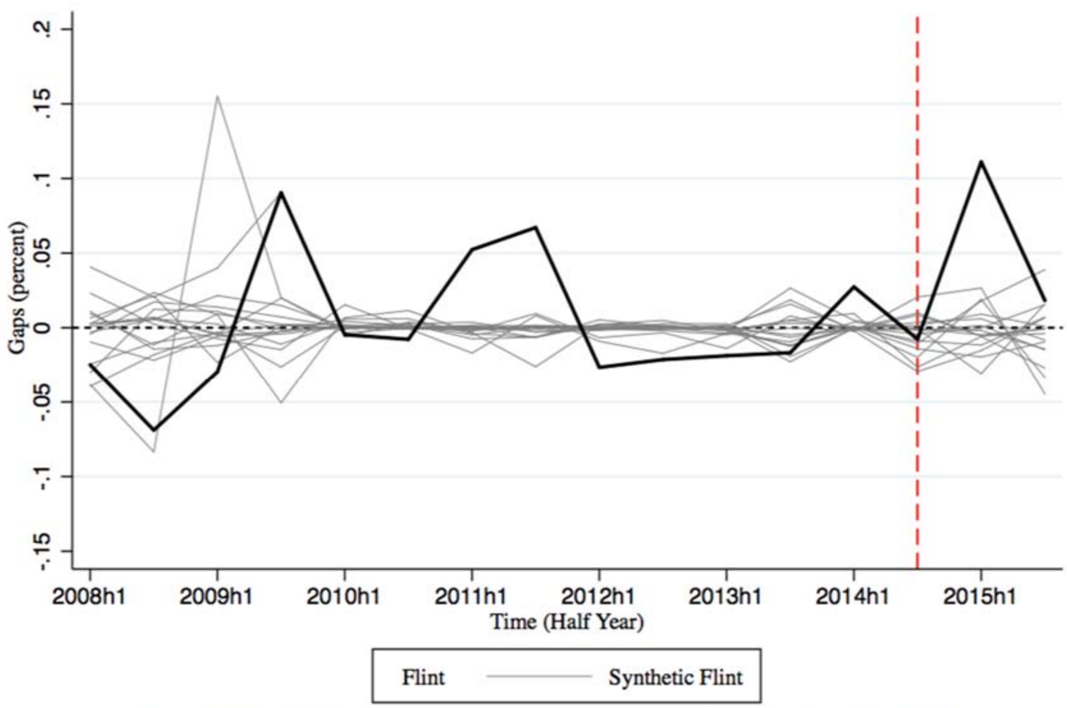


b. Probability of Low Birth Weight Gaps in Flint and Placebo Gaps in 119 Control Cities
(Advantaged Mothers)

Figure A8: Percentage of VLBW Gaps in Flint Infants and Placebo Gaps in Control Cities
(Discarding cities with pre-RMSPE more than two times as high as Flint's)

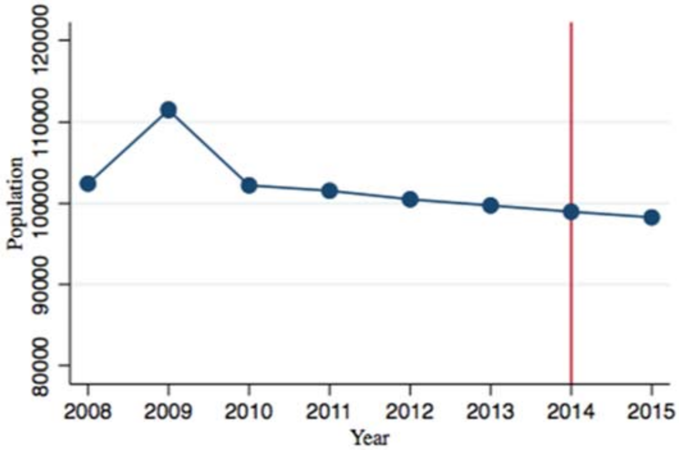


a. Probability of Very Low Birth Weight Gaps in Flint and Placebo Gaps in 113 Control Cities
(Disadvantaged Mothers)

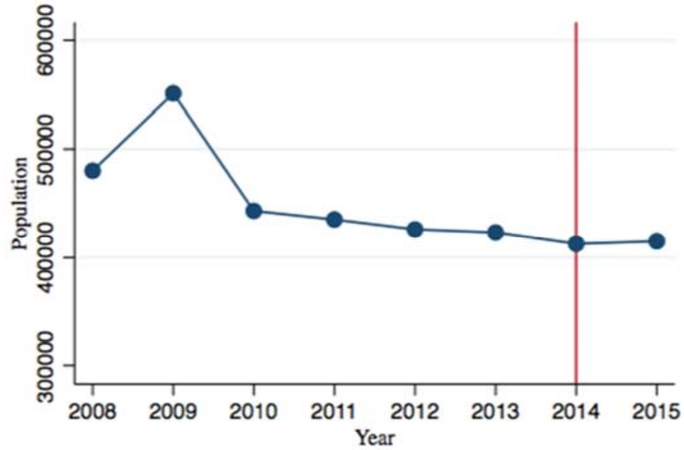


b. Probability of Very Low Birth Weight Gaps in Flint and Placebo Gaps in 19 Control Cities
(Advantaged Mothers)

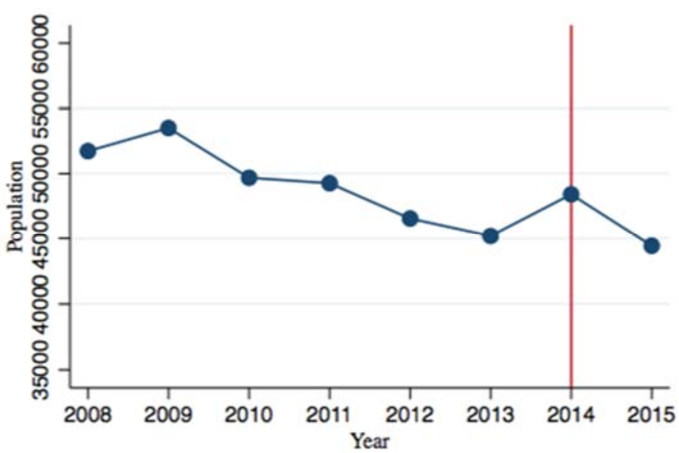
Figure A9: Population in Flint and Six Control Cities



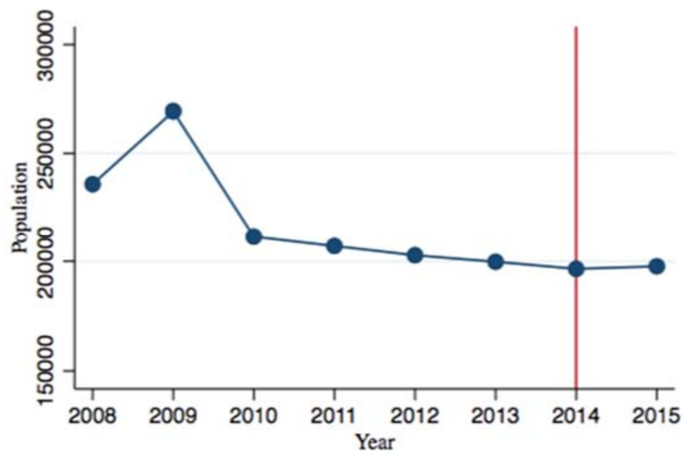
a. Total Population in Flint



b. Total Population in Control Cities

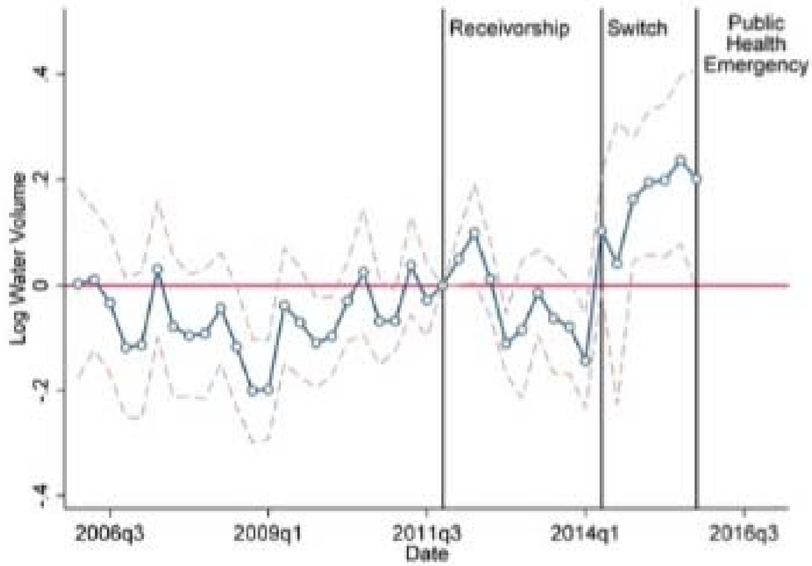


c.The Population of Female Aged 15 to 49 in Flint



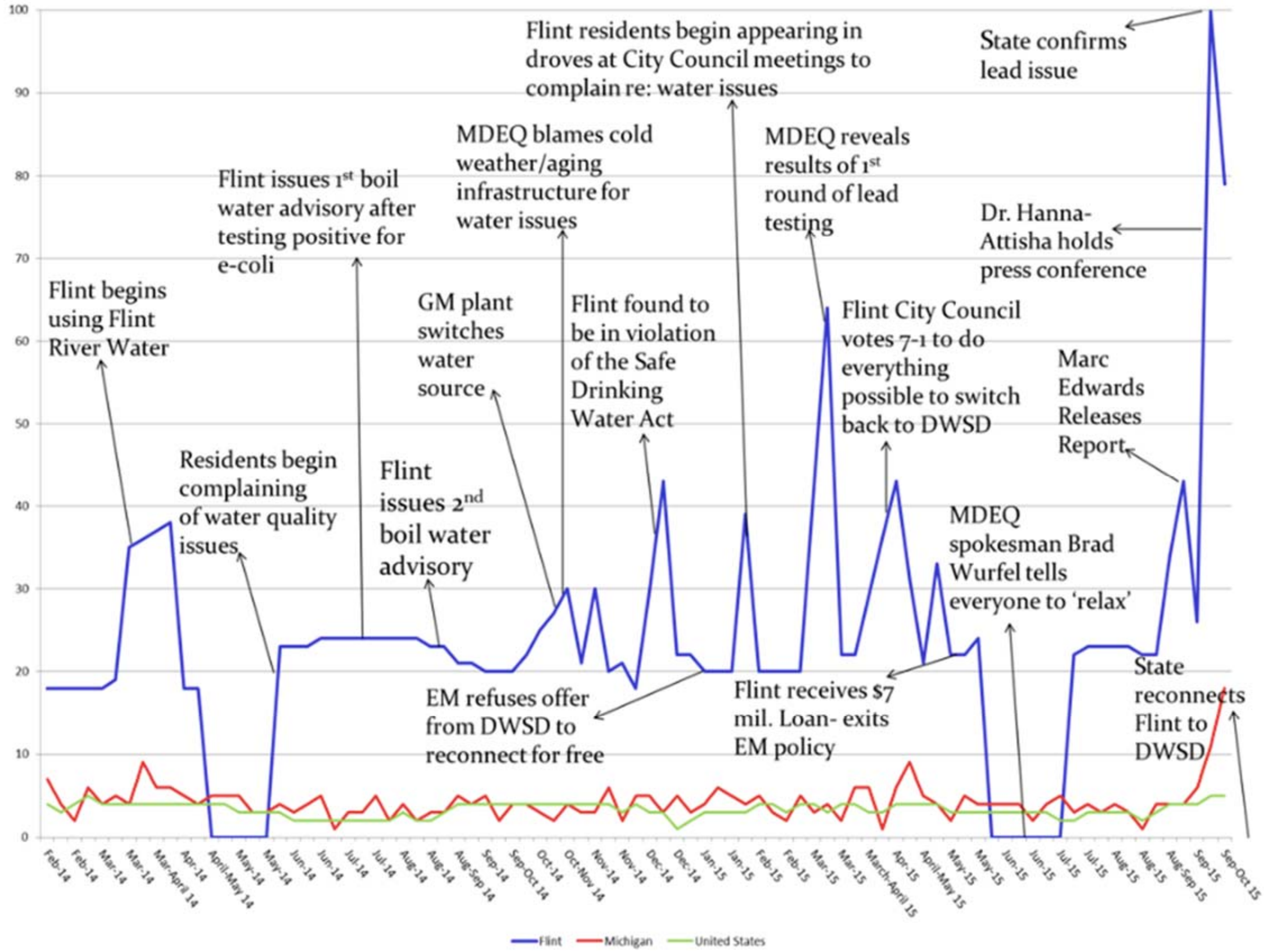
d.The Population of Female Aged 15 to 49 in Control Cities

Figure A10: Bottled Water Sales



Source: The Economic Effects of Environmental Crises: Evidence from Flint, Michigan. (Christensen, Keiser, and Lade, 2017)

Figure A11: Google Search Volume



Source: Institute for Public Policy and Social Research

Table A1: Summary Statistics for Flint and Six Control Cities

	Pre-treatment Period			Post-treatment Period			Difference
	Flint	Control Cities	P-value	Flint	Control Cities	P-value	(Pre&Post)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Birth Outcome							
Birth weight	3124.188	3207.947	0.000***	3107.244	3207.004	0.000***	-16.001
	(603.067)	(610.920)		(604.121)	(601.224)		(13.649)
% LBW	0.114	0.092	0.000***	0.127	0.091	0.000***	0.014**
	(0.317)	(0.289)		(0.333)	(0.287)		(0.007)
% VLBW	0.021	0.019	0.002	0.021	0.017	0.003	0.001
	(0.143)	(0.137)		(0.141)	(0.131)		(0.003)
Length of gestation	38.291	38.574	0.000***	38.340	38.557	0.000***	0.065*
	(3.035)	(2.798)		(3.000)	(2.682)		(0.063)
% Prematurity	0.173	0.132	0.000***	0.176	0.127	0.000***	0.008
	(0.379)	(0.339)		(0.381)	(0.333)		(0.008)
% Male	0.508	0.513	0.308	0.509	0.505	0.708	0.009
	(0.500)	(0.500)		(0.500)	(0.500)		(0.011)
Mother Characteristics							
% Younger than 18	0.060	0.044	0.002***	0.041	0.024	0.004***	0.001
	(0.238)	(0.205)		(0.197)	(0.152)		(0.004)
% Aged between 18 and 35	0.874	0.843	0.003***	0.887	0.852	0.007***	0.004
	(0.331)	(0.363)		(0.316)	(0.355)		(0.008)
% Older than 35	0.066	0.112	0.000***	0.072	0.124	0.000***	-0.005
	(0.237)	(0.316)		(0.259)	(0.330)		(0.007)
% White	0.401	0.426	0.000***	0.406	0.416	0.083*	0.015
	(0.490)	(0.494)		(0.491)	(0.493)		(0.011)
% Black	0.591	0.526	0.000***	0.570	0.515	0.000***	-0.009
	(0.492)	(0.499)		(0.495)	(0.499)		(0.011)
% Less than high school	0.285	0.257	0.000***	0.253	0.205	0.000***	0.020*
	(0.451)	(0.444)		(0.435)	(0.404)		(0.010)
% High school	0.320	0.315	0.191	0.343	0.315	0.024**	0.023**
	(0.467)	(0.464)		(0.475)	(0.464)		(0.010)
% Some college	0.342	0.257	0.000***	0.348	0.286	0.000***	-0.024**
	(0.474)	(0.437)		(0.477)	(0.452)		(0.010)
% Bachelor’s degree	0.040	0.105	0.000***	0.041	0.120	0.000***	-0.014**
	(0.197)	(0.307)		(0.199)	(0.325)		(0.007)
% Graduate school	0.012	0.067	0.000***	0.014	0.073	0.000***	-0.005
	(0.111)	(0.250)		(0.116)	(0.261)		(0.005)
% Married mother	0.210	0.387	0.000***	0.208	0.398	0.000***	-0.013
	(0.408)	(0.487)		(0.406)	(0.489)		(0.011)
Previous termination	0.292	0.300	0.068*	0.300	0.293	0.423	0.016
	(0.455)	(0.458)		(0.459)	(0.455)		(0.010)
Prenatal care	2.990	3.016	0.085*	2.719	2.917	0.000***	-0.171***
	(1.711)	(1.642)		(1.615)	(1.623)		(0.037)
Weight gain	32.105	30.298	0.000***	31.989	28.941	0.000***	1.239***
	(16.735)	(15.821)		(16.887)	(15.582)		(0.359)
Diabetes	0.063	0.046	0.000***	0.077	0.047	0.000***	0.013***
	(0.243)	(0.209)		(0.267)	(0.212)		(0.005)
Chronic hypertension	0.019	0.016	0.033**	0.015	0.016	0.522	-0.003***
	(0.136)	(0.126)		(0.121)	(0.124)		(0.003)
Pregnancy associated hypertension	0.046	0.036	0.000***	0.026	0.045	0.000***	-0.030***
	(0.210)	(0.0.185)		(0.160)	(0.207)		(0.004)
Observation							
Total Births	12788	125604		2637	26728		

Notes: Average monthly births per city have been rounded up from the decimal.
*** Significant at the 1 percent level, ** Significant at the 5 percent level, * Significant at the 10 percent level.

Table A2: The Relationship between Mother's Characteristics and the Flint Crisis

Dependent Variable	(1)	(2)
White Mother	0.029*	0.017
	(0.016)	(0.112)
Weight gain	1.242*	-0.211
	(0.653)	(0.942)
Covariates	Yes	Yes
City FEs	Yes	Yes
Year FEs	Yes	Yes
Month FEs	Yes	Yes
City-Month trends	No	Yes
Observations	167,757	167,757

Note: The results in model (1) are estimated by $y_{ict} = \beta_0 + \gamma * I(F)_i * T_{it} + city_c + \vartheta_t + \varphi_t + \varepsilon_{ict}$, where y_{ict} is the mother's characteristics. Model (2) adds extra city linear trends. We test all maternal characteristics shown in Table A1 and did not find any significant effect.

Table A3: Summary Statistics for Flint Infants:
(Disadvantaged Mothers vs. Advantaged Mothers)

	Pre-treatment Period		Post-treatment Period	
	Disadvantaged Mother	Advantaged Mother	Disadvantaged Mother	Advantaged Mother
Birth Outcome				
Birth weight	3115.516	3382.058	3095.139	3232.56
	(602.713)	(581.238)	(604.200)	(538.303)
% LBW	0.115	0.063	0.128	0.080
	(0.319)	(0.243)	(0.335)	(0.273)
% VLBW	0.021	0.012	0.021	0.010
	(0.144)	(0.109)	(0.143)	(0.100)
Length of gestation	38.285	38.488	38.330	38.280
	(3.056)	(2.353)	(3.024)	(2.442)
% Prematurity	0.175	0.130	0.173	0.180
	(0.175)	(0.337)	(0.378)	(0.386)
% Male	0.509	0.478	0.509	0.500
	(0.500)	(0.500)	(0.500)	(0.503)
Mother Characteristics				
% Younger than 18	0.062	None	0.041	None
	(0.241)		(0.198)	
% Aged between 18 and	0.876	0.822	0.893	0.760
35	(0.329)	(0.382)	(0.309)	(0.429)
% Older than 35	0.062	0.178	0.066	0.240
	(0.241)	(0.383)	(0.249)	(0.429)
% White	0.382	0.945	0.381	0.920
	(0.486)	(0.229)	(0.486)	(0.273)
% Black	0.611	None	0.596	None
	(0.488)		(0.491)	
% Less than high school	0.294	None	0.259	None
	(0.456)		(0.438)	
% High school	0.331	None	0.349	None
	(0.471)		(0.477)	
% Some college	0.354	None	0.366	None
	(0.478)		(0.482)	
% Bachelor’s degree	0.016	0.767	0.021	0.700
	(0.126)	(0.423)	(0.143)	(0.461)
% Graduate school	0.005	0.233	0.005	0.300
	(0.070)	(0.423)	(0.071)	(0.461)
% Married mother	0.189	0.841	0.186	0.800
	(0.392)	(0.366)	(0.389)	(0.402)
Previous termination	0.294	0.221	0.292	0.240
	(0.456)	(0.415)	(0.454)	(0.429)
Prenatal care	2.999	2.704	2.693	2.720
	(1.722)	(1.311)	(1.608)	(1.333)
Weight gain	32.138	30.995	32.140	30.330
	(16.803)	(14.492)	(16.910)	(15.016)
Diabetes	0.063	0.063	0.077	0.070
	(0.243)	(0.242)	(0.266)	(0.256)
Chronic hypertension	0.019	0.022	0.015	None
	(0.135)	(0.146)	(0.123)	
Pregnancy associated	0.046	0.060	0.028	0.040
hypertension	(0.210)	(0.238)	(0.164)	(0.197)
Observation				
Total Births	12355	416	2991	100

Table A4: Testing Potential Population Size Effects of the Flint Crisis

Dependent Variable	(1)	(2)
Total Population	-0.004	-0.017
	(0.015)	(0.110)
Population of females aged 15-49	-0.011	0.004
	(0.033)	(0.041)
Covariates	Yes	Yes
City FEs	Yes	Yes
Year FEs	Yes	Yes
Month FEs	Yes	Yes
City-Month trends	No	Yes
Observations	167,757	167,757

Notes: The sample uses data from January of 2008 to December of 2015. Covariates are the same as those in Table 2.

Table A5: Testing the Potential Mechanism of Mortality Selection of the Flint Crisis

Dependent variable: newborn being male	Full Sample	Disadvantaged Mothers	Advantaged Mothers
Panel A: Difference in Differences Design			
Flint and Post	0.011	0.013	-0.100
	(0.015)	(0.020)	(0.113)
Observations			
Panel B: Synthetic Control Methods			
Treatment Effect	-0.015	-0.033	-0.046
P-value of RMSPE	0.276	0.558	0.459

Notes: Panel A presents OLS estimates of the regressions of a binary variable – newborn being male – on the Regressands using the DID specification (4) in Table 2. Standard errors are clustered at city level using wild bootstrap method with imposing null hypothesis of 0. Panel B shows the results estimated by the Synthetic Control Method. P-value of RMSPE indicates the chances of obtaining the ratio of post root mean squared prediction error (RMSPE) to the pre RMSPE for Flint.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
*Significant at the 10 percent level.

B Description of Synthetic Control Methods

According to Abadie, Diamond, and Hainmueller (2003) (henceforth ADH (2003)), the observed percentage of low-birth-weight infants for city i at time t is:

$$Y_{it} = Y_{it}^N + \alpha_{it} D_{it}, \quad (3)$$

where Y_{it}^N is the percentage of LBW that would be observed for infants in city i at time t in the absence of the crisis, and$D$ indicates whether city i experienced the crisis, where$t = 1, \dots, T_1$ is split into before the crisis ($t = 1, \dots\dots T_0$) and after the crisis ($t = T_0 + 1, \dots, T_1$) . Because only Flint experienced the crisis after April 2014 (denoted by T_0), $D_{it} = 1$ if i is Flint and $t > T_0$, and 0 otherwise.

Let Y_{1t}^I be the percentage of low-birth-weight infants in Flint at a time t ($t > T_0$ when Flint was experiencing the crisis. Then, the effect of the crisis is expressed by

$$\alpha_{1t} = Y_{1t}^I - Y_{1t}^N, \quad (4)$$

where $t = T_0 + 1, \dots, T_1$. Because Y_{1t}^N is unobserved, we need to estimate it. SupposeY_{it}^N is given by a linear factor model as follows:

$$Y_{it}^N = \theta_t + \beta_t X_i + \gamma_t \mu_i + \epsilon_{it}, \quad (5)$$

where θ_t is an unknown common factor across cities, for example, time-fixed effects, X is a vector of pre-treatment observed covariates that are not directly affected by the crisis, $\gamma_t$is a vector of unknown common factors that allows the effects of confounding unobservables to vary with time,μ_i is a vector of unknown factor loading, and ϵ_{jt} is a vector of unobserved transitory shocks at the city level with zero mean.

Based on Equation (5), the outcome for synthetic Flint is defined as the weighted average outcome of each potential control city:

$$\sum_{j=2}^{J+1} w_j Y_{jt} = \theta_t + \beta_t \sum_{j=2}^{j+1} w_j X_j + \gamma_t \sum_{j=2}^{j+1} w_j u_j + \sum_{j=2}^{j+1} w_j \epsilon_{jt}, \quad (6)$$

where w_j is the nonnegative weight for the potential control city j, and $w_2 + \dots + w_{J+1} = 1$.

If there are $(w_2^*, \dots, w_{J+1}^*)$ such that

$$\sum_{j=2}^{J+1} w_j^* Y_{j1} = Y_{11}, \sum_{j=2}^{J+1} w_j^* Y_{j2} = Y_{12}, \dots \sum_{j=2}^{J+1} w_j^* Y_{jT_0} = Y_{1T_0}, \text{ and } \sum_{j=2}^{J+1} w_j^* X_j = X_1, \quad (7)$$

then $\sum_{j=2}^{J+1} w_j^* Y_{jt}$ provides an unbiased estimator ofY_{1t}^N. Then, the estimator ofα_{1t} becomes

$$\hat{a}_{it} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}. \quad (8)$$

To find w_j^* such that $\sum_{j=2}^{J+1} w_j^* X_j = X_1$, let $W^* = (w_2^*, \dots, w_{J+1}^*)'$ be a vector of w_j^*, which minimizes the distance between X_1 (the vector of preintervention characteristics for Flint) and $X_0 W (X_0$ is a matrix that contains the same variables for the unaffected cities):

$$W^* argmini \parallel X_1 - X_0 W \parallel v = \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)} \;, \quad (9)$$

where V is a symmetric and positive semidefinite matrix that represents the weight for each predictor. V can be chosen in many ways. ADH state that the resulting analysis is valid for any predictor weight. This paper chooses V such that the mean squared prediction error (MSPE) of the outcome variable is minimized in the pre-treatment period.

To apply the SCM, the first step is to identify predictors. Besides maternal characteristics, we use average number of birth and birth outcomes in periods before the Crisis as predictors to adjust for differences of birth outcomes between Flint and the control cities. It is common to include outcomes for a part of the pre-treatment period in the matching process to control for heterogeneity in the observed and unobserved factors on the outcome of interest (Abadie et al. 2003; Abadie et al., 2015). However, if matching outcomes for each pre-treatment period, the effects of other predictors are eliminated with barely apparent weights. If so, outcomes for the synthetic unit in the post-treatment period may be biased when other predictors have predictive value for the outcomes. Thus, we only include birth outcomes for some parts of the pre-treatment period. Figure A3 shows Flint experienced a drop in the average birth weight and a sharp increase in the probability of LBW from 2010 to 2012, indicating an unobserved shock on birth outcomes before the Crisis. To capture the effect of this shock, we choose birth outcomes during these periods as some of our predictors.

The next step is to identify the set of potential control cities as the "donor pool" that will synthesize Flint. Similar crises or pollution incidents should not occur in any donor pool city at any time during the study to avoid potential downward bias. Although Flint is the only city that experienced a disclosed crisis, it is not the only city that had water pollution during our study period. At least 10% of Americans regularly drink unhealthy water and around 50 million Americans drink water containing industrial solvents or other chemicals with health risks. More than 20% of public water systems in the U.S. violated provisions of the SDWA over years (Fung et al., 2007). To exclude cities that may have similar crises but were not reported, we choose cities that did not have any health-based drinking water violations of SDWA during the study.[14]

To assess the significance of the estimates, we conduct a series of placebo tests by iteratively applying the SCM to every other city in the donor pool. In each iteration, we reassign the crisis to one of the donor cities, shifting Flint to the donor pool. That is, we proceed as if one of the cities in the donor pool would have had the crisis instead of Flint. This iterative procedure provides distributions of the ratios of post/pre-crisis Root Mean Square Prediction Error (post/pre RMSPE) for Flint and cities in the donor pool. Following Abadie et al. (2003, 2010) and McClelland and Gault (2017), we generate a permutation p-value based on the location of Flint's post/pre RMSPE ratios in the distributions of birth weight and the rate of LBW, respectively.

[14] We observed over 150 major cities across the U.S. with health-based drinking water violations of SDWA from 2008 to 2015.