Published in final edited form as:

Sci Total Environ. 2017 July 15; 590-591: 139–153. doi:10.1016/j.scitotenv.2017.02.183.

Monitoring the Aftermath of Flint Drinking Water Contamination Crisis: Another Case of Sampling Bias?

Pierre Goovaerts

BioMedware Inc, Ann Arbor, MI 48106

Abstract

The delay in reporting high levels of lead in Flint drinking water, following the city's switch to the Flint River as its water supply, was partially caused by the biased selection of sampling sites away from the lead pipe network. Since Flint returned to its pre-crisis source of drinking water, the State has been monitoring water lead levels (WLL) at selected "sentinel" sites. In a first phase that lasted two months, 739 residences were sampled, most of them bi-weekly, to determine the general health of the distribution system and to track temporal changes in lead levels. During the same period, water samples were also collected through a voluntary program whereby concerned citizens received free testing kits and conducted sampling on their own. State officials relied on the former data to demonstrate the steady improvement in water quality. A recent analysis of data collected by voluntary sampling revealed, however, an opposite trend with lead levels increasing over time. This paper looks at potential sampling bias to explain such differences. Although houses with higher WLL were more likely to be sampled repeatedly, voluntary sampling turned out to reproduce fairly well the main characteristics (i.e. presence of lead service lines (LSL), construction year) of Flint housing stock. State-controlled sampling was less representative; e.g., sentinel sites with LSL were mostly built between 1935 and 1950 in lower poverty areas, which might hamper our ability to disentangle the effects of LSL and premise plumbing (lead fixtures and pipes present within old houses) on WLL. Also, there was no sentinel site with LSL in two of the most impoverished wards, including where the percentage of children with elevated blood lead levels tripled following the switch in water supply. Correcting for sampling bias narrowed the gap between sampling programs, yet overall temporal trends are still opposite.

Graphical abstract

Corresponding author phone: 517-900-9283; fax: 734-913-2201; goovaerts @biomedware.com. Corresponding author address: BioMedware, Inc. PO Box 1577, Ann Arbor, MI 48106 (USA).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Sampling bias; Blood lead levels; Lead-and-Copper Rule; Poverty; Lead service lines

Introduction

The drinking water contamination crisis in Flint, Michigan was a painful reminder that biased monitoring and sampling procedures can be used to hide the true extent of environmental disasters. The delay in reporting high lead levels following April 2014's change in water supply, which resulted in water with high chloride and no corrosion inhibitor flowing through the aging Flint water distribution system (Flint Water Advisory Task Force, 2016), was partially caused by the biased selection of sampling sites. Flint's water testing from late 2014 missed the bulk of the city's lead pipe network, which along with home plumbing fixtures (e.g., leaded brass and solder), is the primary source of lead being leached from chlorine-induced corrosion. Instead, according to Milman, (2016) the sampling targeted properties on the eastern and western fringes of the city which, in some cases, were a long way from any apparent source of lead. Even more troubling was the news that such sampling practices were being used by other public water systems throughout the country (Milman and Glenza, 2016).

According to EPA there is no safe level of lead in drinking water as this toxic metal can be harmful to human health even at low exposure levels. Yet, the Lead and Copper Rule (LCR, US EPA, 1991, 2002, 2016a) allows a sizeable number of first-draw 1-L water samples collected from high-risk homes to exceed the action level of 15 $\mu\text{g/L}$ for lead. Indeed, as long as this level is exceeded at no more than 10% of sampled residences, the public water system is in compliance, with no legal requirement to take action and notify local populations. High-risk homes are defined as sites where elevated levels of lead are likely to be found based on the presence of lead service lines (LSL), lead pipes, or copper pipes soldered with lead installed after 1982 but before 1985 when solder containing high concentrations of lead was banned. Ideally, water should be sampled from Single Family Residences with half samples collected at LSL sites and half from sites with lead pipes or copper pipes with lead solder.

The LCR is statistical in nature as it applies to a sampled set instead of a single measurement and involves both a percentage threshold (10%) and a chemical threshold (15 µg/L). Since LCR compliance is controlled by how water lead levels (WLL) recorded at a number of sites measure up with respect to an action level, water testing can be manipulated at two different levels to mask potential problems. At the residence-level WLLs can be under-estimated by adopting different tactics known to lower the amount of lead in water samples (Milman and Glenza, 2016). These include running faucets prior to the 6 hrs. stagnation period (pre-stagnation flushing), removing or cleaning faucet filters called “aerators” where lead particles can be trapped, or sampling at unrealistically low flow rates to reduce the amount of lead and other material that is dislodged from pipes (e.g., use narrow-necked bottles that cannot be filled at normal flow rates). The second type of manipulation can occur at the level of the water system and is achieved through the biased selection of sampling sites to ensure that the 10% cutoff is not exceeded. An obvious way is to avoid sampling tier 1 category houses (i.e. single family structures containing copper pipes with lead solder installed after 1982 or containing lead pipes; and/or are served by a lead service line), which can go as far as asking water department employees to test water safety in their own homes (Milman and Glenza, 2016). Another tactic is to “bump out” a test result that found very high levels of lead by testing more homes (Felton, 2016).

Since Flint returned to its pre-crisis source of drinking water on October 16, 2015, close to 25,000 water samples have been collected and tested for lead and copper in more than 10,000 residences. Most of these samples (80%) were collected through voluntary or homeowner-driven sampling whereby concerned citizens received free testing kits from local water distribution centers and conducted sampling on their own following instructions provided by MDEQ. This type of crowd sourcing was supplemented by a State-controlled monitoring, called sentinel program. In a first phase that lasted two months, this program aimed to determine the general health of the distribution system and to track temporal changes in lead concentrations through the biweekly sampling of more than 600 sentinel sites chosen by the EPA and MDEQ (Flint Safe Drinking Water Task Force, 2016). The initial set of sentinel sites was selected from a pool of 1,951 volunteer sites identified during the door-to-door water distribution; in particular it included all 156 sites with lead or lead combination service lines according to City records (Sentinel Site Selection, 2016). Other sites were added according to the following criteria: i) spatial distribution to ensure coverage of all nine wards, ii) areas predicted to have high blood levels based on Hanna-Attisha *et al.* (2016), and iii) environmental justice considerations, specifically lead paint indicators, minority population, and low income derived from EPA Environmental Justice Mapping and Screening Tool, known as EJSCREEN (US EPA, 2016b). This initial set evolved between sampling rounds as some residents stopped participating while others asked to be included in the network (Bryce Feighner, personal communication, February 2, 2017). Once again, samples were collected by homeowners although after training by a sentinel team. All water samples collected during the sentinel and voluntary residential sampling programs were tested for lead and copper by MDEQ Drinking Water Analysis Laboratory, and in the spirit of transparency, results have been posted periodically at <http://www.michigan.gov/flintwater>.

State officials relied on the sentinel data to demonstrate the steady improvement in water quality since the source water switch. In particular, the vast majority of the sentinel

properties (i.e., well over 90%) were found to be at or below the EPA action level (Calley, 2016). A recent analysis of the data collected at non-sentinel sites during the same time period revealed, however, an actual increase in lead levels above 15 µg/L, averaging at some point twice the percentages reported by the sentinel program (Goovaerts, 2017). Despite the lack of control on the selection of non-sentinel sites, one should expect both sampling programs to share the same objective of characterizing WLL in Flint housing stock in general. A legitimate question is thus: whether sampling bias could be the culprit for such opposite trends.

Recent analyses of Flint WLL data alluded to the potential lack of representativeness of sentinel sites. For example, fewer pre-1940 houses were sampled by the sentinel program compared to non-sentinel sites and Flint housing stock, while the reverse trend was observed for LSL (Goovaerts, 2017). Another finding was that results of the two sampling programs differed the most in wards with the greatest percentage of habitants living below the poverty line, which also turned out to be less densely sampled than the least disadvantaged wards. Similarly, Abernethy and Schwartz (2016) found that lower-value homes in Flint tend to be those with the lowest rates of water sampling.

The main objective of this paper is to compare the housing characteristics and geographical distribution of residences sampled by the voluntary and sentinel sampling programs in the aftermath of the Flint drinking water crisis. Unlike the temporal trend analysis described in Goovaerts (2017), the focus is here on a shorter period (2/16/2016–4/15/2016) when the sentinel sampling program aimed to assess the general health of the distribution system before targeting high-risk areas in an extended phase (Calley, 2016). Another difference is the study of relationships among all putative factors (presence of LSL, construction year, poverty level) and water lead levels through frequency analysis, followed by a ward-level exploration of potential sampling bias with particular attention to poverty level and percentages of children with elevated blood lead levels (EBLL) as reported in Hanna-Attisha *et al.* (2016). Last, 2010 census tract poverty levels analyzed in Goovaerts (2017) are replaced by 2015 block group values, as these more recent and precise estimates were used when designing the sentinel sampling network with the help of EJSCREEN software.

2. Data Sources and Methods

2.1 Datasets

The database downloaded from <http://www.michigan.gov/flintwater> and described in Goovaerts (2017) was used for studying the two sampling designs. The focus of the present analysis was on residential testing results recorded over the period 2/16/2016–4/15/2016 when the sentinel sampling program took place. This two-month period was segmented into five time intervals based on the dates for the first and last measurement within each of the five sentinel sampling rounds (S1–S5). Limits of time intervals were adjusted whenever there was a gap of a few days between the end of a sentinel sampling round and the beginning of the next one (Table 1). Data collected at non-sentinel sites were then allocated to one of the five time intervals to facilitate the comparison of results of both sampling programs.

At 204 non-sentinel sites more than one water sample was collected on the same day; e.g., typically multiple taps are being sampled in houses for sale. To avoid allocating too much weight to these replicates each of these observations was assigned a weight equal to one divided by the number of repeated samples on that day. A similar approach was applied to 45 sentinel sites where more than one sample was collected within the same sampling round. The final datasets include 3,123 and 4,645 WLL data collected at 759 sentinel sites and 4041 non-sentinel sites, respectively.

The set of all 51,045 residential tax parcels located within the City of Flint is viewed as the population of interest. Lead in drinking water mainly comes from lead-based solder and lead-containing plumbing fixtures (premise plumbing) in addition to lead service lines bringing water from street main water breaks to the property (Lee *et al.*, 1989; Cartier *et al.*, 2011; Clark *et al.*, 2015). Plumbing material is usually related to the installation year of a plumbing system, which can be approximated by the year of construction. For example, lead service lines were mostly installed before the 1930s while most faucets purchased prior to 1997 were constructed of brass or chrome-plated brass containing up to 8 percent lead (Rabin, 2008; US EPA, 2006). Poor workmanship as well as lack of regular maintenance can also lead to more corrosion and leaching, and the presence of lead particulates, such as disintegrating brass or detaching pieces of old solder (Wang *et al.*, 2014). A representative sample would thus be expected to reproduce the main housing characteristics suspected to influence WLLs, such as presence of LSL, construction year, or socio-economic status of residents. Finally, measurements should be uniformly distributed within the city boundaries to account for any other putative factors likely to be spatially structured (e.g., water travel time between the treatment plant and home plumbing system). For example, during their investigation of the Washington, D.C. drinking water crisis, Edwards *et al.* (2009) found that the relative risk of exposure to high lead in water was a strong function of zip code, resulting in “hot spots” neighborhoods. The spatial coverage of WLL data was here assessed using the percentage of data collected within each of the nine city wards since these geographical units were used in the seminal paper on children EBLL that triggered the emergency response (Hanna-Attisha *et al.*, 2016).

The following housing characteristics described in detail in Goovaerts (2017) were derived for each sampling site: 1) the type of service line (SL) retrieved from a digital map of Flint’s lead water pipe, and 2) the year the house was built. For the sentinel network, on-site data on the composition of service lines were also collected by a plumber during the sentinel team visit. Socio-economic status was assessed using 2015 ACS (American Community Survey) 5-year estimates of the percentage of the block group population living in households where the income is less than or equal to twice the federal “poverty level.” The rationale for using twice the poverty threshold rather than just the poverty threshold is listed in Appendix B of EJSCREEN technical documentation (US EPA, 2016b); in particular the fact that today’s poverty thresholds are too low to adequately capture the populations adversely affected by low income levels. These data were downloaded from https://factfinder.census.gov/faces/nav/jsf/pages/download_center.xhtml (variable ACS_15_5YR_C17002). The last variable used in the study was the ward-level percentage of elevated blood lead levels (i.e., blood level > 5 µg/dL) recorded in children between 1/1/2015 and 9/15/2015, that is after the

water source change from Detroit-supplied Lake Huron water to the Flint River (Table 2 in Hanna-Attisha *et al.*, 2016).

2.2 Frequency analysis of housing characteristics

The visualization and comparison of housing characteristics for the reference population (Flint housing stock) and the two sample sets (sentinel and non-sentinel sites) were conducted using frequency analysis and kernel smoothing. Let N be the number of residences, denoted by their centroid's geographical coordinates $\mathbf{u}_\alpha$, within any given database, while $n(\mathbf{u}_\alpha)$ is the number of times a residence α has been visited over the two-month study period ($n(\mathbf{u}_\alpha)=1$ for the reference population). The total number of WLL data, denoted N' , is thus $\sum_{\alpha=1}^N n(\mathbf{u}_\alpha)$. Each residence is characterized by its built year, $BY(\mathbf{u}_\alpha)$, the type of service line, $SL(\mathbf{u}_\alpha)$, and the block group poverty level, $GP(\mathbf{u}_\alpha)$.

The frequency distribution for the two continuous variables (i.e., built year, poverty level) was constructed using a rectangular kernel of size 11 (i.e., each observation within the window of size 11 receives the same weight). For example, the sampling frequency of residences built in year y_1 was computed as:

$$f(BY=y_1) = \sum_{\alpha=1}^N n(\mathbf{u}_\alpha) \sum_{i=-5}^5 i_{BY}(\mathbf{u}_\alpha; y_1+i) \quad (1)$$

where $i_{BY}(\mathbf{u}_\alpha; y_1+i) = 1$ if $BY(\mathbf{u}_\alpha) = y_1+i$ and zero otherwise. In other words, the number of residences built in 1935 that were sampled was calculated as the number of WLL samples collected in houses built between 1930 and 1940. The choice of the kernel width (11 years) was somewhat subjective and aimed to strike a balance between a too wide window causing information loss and a too narrow window resulting in unreliable estimates. Results based on fewer than 50 observations were discarded. A similar formula was used for poverty level (width = 11%). Relative frequencies were then derived by dividing quantity (1) by $11 \times N'$.

The relationship between any two housing characteristics was explored using conditional frequencies. For example, the relative frequency of lead service lines within houses built in year y_1 was computed as:

$$f(SL=\text{lead} | BY=y_1) = \frac{1}{N_{y_1}} \sum_{\alpha=1}^N \sum_{i=-5}^5 i_{BY}(\mathbf{u}_\alpha; y_1+i) \times i_{SL}(\mathbf{u}_\alpha; \text{lead}) \quad (2)$$

where $i_{SL}(\mathbf{u}_\alpha; \text{lead}) = 1$ if $SL(\mathbf{u}_\alpha) = \text{lead}$, and zero otherwise. The denominator N_{y_1} is computed as:

$$N_{y_1} = \sum_{\alpha=1}^N \sum_{i=-5}^5 i_{BY}(\mathbf{u}_\alpha; y_1+i) \quad (3)$$

Note that unlike for sampling frequencies (Eq. 1) each sampled residence is used only once for the computation of conditional frequencies, i.e., $n(\mathbf{u}_\alpha)=1$ (replicates are not used). Once again, results based on fewer than 50 observations were discarded.

2.3 Conditional analysis of water lead levels

The analysis started with the coding of each WLL data $(\mathbf{u}_\alpha; S_k)$, collected at residence $\mathbf{u}_\alpha$ during sampling round S_k , into an indicator of being greater or not than the threshold $z_c = 15$ $\mu\text{g/L}$:

$$i_k(\mathbf{u}_\alpha; z_c) = \begin{cases} 1 & \text{if } z(\mathbf{u}_\alpha; S_k) > z_c \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

To account for temporal changes in lead levels over the two-month period (Goovaerts, 2017) and their strongly positively skewed distribution, the percentile of each WLL data $z(\mathbf{u}_\alpha; S_k)$ within the corresponding sampling round S_k was also computed as:

$$p(\mathbf{u}_\alpha; S_k) = F_k^{-1}(z(\mathbf{u}_\alpha; S_k)) \quad (5)$$

where $F_k(\cdot)$ is the cumulative distribution function of WLL data for the k -th sampling round. For example, $p(\mathbf{u}_\alpha; S_k) = 0.75$ indicates that the level measured at residence $\mathbf{u}_\alpha$ was greater than 75% of WLL data collected during that sampling round.

The impact of housing characteristics on water lead levels was then assessed using conditional frequencies similar to the ones described in Equation (2). For example, the impact of construction year on the proportion of WLL above $z_c = 15$ $\mu\text{g/L}$ or the relative magnitude of WLL as measured by percentiles (Eq. 5) was explored using the following relative frequencies:

$$f(z > z_c | BY = y_1) = \frac{1}{N'_{y_1}} \sum_{\alpha=1}^N \sum_{k=1}^5 \sum_{i=-5}^5 i_{BY}(\mathbf{u}_\alpha; y_1 + i) \times i_k(\mathbf{u}_\alpha; z_c) \times \delta_{\alpha k} \quad (6)$$

$$f(p | BY = y_1) = \frac{1}{N'_{y_1}} \sum_{\alpha=1}^N \sum_{k=1}^5 \sum_{i=-5}^5 i_{BY}(\mathbf{u}_\alpha; y_1 + i) \times p(\mathbf{u}_\alpha; S_k) \times \delta_{\alpha k} \quad (7)$$

where N'_{y_1} is the number of WLL data collected in houses built during the time period $[y_1 - 5, y_1 + 5]$. $\delta_{\alpha k} = 1$ if residence α was sampled during k -th round, and zero otherwise. Similar expressions were used to investigate the impact of block group poverty levels on WLL. In all cases, results based on fewer than 50 observations were discarded.

3. Results and Discussion

3.1 Sampling design

Fig. 1A shows the location of all residential parcels (759 sentinel sites and 4,041 non-sentinel sites) that were sampled over the two-month period. Despite the much smaller number of sentinel sites (15.8%), their repeated sampling leads to a more balanced proportion of WLL data relative to non-sentinel sites (41.3%); see Table 1. To visualize the relative distribution of sentinel versus non-sentinel sites, indicators of presence/absence of sentinel sites were created for all 4,800 sampled locations before being interpolated using kriging (Goovaerts, 1997). The map of indicator kriging estimates (Fig. 1B) illustrates the spatial clustering of sentinel sites, in particular along the boundary that Ward 2 shares with wards 3 and 6. On the other hand, much fewer sentinel sites relative to non-sentinel sites are located in wards 1, 3, 4 and 5.

The four variables used to explore any potential sampling bias are mapped in Figs. 1C–F. Information is available at three different geographies: 1) tax parcel units for housing characteristics: (composition of service line, construction year), 2) block groups (BG) for percentages of families living below twice the poverty line in 2015, and 3) Flint wards for percentage of EBLI recorded in children the first 9 months of 2015. Visual comparison of these maps already yields some interesting findings (see Section 3.4 for a more quantitative analysis):

- Ward 5, which includes the largest percentage of children with EBLI (15.7%, Table 4) and a large proportion of houses built before 1935 (82.09%, Table 5), has a very low density of sentinel sites (5.01%).
- According to Hanna-Attisha *et al.* (2016), the area of intersection between wards 3, 4, and 5 (in the east side of the city) also appeared to have high EBLI. This area circled in Fig. 1C is characterized by older homes, higher poverty, and very few sentinel sites.
- Compared to other poor areas of the city, Ward 1 has houses that are much more recent (i.e., only 13.71% of pre-1935 houses, Table 5).
- Main clusters of sentinel sites appear to be preferentially located in areas of moderate to lower poverty.

3.2 Temporal changes

Although none of the four covariates displayed in Fig. 1C–D changed over the two-month study period, the location of sampling sites did, and the temporal changes in the characteristics of sampled residences is summarized in Fig. 2. The time series of percentages of WLL above 15 $\mu\text{g/L}$ (see Table 1 for precise numbers) highlight the widening gap between results of the two sampling programs (Fig. 2A). Not only does the voluntary sampling program indicate increasing levels of lead over time, it also reflects the failure to meet the LCR action level of 10% (horizontal dashed line) at any moment during the two-month period. On the other hand, according to the sentinel monitoring program water quality

has been steadily improving and the threshold of 10% of WLL data above 15 µg/L was never exceeded.

Given the lack of control on the voluntary sampling program a legitimate concern is that homeowners who found high levels of lead in their first samples would be prone to acquire additional testing kits and repeat the sampling within the two month period. In addition, whenever a test result exceeded 15 µg/L, the residential testing procedure was to offer a follow-up test to see if levels were coming down and if remediative efforts were working (Testing Plan, Process & Protocols, 2016). On the other hand houses with low lead levels would be less likely to be tested again, leading over time to a biased selection of houses and an inflated percentage of WLL above 15 µg/L.

This potential temporal bias was investigated by identifying for each sampling round the sites that were visited for the first time or sampled repeatedly. As expected, the proportion of first time visits has declined over time, in particular for the sentinel sampling program which aims to track temporal changes at fixed sites. For example, Table 2 indicates that 80% of sites sampled voluntarily in round S5 (4/6/2016–4/15/2016) had never been tested in rounds S1–S4, while this was true only for 1.5% of sentinel sites. Interestingly, WLL exceeded 15 µg/L at a much lower rate during these first time visits compared to residences that had been sampled previously; note, however, that some statistics can be unreliable for a small number of homes; e.g. the 27.1% recorded at round S4 is based on 42 houses. Conversely, the sentinel sites visited for the first time in rounds S2–S4 had higher lead levels, which confirms the targeting of houses at risk as illustrated by the increasing percentages of sentinel sites with lead service lines (Fig. 2B). Excluding repeated measurements at non-sentinel sites narrowed the gap between the two times series (Table 2, columns and Fig. 7B), yet it did not alter temporal trends: the percentage of WLL recorded above 15 µg/L at non-sentinel sites is still increasing for rounds S1 through S4.

Fig. 2B shows that WLL and percentages of sentinel sites with lead service lines have moved in opposite directions over the five sampling rounds. Indeed, the sentinel program has sampled an increasing number of houses with LSL, a trend that appears even stronger when using the more accurate on-site data on composition of service lines instead of the digital data. As expected the voluntary sampling data more closely reproduce the percentage of LSL found in Flint housing stock (7.32%), which is lower than the results of preferential sampling by the sentinel program. Lead service lines are widely considered the main source of lead in drinking water. The fact that WLLs have exceeded 15 µg/L at fewer and fewer sentinel sites, even though these sites have increasingly included LSL, could thus be viewed as the sign of continuous improvement in water quality. Yet, lead levels recorded at non-sentinel sites have increased despite the lower sampling frequency of lead service lines.

Another source of lead in drinking water is lead fixtures and pipes present within old houses (premises plumbing). In comparison to LSL temporal trends in construction year of sampled houses better match WLL changes over the five sampling rounds (Figs. 2C,D). The sentinel program has been sampling fewer and fewer pre-1935 houses, while houses built between 1935 and 1950 became an increasingly large fraction of the sampling pool. Pre-1935 houses have also been under-sampled by the voluntary program (average = 42.8% in Flint housing

stock), yet the sampling deficit has decreased over time. On the other hand, the share of 1935–1950 houses in Flint housing stock is well reproduced in the voluntary sampling and is half the percentage observed in the State-controlled monitoring plan. There is no official reason for this sampling bias by the sentinel program as construction year was not one of the selection criteria. Percentage of housing units built before 1960 is, however, used as an indicator of potential exposure to lead paint in EJSCREEN software (US EPA, 2016b).

The last covariate, which might indirectly inform on housing condition, including quality of premise plumbing, is socio-economic status which was here assessed using block group percentages of habitants living below twice the poverty line. According to Figs. 2 E,F both sampling programs have over-sampled houses located in the least disadvantaged block groups (poverty level < 55%), while block groups with poverty levels above 75% have been under-sampled. These results indicate that citizens living in the most impoverished areas used fewer testing kits for voluntary sampling. Interestingly, the sentinel sampling program led to very similar statistics despite a very different selection procedure. One hypothesis is that as poverty level increased citizens were less likely to volunteer to be part of the sentinel monitoring network.

3.2 Sampling bias: housing characteristics

The relative frequency distributions in Figs. 3A,B illustrate the limitations of relying on average statistics on construction year and BG poverty level (e.g., means plotted in Fig. 2) to compare residences within the Flint housing stock and the two sample sets. In particular, the histogram for construction year is clearly bimodal. For the reference population (Fig. 3A, black solid line), the two modes are years 1927 and 1955. The first mode is well reproduced by both sampling distributions, albeit with a slightly smaller frequency. On the other hand, the sentinel sampling program oversampled houses built around 1945 (Fig. 3A, red dashed line). The second mode for the voluntary program coincides with Flint housing stock value, although at a higher frequency (Fig. 3A, green dashed line).

Discrepancies between the three frequency distributions are larger for poverty level. Both sentinel and voluntary programs over-sampled houses in the least disadvantaged block groups (35–50% poverty level) while block groups with poverty level in the 70–95% range were under-sampled, a fact already stressed by the time series in Figs. 2 E,F. The bias is, however, larger for the sentinel program which appears to have sampled uniformly block groups with poverty levels ranging between 40% and 70%; compare black solid line to red dashed line in Fig. 3B.

An important feature controlling water lead levels is the presence of lead service lines. Fig. 3C (black solid line) shows that LSL tend to be more frequent in older homes; the increase observed for post-2000 houses can be disregarded as it is based on a very small percentage of Flint housing stock (0.6%). Despite the lack of control on site selection, the voluntary sampling program captures very closely the relationship between construction year and presence of LSL. Plotting these conditional frequencies for the sentinel program reveals a very strong sampling bias, which could not be detected by looking at individual statistics for construction year (Fig. 3A) and percentage of LSL (Fig. 2B). Indeed, the vast majority of

sentinel sites with LSL were built between 1935 and 1950. Using the more accurate on-site data accentuates the magnitude of the bias.

The bias attached to the sampling of LSL by the sentinel program extends to poverty level (Fig. 3D). Because older homes tend to be located in poorer neighborhoods (Fig. 3E) the percentage of houses with lead service lines tends to increase with poverty level (black solid line). The sentinel program has over-sampled LSL in the block groups with poverty levels ranging between 40% and 70%, which is consistent with the interpretation of Fig. 3B. Sites sampled by the voluntary program more closely mimic the characteristic of Flint housing stock, including the largest percentage of LSL in block groups with poverty levels above 90%, a fact that was not captured by the sentinel program.

The last conditional frequency distribution (Fig. 3E, black solid curve) reflects the larger frequency of older houses (pre-1915 construction year) in poorer neighborhoods, as well as the existence of public housing complexes built post-1970. Results for the sentinel program (lower red dashed curve) confirm the under-sampling of lower income block groups, in particular for houses built between 1935 and 1955, which overlaps with the sampled housing segment with LSL. The voluntary program displays a similar sampling bias. Note that the curve for the sentinel program ends around year 1965 as only 21 sentinel sites (2.77%) were post-1965 constructions, while the voluntary sampling pool included 363 homes (8.98%) built after 1965.

3.3 Impact of housing characteristics on water lead levels

Kernel smoothing was used to explore the potentially non-linear impact of built year and poverty level on the magnitude of water lead levels (i.e., sampling round percentile) and likelihood of exceeding two thresholds: 1 $\mu\text{g/L}$ and the action level of 15 $\mu\text{g/L}$. The larger the deviation from the 50th percentile (median), depicted by the horizontal dashed line in Figs. 4A,B, the greater the average impact of housing characteristics. Discrepancies between results of both sampling programs are the most important for construction year: percentiles and percentages of data above 1 $\mu\text{g/L}$ are larger at voluntarily sampled houses built prior to 1935 and smaller post 1955; see Fig. 4 (left column). The trend beyond 1975 can be ignored as the voluntary sampling pool includes only 99 homes (2.45%) built after 1975.

Interestingly construction year exerts opposite effects on the percentage of WLL above 15 $\mu\text{g/L}$ recorded at the two types of sampling sites: This percentage displays an expected decline for newer homes that were sampled voluntarily, while it increases at sentinel sites. Despite the oversampling of LSL at sentinel sites built between 1935 and 1955 (Fig. 3C) the percentage of data above 15 $\mu\text{g/L}$ is still smaller than at non-sentinel sites. These results suggest that higher WLL originate from lead fixtures and pipes present within old houses (premise plumbing) as opposed to LSL. This confirms earlier findings that home lead service lines may not be the largest contributor of lead in Flint, and lead contamination may be caused by interior plumbing (Dolan, 2016; Goovaerts, 2017).

Poverty level appears to have little impact on WLL recorded within the voluntary sampling program: most curves in Fig. 4 (right column, green curves) are flat; the steep increase observed for poverty level below 30% (Fig. 4F) can be disregarded as these frequencies are

based on only 2.41% of samples (112 observations). The percentile and percentage curves display larger fluctuations for the sentinel program (Fig. 4, right column, red curves), which is caused by strong disparities in percentage of lead service lines sampled over different classes of poverty level (Fig. 3D). In particular the oversampling of houses with LSL for the 55–70% poverty range (Fig. 3B) results in larger values (Fig. 4, right column). As for construction year, percentile curves that account for the whole range of the data instead of focusing on a specific threshold, in particular 15 µg/L, overlap over mid-range values (i.e. poverty from 40 to 60%), indicating that WLLs are not systematically higher at non-sentinel sites. This is confirmed by the fact that WLL exceeds 1 µg/L at more sentinel sites than non-sentinel sites (horizontal dashed lines, Figs. 4C,D), while the reverse is true for 15 µg/L (horizontal dashed lines, Figs. 4E,F).

This graphical interpretation was supplemented by a regression analysis to predict the probability of exceeding three thresholds (1, 15, and 25 µg/L) on the basis of two housing characteristics (type of service line, year of construction) and block group poverty level. Because sentinel sites in particular have repeated samples, Generalized Estimating Equations (GEE) regression (Liang and Zeger, 1986) with logit link function and exchangeable correlation structure was used to fit this model (SAS Institute Inc, 2011). The impact of each covariate was quantified using the odds ratio (OR) which represents the odds that the outcome (exceedance of WLL threshold) will occur given a particular event (e.g. presence of LSL), compared to the odds of the outcome occurring in the absence of that event (e.g., service lines composed of unknown or other than lead material).

For both sampling programs the impact of housing characteristics and poverty, as measured by odds ratios, increases for lower thresholds (Table 3). For example, the presence of lead service lines doubles and quadruples the likelihood of measuring WLL above 1 µg/L at voluntary (OR=2.06) and sentinel sites (OR=3.78), respectively. For the former the impact of construction year is even higher, in particular when comparing pre-1935 houses to post-1950 houses (OR=3.22). Odds ratios are also highly significant for all thresholds ($\alpha=0.01$), reflecting the greater influence of construction year relative to LSL. On the other hand, the impact of construction year is lower at sentinel sites (OR=2.01) and is only significant at $\alpha=0.05$ for 1935–1950 houses (OR=1.39). Since the vast majority of sentinel sites with LSL were built between 1935 and 1950, the impact of construction year and LSL cannot be disentangled and this might explain a lower OR for construction year being compensated by a higher OR for LSL. Note that similar results were obtained when using the more accurate LSL on-site records instead of digital data, which invalidates the hypothesis that the larger impact of built year vs LSL at non-sentinel sites simply reflects the larger accuracy of construction year vs digital data on the presence of lead service lines. This is also confirmed by the fact that a regression model using only LSL as covariate led to higher OR for all three thresholds (i.e. 2.56 vs 2.06, 1.80 vs 1.53, and 1.34 vs 1.13) relative to the case where all covariates were included.

Discrepancies between the two types of sampling widen as the threshold increases. In particular, the predictive power of construction year decreases substantially at sentinel sites: the odds ratio becomes smaller than 1 and non-significant for WLL thresholds of 15 and 25 µg/L (Table 3). This result is consistent with the interpretation of Figs. 4C,E: The gap

between the two conditional frequency curves increases for higher thresholds and older construction years, while the impact of construction year becomes negative for the percentage of data above 15 µg/L.

According to odds ratios the likelihood of exceeding WLL thresholds at sentinel sites increases with poverty level ($OR > 1$). Note that ORs are the largest for a threshold of 1 µg/L, and results are significant only for the 55–75% poverty category which includes most sentinel sites with LSL (Fig. 3D). As already illustrated by frequency curves in Fig. 4 (right column), poverty level has little impact on WLL recorded within the voluntary sampling program.

3.4 Ward-level analysis

The frequency analysis in Sections 3.2–3.3 was aspatial in that sample locations were ignored during the interpretation. The description of spatial patterns in Fig. 1 was supplemented by the computation of ward-level statistics regarding sampling density, water and blood lead levels, housing characteristics, and poverty levels (Tables 4&5).

Each ward represents between 8.76% (Ward 7) and 14.24% (Ward 3) of residential parcels in the city of Flint. This range is much wider for the sentinel program (5.18–20.35%), which reflects the oversampling of specific areas in the city despite the initial aim to cover adequately all nine wards (Sentinel Site Selection, 2016). The voluntary sampling program yields an intermediate range: 6.91% to 15.26%. Interestingly, Ward 7, which includes the smallest percentage of residential parcels in the city, was the most frequently sampled by both programs (Table 4). One explanation is that Ward 7 had the largest number (41 out of 148) of sites with lead or lead combination service lines within the initial pool of 1,951 prospective sentinel sites. The second largest number of LSL (21 sites) within that pool was recorded for Ward 6 which was the third most densely sampled ward (15.11%) by the sentinel program. In addition wards 6 and 7 included, respectively, the second and third largest percentages of children with elevated blood lead levels (Table 4). These sentinel sites tend, however, to be clustered in the least disadvantaged part of Ward 7, see circled area in Fig. 1E. In fact, there is a general tendency for the sentinel program, and to a lesser extent the voluntary sampling program, to collect more WLL data in wards where fewer habitants live below the poverty line (Fig. 5A). The steady decline observed for the voluntary sampling program is particularly noteworthy. The trend displayed by both sampling programs (dashed lines, Fig. 5A) is opposite to what is observed for Flint housing stock: more residential parcels exist in wards with higher poverty (solid line, Fig. 5A). These results indicate that citizens living in the most impoverished wards have used fewer testing kits for voluntary sampling, and they might have been less likely to request their inclusion in the sentinel sampling program after it started. This could explain why Ward 5, which has the second largest poverty level (73.24%) among all nine wards, was the least densely sampled by the sentinel program (5.01%) despite featuring by far the largest percentage of elevated blood lead levels (15.7%).

This socio-economic bias is not limited to the number of monitoring sites but extends to the characteristics of these sites. Indeed, Fig. 5B shows that relative to Flint housing stock (solid line) the sentinel program has under-sampled lead service lines in the two wards with

poverty levels above 70% (red dashed line) while these same wards were over-sampled in the voluntary sampling program. In particular, the sentinel program did not sample any house with LSL in Ward 5, although the percentages of lead service lines (9.95%, Table 5) and elevated blood lead levels (Table 4) were the largest in Flint. One potential culprit is the fact that 82% of houses in Ward 5 were built before 1935 (Table 5), whereas the sentinel program has almost exclusively sampled LSL in houses built between 1935 and 1950 (Fig. 3C). The other case is Ward 3, which has the highest poverty level (74.60%) and 8.56% of houses with LSL according to digital data. Although the sentinel sampling program targeted the poorest block groups in that ward (92.13% of sentinel sites are in block groups with poverty level > 75%), the sampling rate for LSL was only 3.24% compared to 10.63% for the voluntary program. Ward 1 is the second ward with no LSL sampled by the sentinel program while having the 4th largest poverty level (66.33%). However, as noticed on Fig. 1C, the majority of houses in this ward were built after 1935 (87%) and only 3.27% of houses have LSL.

The largest percentage of existing LSL sampled by the sentinel program is recorded in Ward 9: 24.06% of sentinel sites there possessed lead service lines, which is close to four times the percentage of houses with LSL (7.07%) in that ward (Table 5), a number well captured by voluntary sampling (8.70%). There is no apparent reason for such an over-sampling as this ward is average for all statistics (i.e. poverty level, percentage of EBL, and older homes) and had one of the lowest number of houses with lead or lead combination service lines (12 out of 148) within the initial pool of 1,951 prospective sentinel sites. Sentinel sites with LSL are, however, clustered in Ward 9 (Fig. 5E, dashed circles), which suggests that practical convenience might have been the culprit for the sampling of such a large number of lead service lines in that ward.

One consequence of such socio-economic sampling bias is the tendency for the most impoverished wards to report fewer WLL data above 15 µg/L at sentinel sites (Fig. 5C), hence, the widening gap between both sampling programs as the poverty level increases (Fig. 5D). Although Goovaerts (2017) had already noticed the largest ratios between percentages of WLL data > 15 µg/L recorded at non-sentinel sites vs sentinel sites in impoverished wards 1 and 5, the link with the biased sampling of lead service lines (no LSL sampled in both wards, Table 5) is a new finding. It is noteworthy that for all nine wards the percentage of WLL data above 15 µg/L is systematically smaller for the State-controlled sampling program compared to the homeowner-driven sampling (Fig. 5C).

3.5 Correcting for sampling bias

To correct for the sampling bias detected in Sections 3.1–3.4, the time series of percentages of WLL data above 15 µg/L (Fig. 2A) were standardized using the procedure described in Goovaerts (2017). Standardization is a common approach for controlling confounding in population studies or data from disease registries (Waller and Gotway, 2004). It is defined as a weighted average of stratum-specific rates. In the present case-study, Flint's 51,045 residential parcels and the WLL data collected by both sampling programs were stratified on the basis of four covariates: 1) presence/absence of lead service lines, 2) construction year (3 classes: pre-1935, 1935–1950, and post-1950), 3) block group poverty level (<

55%, 45–75%, > 75%), and 4) ward. The choice of categories of construction year and poverty level was guided by results displayed in Figs 3C and D regarding the oversampling of some segments of Flint housing stock by the sentinel program. For example, Fig. 6 (top) shows the stratification of WLL data collected during sampling round 1 at non-sentinel sites on the basis of the first two covariates. The percentage of data above 15 µg/L was computed for each of the six categories. As expected, fewer observations exceed the action level in houses without LSL: 9.84% vs 14.5%. Accounting for construction year indicates however that the presence of LSL has no impact on results for pre-1935 houses (15.7% in both cases) while only half that percentage (7.23%) exceed 15 µg/L in post-1950 houses without LSL.

Residences tested during the first round of voluntary sampling represent almost perfectly Flint housing stock when it comes to presence of lead service lines: 7.69% vs 7.32% (Fig. 6). However, as discussed previously (Fig. 2C), older houses were under-sampled by both sampling programs; e.g., for pre-1935 houses: 3.56% vs 4.46% (LSL) and 25.4% vs 38.38% (no LSL). Similarly, houses built more recently (i.e., post-1950) were over-sampled: 1.55% vs 0.80% (LSL) and 49.3% vs 33.86% (no LSL). After correction for this sampling bias, the percentage above 15 µg/L increased from 10.20% to 11.35% (Fig. 6) since the under-sampled older houses were the ones with the greatest water lead levels while lower WLLs were measured in over-sampled post-1950 houses.

A similar correction was conducted for all five rounds of both sampling programs, using between one and four covariates (Fig. 7). This correction was applied after adjusting the time series of the voluntary test results for temporal bias by eliminating repeated measurements (Table 2), lowering percentages above 15 µg/L in particular for the last two sampling rounds (Fig. 7A). Adjusting for the presence of lead service lines had a negligible impact (Fig. 7B) since the percentage of residences with LSL is still small even after over-sampling (Fig. 2B). Adding construction year as a covariate caused bigger changes (Fig. 7C), in particular for the voluntary sampling program since, according to the odds ratio (Table 3), this covariate has a greater influence on WLL. Interestingly, the adjusted percentage of WLLs recorded above 15 µg/L during the first sampling round is greater at sentinel sites compared to non-sentinel sites. However, the fact that only 6.7% of sentinel sites with LSL were built before 1935 or after 1950 hampered the reliable estimation of WLLs for these strata. In comparison, these strata represent 71.84% in Flint housing stock: $(2277+407)/3736$ in Fig. 6.

Incorporating BG poverty into the set of confounding factors slightly narrowed the gap between the two time series (Fig. 7D). The impact of the socio-economic sampling bias was, however, attenuated by the small influence exerted by poverty level on WLL data (see odds ratios in Table 3). After adjusting for all four covariates, the time series for the sentinel program displays the largest decrease over the five sampling rounds (Fig. 7E). This correction had the opposite impact on the other time series as the percentage of voluntary tests above 15 µg/L increased. Although the bias-correction procedure narrowed the gap between results of both sampling programs, the time series still differ during the latest sampling rounds, even after accounting for the uncertainty attached to these estimates; see 90% confidence intervals in Fig. 7E

4. Conclusions

It is common in environmental studies to rely on sampling to characterize populations that are too large to be measured exhaustively; in other words a measurement cannot be collected at every location in the spatial domain (Myers, 1997). When designing sampling schemes one should keep in mind the objectives of the study (i.e., which questions are we trying to answer?), as well as the definition of the population to be studied. Attention to sampling design is particularly critical when analyzing data collected by others and sparsely documented, as in the case of Flint sentinel monitoring network. The analyst should thus question whether that sample is representative of the underlying population and explore ways to correct sample statistics whenever biased or preferential sampling is suspected.

The delay in reporting high levels of lead in Flint drinking water and the resulting extent of the ensuing environmental crisis and public health threat were partially caused by the biased selection of sampling sites. Therefore, one could have expected a greater transparency in the selection of sites for monitoring post-crisis water lead levels. For example, despite the frequent posting of sentinel testing results, addresses of sentinel sites never included street numbers, vital information that had to be retrieved through data mining (Goovaerts, 2017). This paper presented the first detailed analysis of the monitoring network put in place by the State in the aftermath of the Flint water drinking crisis. Results of an exploratory spatial data analysis were combined with on-line descriptions of the sentinel sampling plan (e.g., Testing Plan, Process & Protocols, 2016; Sentinel Site Selection, 2016), and personal communication from the director of Drinking Water and Municipal Assistance Division (DWMAD) at the Michigan Department of Environmental Quality.

A key finding of the exploratory data analysis was that the impact of lead service lines on water lead levels was mainly investigated for houses built between 1935 and 1950 in less disadvantaged areas of the city. In addition, there was no sentinel site with LSL in two of the most impoverished wards, including where the percentage of children with high blood lead levels tripled following the switch in water supply. Such bias seems surprising as socio-economic status and predicted blood lead levels were some of the criteria used during the selection of sentinel sites. One must, however, keep in mind that the network includes only sites where homeowners volunteered to participate in the bi-weekly sampling. Also, the sentinel network evolved between sampling rounds as some residents stopped participating, while others asked to be included, which might have created clusters of sentinel sites in areas of higher socio-economic status. Indeed, analysis of the voluntary sampling program indicated the tendency of citizens living in the most impoverished wards to use fewer testing kits. The oversampling of lead service lines for houses built between 1935 and 1950 is more puzzling and is a major drawback as it hampers our ability to disentangle the effects of LSL and premise plumbing (lead fixtures and pipes present within old houses) on WLL.

Despite the lack of sampling strategy, voluntary testing turned out to capture the main characteristics (i.e., presence of lead service lines, construction year) of Flint housing stock much more closely than the sentinel program. A sampling bias was, however, detected as homeowners who found high levels of lead in their first samples were more likely to acquire additional testing kits, while houses with low lead levels were less likely to be tested again.

Over time this led to an inflated percentage of water lead levels that tested 15 µg/L and higher. Correcting this bias narrowed the gap between results of the sentinel and voluntary sampling programs. Yet, even after adjusting for other covariates, such as housing characteristics and socio-economic status, lead levels measured at sentinel sites in sampling rounds 4 and 5 still exceed 15 µg/L at a statistically significant lower rate than samples collected voluntarily. Temporal trends also remained drastically different: The percentage of sentinel data above 15 µg/L decreased steadily over time, while it increased for data collected on a voluntary basis during most of the two-month period.

Caveats of the analysis include the uncertainty attached to the digital data on composition of service lines and the sensitivity of the results to the type (e.g., household vs individual income) and resolution (i.e., census tract vs block group) of poverty estimates. It is noteworthy that important results for the sentinel sampling program, in particular the lack of sampled lead service lines in two wards, did not change when using the more accurate on-site LSL data. On the other hand, relying on more accurate and recent block group poverty estimates, similar to the ones used by the State when designing the sentinel network, greatly attenuated the socio-economic bias detected in Goovaerts (2017). Our analysis, however, confirmed earlier findings (Dolan, 2016; Goovaerts, 2017) that home lead service lines may not be the largest contributor of lead in Flint, and lead contamination may be caused by interior plumbing. This would explain why water collected in March 2016 in a residence without a lead service line tested at 1,000 µg/L (Johnson, 2016). This information is critical as substantial resources are currently being spent on the replacement of lead service lines in Flint (Moore, 2016) and the selection of lead lines targeted for replacement has been questioned (Ridley, 2016).

Acknowledgments

The author is grateful to Troy Rosencrants from the Department of Earth and Resource Science (University of Michigan – Flint) for providing the shape files with tax parcel information and service line data. This research also greatly benefited from information about Flint and the sentinel sampling program shared by Dr. Rick Sadler from MSU and Mr. Bryce Feighner from MDEQ. The author would like to express his gratitude to Dr. Colleen Vallo for proof-reading the manuscript. This research was funded by grant 1R43CA192520-01A1 from the National Cancer Institute. The views stated in this publication are those of the author and do not necessarily represent the official views of the NCI.

References

- Abernethy, J., Schwartz, E. Flint fiasco calls for heavy-hitting data analysis. GreenBiz. 2016 Sep 28. Retrieved from <https://www.greenbiz.com/article/flint-fiasco-calls-heavy-hitting-data-analysis>
- Calley, B. Understanding Recent Flint Water Test Results. 2016. Available at <https://medium.com/@LtGovCalley/understanding-recent-flint-water-test-results-fbb69cf4b5d5>
- Cartier C, Laroche L, Deshommes E, Nour S, Richard G, Edwards M, Prévost M. Investigating dissolved lead at the tap using various sampling protocols. J Am Water Works Ass. 2011; 103:55–67.
- Clark BN, Masters SV, Edwards MA. Lead release to drinking water from galvanized steel pipe coatings. Environ Eng Sci. 2015; 32(8):713–721. DOI: 10.1089/ees.2015.0073
- Dolan, M. Study: Flint lead contamination goes beyond service pipes. Detroit Free Press; 2016 Sep 8. Available at: <http://www.freep.com/story/news/local/michigan/flint-water-crisis/2016/09/08/study-flint-lead-contamination-goes-beyond-service-pipes/89994636/> Accessed November 23, 2016

- Edwards M, Triantafyllidou S, Best D. Elevated blood lead levels in young children due to lead-contaminated drinking water: Washington, DC, 2001–2004. *Environ Sci Technol.* 2009; 43(5): 1618–1623. [PubMed: 19350944]
- Felton, R. Michigan official suggested gaming water tests to ‘bump out’ lead results. *The Guardian*; 2016 Apr 27. Retrieved from <http://www.theguardian.com>
- Flint Safe Drinking Water Task Force Recommendations on MDEQ’s Draft Sentinel Site Selection. Feb. 2016 Retrieved from https://www.epa.gov/sites/production/files/2016-02/documents/task_force_recommendations_on_sentinel_site_selection_2-16.pdf on August 20, 2016
- Flint Water Advisory Task Force Report. Mar. 2016 Retrieved from https://www.michigan.gov/documents/snyder/FWATF_FINAL_REPORT_21March2016_517805_7.pdf on June 16, 2016
- Goovaerts, P. *Geostatistics for Natural Resources Evaluation*. Oxford University Press; New-York, NY: 1997.
- Goovaerts P. The drinking water contamination crisis in Flint: Modeling temporal trends of lead level since returning to Detroit Water System. *Sci of the Total Environ.* 2017;581–582. 66–79.
- Hanna-Attisha M, LaChance J, Sadler RC, Champney Schnepf A. Elevated blood lead levels in children associated with the Flint drinking water crisis: A spatial analysis of risk and public health response. *Am J Public Health.* 2016; 106:283–290. [PubMed: 26691115]
- Johnson, J. 2016 Mar 11. Removing lead pipes may not solve Flint’s water crisis. Available at: http://www.mlive.com/news/flint/index.ssf/2016/03/removing_lead_pipes_wont_solve.html Accessed November 27, 2016
- Lee RG, William CB, David WC. Lead at the tap: Sources and control. *J Am Water Works Ass.* 1989; 81(7):52–62.
- Liang KY, Zeger SL. *Longitudinal Data Analysis Using Generalized Linear Models*. Biometrika. 1986; 73:13–22.
- Milman, O., Glenza, J. At least 33 US cities used water testing ‘cheats’ over lead concerns. *The Guardian*; 2016 Jun 2. Retrieved from <http://www.theguardian.com>
- Milman, O. Tests on Flint water targeted homes far from network of lead pipes. *The Guardian*; 2016 Jun 2. Retrieved from <http://www.theguardian.com>
- Moore, K. Mayor Weaver kicks off third Phase of FAST start pipe replacement initiative. Oct 31. 2016 Press release Available at <https://www.cityofflint.com/2016/10/31/mayor-weaver-kicks-off-third-phase-of-fast-start-pipe-replacement-initiative/>
- Myers, JC. *Geostatistical Error Management: Quantifying Uncertainty for Environmental Sampling and Mapping*. John Wiley & Sons; New-York, NY: 1997.
- Rabin R. The lead industry and lead water pipes “A MODEST CAMPAIGN”. *Am J Public Health.* 2008; 98:1584–1592. [PubMed: 18633098]
- Ridley, G. More than a third of Flint homes that got pipe fixes did not have high lead. 2016 Jul 14. Available at: http://www.mlive.com/news/flint/index.ssf/2016/07/lead_lacking_from_some_flint_h.html Accessed November 27, 2016
- SAS Institute Inc. *SAS/STAT 9.3 User’s guide*. SAS Institute Inc; Cary, NC: 2011.
- Sentinel Site Selection. 2016 Feb 9. Retrieved from http://www.michigan.gov/documents/flintwater/Sentinel_Site_Selection_2-9-2016_Final_525077_7.docx on February 2, 2017
- Testing Plan Process & Protocols DRAFT. 2016 Jan 27. Retrieved from <http://docs.house.gov/meetings/IF/IF14/20160413/104765/HHRG-114-IF14-Wstate-CreaghK-20160413-SD003.pdf> on February 2, 2017
- US Environmental Protection Agency, Office of Water. Lead and Copper Rule 40 CFR Part 141 Subpart I 1991. 1991. Available at: <https://www.epa.gov/dwreginfo/lead-and-copper-rule> Accessed July 18, 2016
- US Environmental Protection Agency, Office of Water. Lead and Copper Monitoring and Reporting Guidance for Public Water Systems. 2002. Available at: <https://www.epa.gov/dwreginfo/lead-and-copper-rule-compliance-help-public-water-systems> Accessed August 30, 2016
- US Environmental Protection Agency. 3Ts for Reducing Lead in Drinking Water in Schools. Revised Technical Guidance. 2006. Available at: https://www.epa.gov/sites/production/files/2015-09/documents/toolkit_leadschools_guide_3ts_leadschools.pdf Accessed November 23, 2016

- US Environmental Protection Agency, Office of Ground Water & Drinking Water. Memorandum: Clarification of Recommended Tap Sampling Procedures for Purposes of the Lead and Copper Rule. 2016a. Available at: <https://www.epa.gov/dwreginfo/memo-clarifying-recommended-tap-sampling-procedures-lead-and-copper-rule> Accessed August 30, 2016
- US Environmental Protection Agency, Office of Policy. EJSCREEN Technical Documentation. 2016b. Available at: https://www.epa.gov/sites/production/files/2016-07/documents/ejscreen_technical_document_20160704_draft.pdf Accessed February 2, 2017
- Waller, LA., Gotway, CA. Applied Spatial Statistics for Public Health Data. John Wiley & Sons; Hoboken, NJ: 2004.
- Wang Z, Devine H, Zhang W, Waldroup K. Using a GIS and GIS-assisted water quality model to analyze the deterministic factors for Lead and Copper corrosion in drinking water distribution systems. J Environ Eng. 2014; 140(9):A4014004.

HIGHLIGHTS

- State-controlled monitoring collected biweekly water samples at 739 sentinel sites.
- Sentinel sites are not as representative of Flint housing stock as voluntary sites.
- Unlike sentinel program voluntary sampling indicates increase in lead levels.
- State did not sample houses with lead service lines in two of the poorest wards.
- Interior plumbing might contribute more to lead in Flint water than service lines.

Fig. 1. Spatial distribution of sampling sites (A, red dots indicate sentinel sites) and probability of occurrence of sentinel sites mapped by indicator kriging (B). Data layers include: construction year (C) and composition of service line (D) for each residential tax parcel, block group percentages of families living below twice the poverty line in 2015 (E), and ward-level percentage of elevated blood lead levels recorded in children during the first 9 months of 2015 (F). The boundaries of 40 census tracts are overlaid on map D, while all

other maps show boundaries of nine Flint wards. Locations of sentinel sites are also overlaid on the poverty map E to illustrate the clustering of sites in areas of lower poverty.

Fig. 2.

Temporal changes in WLLs and characteristics of houses visited by the sentinel and voluntary sampling programs during the five rounds (S1–S5) defined in Table 1. The following statistics were computed from data collected during each sampling round: (A) % WLL data above 15 $\mu\text{g/L}$, (B) % houses with lead service lines (digital and on-site data are plotted for sentinel sites), (C,D) % houses built before 1935 or between 1935 and 1950, and (E,F) % houses located in block groups (BG) with lower (< 55%) or higher (> 75%)

percentages of habitants living below twice the poverty line. Statistics for Flint's 51,045 residential parcels are depicted by horizontal dashed lines.

Fig. 3.

(A,B) Frequency distributions of construction year and block group poverty level for all residential parcels in Flint, as well as houses visited by the sentinel and voluntary sampling programs between 2/16/2016 and 4/15/2016. The same data were used to compute conditional distributions that illustrate the impact of construction year and poverty level on the presence of lead service lines (C,D), and the relationship between construction year and poverty level (E). All distributions were smoothed using a kernel of size 11, and results based on less than 50 observations were discarded.

Fig. 4.

Conditional frequency distributions illustrating the impact of construction year and block group poverty level on the magnitude of WLLs, measured by their percentile for each sampling round (A,B), and the percentage of WLL data above 1 µg/L (C,D) and 15 µg/L (E,F). Data were collected by the sentinel and voluntary sampling programs between 2/16/2016 and 4/15/2016. All distributions were smoothed using a kernel of size 11, and results based on less than 50 observations were discarded. Horizontal dashed lines represent

either the 50th percentile (A,B) or the average overall percentage of WLL data above 1 µg/L (C,D) or 15 µg/L (E,F) at both types of sampling sites.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 5.

Impact of poverty level (i.e., percentage of habitants living below the poverty line) on statistics computed for each of the 9 wards in Flint: A) relative percentage of residential parcels, B) percentage of houses with lead service lines, C) percentages of water lead levels above 15 µg/L, and D) relative differences between curves displayed in (C). To facilitate visualization of trends the nine ward-level statistics are joined by line segments (piecewise linear function). A solid line is used for statistics computed from Flint's 51,045 residential parcels, while dashed lines correspond to data collected by the sentinel and voluntary

sampling programs between 2/16/2016 and 4/15/2016. (E) Locations of sentinel sites overlaid on the construction year map to illustrate the clustering of sampled LSL (red dots) in neighborhoods with houses built between 1935 and 1950.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 6.

Stratification of 1431 WLL data collected during the first round of voluntary sampling on the basis of the existence of lead service lines and construction year (3 categories). The percentage of data above 15 µg/L ranges from 7.23% to 15.7%, depending on the stratum. A similar stratification into six categories was conducted on the 51,045 residential parcels in Flint. This information was used to adjust the overall percentage of WLL data above 15 µg/L (10.20%), leading to a rate (11.35%) that better represents Flint housing stock.

Fig. 7.

Impact of adjusting for an increasing number of covariates on the time series of the percentage of WLL data recorded above 15 µg/L at non-sentinel (voluntary sampling) and sentinel sites: (A) resampling (more likely for residences with high WLL); (B) resampling and presence of LSL; (C) resampling, LSL, and construction year; (D) resampling, LSL, construction year, and poverty level, and (E) resampling, LSL, construction year, poverty

level, and ward. (F) Fully adjusted time series with the corresponding 90% confidence interval.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Datasets available for the comparison of sentinel and voluntary sampling programs of water lead levels in Flint. Statistics for each of the five sampling rounds include the number of data available including repeated samples, the percentage of WLL above 15 µg/L, and the 90th percentile (P90).

Round	Sampling period	Sentinel Sampling		Voluntary Sampling			
		Data (n)	%WLL > 15 µg/L	P90 (µg/L)	Data (n)	%WLL > 15 µg/L	P90 (µg/L)
S1	2/16/2016–2/24/2016	607	9.51	14.0	1481	9.88	15.0
S2	2/25/2016–3/14/2016	607	8.33	13.0	1337	10.29	17.0
S3	3/15/2016–3/28/2016	652	7.92	12.0	889	11.89	19.0
S4	3/29/2016–4/5/2016	640	7.29	10.0	457	14.02	37.0
S5	4/6/2016–4/15/2016	617	6.51	10.0	481	11.66	19.0
Total	2/16/2016–4/15/2016	3123	7.90	12.0	4645	10.95	18.0

Table 2

Percentage of sentinel and non-sentinel sites sampled for the first time in each round and the corresponding percentage of WLL recorded above 15 µg/L. Statistics for all sites and sites that were sampled previously (repetitions) are also listed.

Sampling Round	Sentinel Sampling			Voluntary Sampling		
	% first time sampling	% WLL > 15 µg/L		% first time sampling	% WLL > 15 µg/L	
		All sites	First sampling	Repetition	All sites	First sampling
S1	100	9.51	9.51	–	100	9.88
S2	14.14	8.33	9.53	8.14	95.6	10.10
S3	7.52	7.92	14.59	7.37	90.4	10.99
S4	3.80	7.29	8.33	7.25	90.0	12.53
S5	1.47	6.51	44.44	5.95	80.3	9.52
					11.66	20.26

Table 3

Odds ratio for covariates of the GEE regression models fitted separately to the data collected by the sentinel and voluntary sampling programs between 2/16/2016 and 4/15/2016. In both cases, three thresholds were considered: 1, 15 and 25 µg/L. Digital data on composition of service lines are used for all datasets to facilitate comparison among regression models.

Effects	Sentinel Sampling		Voluntary Sampling	
	1 µg/L	15 µg/L	1 µg/L	15 µg/L
SL: lead vs others/unknown	3.78**	1.90*	1.35	2.06**
Built year: <1935 vs >1950	2.01**	0.76	0.88	1.53*
Built year: 1935–1950 vs >1950	1.39*	0.62	0.61	2.09**
Poverty: 55–75% vs <55%	1.51**	1.36	1.48**	2.20**
Poverty: >75% vs <55%	1.05	0.97	1.34*	0.91
			0.80*	0.87

* Significantly different from 1 at $\alpha=0.05$

** Significantly different from 1 at $\alpha=0.01$

Number of residential parcels and statistics on water lead levels recorded between 2/16/2016 and 4/15/2016 by the sentinel and voluntary sampling programs within each ward in Flint. Percentages of elevated blood lead levels recorded in children are from Table 2 in Hanna-Attisha *et al.* (2016) and correspond to the period 1/1/2015 and 9/15/2015, that is after the water source change from Detroit-supplied Lake Huron water to the Flint River.

Table 4

Statistics	Flint Ward								
	1	2	3	4	5	6	7	8	9
Total number of residential parcels	6296	6406	7270	5257	6311	4784	4472	5637	4612
Number of Sentinel samples	159	477	216	172	154	464	625	430	374
Number of voluntary samples	429	450	301	506	350	497	665	654	506
% Flint residential tax parcels	12.33	12.55	14.24	10.30	12.36	9.37	8.76	11.04	9.03
% sentinel dataset	5.18	15.53	7.03	5.60	5.01	15.11	20.35	14.00	12.18
% voluntary dataset	9.84	10.33	6.91	11.61	8.03	11.40	15.26	15.01	11.61
% sentinel data > 15 µg/L	1.26	10.17	5.32	4.36	6.49	8.94	11.04	8.02	4.81
% voluntary data > 15 µg/L	6.84	11.65	9.97	7.91	14.05	13.04	10.66	13.08	10.87
% elevated blood lead levels	2.8	1.4	4.5	1.7	15.7	9.3	5.9	1.4	1.6

Table 5

Characteristics (presence of lead service lines, construction year, poverty level) of houses visited by the sentinel and voluntary sampling programs within each ward in Flint. Statistics are on all samples collected between 2/16/2016 and 4/15/2016, including repeated sampling of the house in different sampling rounds. Reference values were computed for Flint housing stock to any sampling bias. Digital data on composition of service lines are used for all three datasets facilitate comparison.

Statistics	Flint Ward								
	1	2	3	4	5	6	7	8	9
<i>% Lead service lines</i>									
Flint housing stock	3.27	8.34	8.56	5.48	9.95	9.51	6.46	6.88	7.07
Sentinel monitoring	0	9.01	3.24	1.74	0	10.56	10.88	5.58	24.06
Voluntary sampling	0.93	7.33	10.63	4.55	13.14	8.45	7.82	7.34	8.70
<i>% Pre-1935 houses</i>									
Flint housing stock	13.71	28.94	58.40	36.37	82.09	50.69	37.54	35.52	36.99
Sentinel monitoring	18.87	31.45	51.39	32.56	92.86	34.70	29.60	32.09	19.52
Voluntary sampling	7.69	16.89	47.51	24.70	74.00	43.86	33.53	27.68	26.28
<i>Block group (BG) poverty level (mean)</i>									
Flint housing stock	66.33	63.15	74.60	65.04	73.24	67.27	56.66	59.07	64.44
Sentinel monitoring	69.94	63.64	81.36	65.11	80.13	63.79	51.20	53.24	62.45
Voluntary sampling	62.12	60.43	73.82	61.76	70.99	66.39	52.80	56.93	60.49
<i>% houses in > 75% poverty BG</i>									
Flint housing stock	39.42	18.81	65.78	22.58	53.38	33.86	23.48	17.35	28.95
Sentinel monitoring	58.49	7.34	92.13	37.21	68.83	17.24	8.80	9.07	18.45
Voluntary sampling	34.50	12.22	61.13	20.75	44.00	32.39	15.64	12.54	17.19