Science of the Total Environment xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

The drinking water contamination crisis in Flint: Modeling temporal trends of lead level since returning to Detroit water system

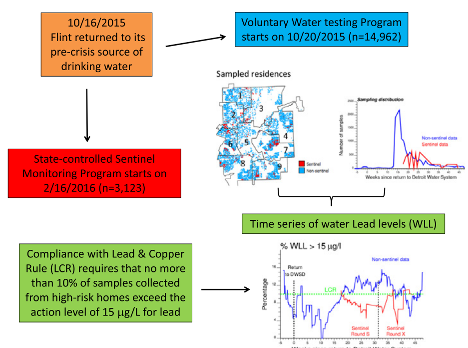
Pierre Goovaerts

BioMedware Inc., PO Box 1577, Ann Arbor, MI 48106, USA

HIGHLIGHTS

- Results from State-controlled and voluntary samplings are statistically different.
- Temporal trends in WLL vary greatly within the city of Flint.
- Timing and rates of change were estimated using joinpoint regression.
- Percentage of non-sentinel data above 15 μg/L recently started declining.
- Compliance with Lead and Copper Rule is still questionable.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 22 July 2016
Received in revised form 25 September 2016
Accepted 26 September 2016
Available online xxxx

Editor: D. Barcelo

Keywords:
Joinpoint regression
Generalized estimating equations
Sampling
Lead and copper rule
Lead service lines

ABSTRACT

Since Flint returned to its pre-crisis source of drinking water close to 25,000 water samples have been collected and tested for lead and copper in >10,000 residences. This paper presents the first analysis and time trend modeling of lead data, providing new insights about the impact of this intervention. The analysis started with geocoding all water lead levels (WLL) measured during an 11-month period following the return to the Detroit water supply. Each data was allocated to the corresponding tax parcel unit and linked to secondary datasets, such as the composition of service lines, year built, or census tract poverty level. Only data collected on residential parcels within the City limits were used in the analysis. One key feature of Flint data is their collection through two different sampling initiatives: (i) voluntary or homeowner-driven sampling whereby concerned citizens decided to acquire a testing kit and conduct sampling on their own (non-sentinel sites), and (ii) State-controlled sampling where data were collected bi-weekly at selected sites after training of residents by technical teams (sentinel sites). Temporal trends modeled from these two datasets were found to be statistically different with fewer sentinel data exceeding WLL thresholds ranging from 10 to 50 μg/L. Even after adjusting for housing characteristics the odds ratio (OR) of measuring WLL above 15 μg/L at non-sentinel sites is significantly >1 (OR = 1.480) and it increases with the threshold (OR = 2.055 for 50 μg/L). Joinpoint regression showed that the city-wide percentage of WLL data above 15 μg/L displayed four successive trends since the return to Detroit Water System. Despite the recent improvement in water quality, the culprit for differences between sampling programs needs to be identified as it impacts exposure assessment and might influence whether there is compliance or not with the Lead and Copper Rule.

E-mail address: goovaerts@biomedware.com.

http://dx.doi.org/10.1016/j.scitotenv.2016.09.207
0048-9697/© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The drinking water contamination crisis in Flint, Michigan has attracted national attention since extreme levels of lead were recorded following a switch in water supply that resulted in water with high chloride and no corrosion inhibitor flowing through the aging Flint water distribution system. This happened 15 years after another well-publicized "lead in drinking water crisis" that impacted Washington, DC following a change in disinfectant that altered the water chemistry and caused lead to leach from lead service line pipes (Edwards & Dudi, 2004). In both cases, the resulting contamination increased significantly the number of children with elevated blood lead levels and an emergency response was initiated (Edwards et al., 2009; Hanna-Attisha et al., 2016). Other water drinking crises of smaller magnitude included Greenville and Durham (North Carolina), or Lakehurst Acres, a public housing development in Maine (Renner, 2009).

The series of events and decisions that led to the Flint water crisis are now well documented (Flint Water Advisory Task Force, 2016) and briefly summarized hereafter. In April 2014, the City of Flint, Michigan switched its public water supply from the Detroit Water and Sewerage Department's system (DWSD) to the Karegnondi Water Authority (KWA), drawing and treating water from the Flint River. The monitoring of lead and copper in the distribution water started in July 2014, and the following month an *E. coli* violation resulted in residents being instructed to boil their drinking water. In December 2014 water samples showed elevated levels of lead, copper, as well as trihalomethanes (disinfection by-product of chlorine). A public health emergency was declared and residents were told to stop boiling their water and not drink it until testing or installation of approved water filters. Over this 9-month time period Flint residents expressed concerns about the quality of the water and changes in their health from skin rashes to symptoms of deteriorating-health. In January 2015, 42 cases of Legionellosis in Genesee County were confirmed. In February 2015, a sample reported a water lead level (WLL) of 104 μg/L, far above the EPA action level of 15 μg/L (ATSDR, 2010); yet according to MDEQ (Michigan Department of Environmental Quality) all other samples in the monitoring period (7/1/2014–12/31/2014) were compliant. Concerns about how the samples were collected (after flushing water through the taps) and the sampling strategy (using samples not necessarily drawn from highest-risk homes) were however raised (Flint Water Advisory Task Force, 2016). In March 2015 the Flint water treatment plant admitted the lack of corrosion control treatment (CCT). In July 2015 public concerns were raised that lead and copper were being leached from corrosion (chlorine-induced) in the underground service line and home plumbing fixtures as a result of not using CCT. In August and September 2015 the water was resampled by Virginia Tech's team and lead was determined to be "a very serious problem"; in particular 16.6% of the 271 first-draw water samples had in excess of 15 μg/L of lead (FlintWaterStudy.org, 2015). In September and October 2015 elevated childhood blood lead levels were confirmed and an emergency response was initiated (Hanna-Attisha et al., 2016) leading the city to switch back to the DWSD water supply on October 16, 2015.

The Lead and Copper Rule (LCR, US EPA, 1991, 2002a, 2016) sampling is intended to measure the lead and copper levels in drinking water to assess the effectiveness of CCT utilized by public water systems. Under that rule, first-draw 1-L water samples must be collected after a minimum of 6 h. of stagnation (e.g., overnight stagnation) and without pre-flushing the tap prior to the stagnation period (i.e., no pre-stagnation flushing). Compliance with that rule requires that no more than 10% of samples collected from high-risk homes exceed the action level of 15 μg/L for lead and 1.3 mg/L for copper. This is equivalent to requiring that the 90th percentile or P90 (concentration exceeded by 10% of samples) is no > 15 μg/L for lead or 1.3 mg/L for copper. If non-compliance is observed the system must undertake a number of additional actions to control corrosion. If the action level for lead is exceeded, the system must also inform the public about steps they should take to protect their health and may have to replace lead service lines under their control.

The LCR includes a tiering system for prioritizing the selection of sampling sites based on the likelihood of the sites to release elevated levels of lead; e.g., sites with lead service lines (LSLs), lead pipes, or copper pipes with lead solder. 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 delay in reporting high levels of lead in Flint drinking water 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, instead targeting properties on the eastern and western fringes of the city which, in some cases, were a long way from any apparent source of lead (Milman, 2016c). Flint is not the only city that neglected to follow EPA guidance as 33 cities across 17 US states have reportedly used water testing "cheats" that potentially conceal dangerous levels of lead (Milman and Glenza, 2016). Besides not sampling tier 1 category houses, underestimation of WLLs can be achieved by asking testers: 1) to run faucets prior to the 6 h. stagnation period (pre-stagnation flushing) which removes water that may have been in contact with LSLs for extended periods, which is when lead typically leaches into drinking water, 2) to remove or clean faucet filters called "aerators" which can collect lead particles, 3) to slowly fill sample bottles or use narrow-necked bottles that do not allow for high flow rates, thereby attenuating the release of particulate and colloidal lead, and 4) to conduct sampling in cooler months when lead concentrations in water are lower because lead dissolves less readily in chilled water (Renner, 2009). These practices led EPA to issue a memorandum in February 2016 to clarify the recommended tap sampling procedures (US EPA, 2016).

Almost all public water systems in the US rely on residents to collect compliance samples and sampling instructions often differ, in particular when it comes to water usage during the stagnation period (e.g., instruction not to use any water from the tap being sampled or from the entire household) and the application of pre-flushing (Del Toral et al., 2013). In other countries, such as Canada or France, sampling must be conducted by a trained technician, making a long stagnation period difficult to implement (Cartier et al., 2011). According to regulatory protocols in Europe (Hoekstra et al., 2009), it should be feasible for a sampler to take at least 10 samples during a normal working day. Random daytime samples (without prior stagnation times) and samples taken after a fixed stagnation time of 30 min. Are thus taken in large numbers because this type of sampling is not reproducible if few samples are collected.

As stressed by Cartier et al. (2011) all sampling methods have benefits and disadvantages. One of the benefits of sampling by residents is the ability to collect a large number of data within a short amount of time. During the first eleven months following Flint's return to its pre-crisis source of drinking water, close to 25,000 water samples have been collected almost daily, mainly by homeowners using two different sampling initiatives. The majority of these samples (80%) were collected through a voluntary or homeowner-driven sampling whereby concerned citizens decided to acquire a free testing kits available to residents at local water distribution centers and conduct sampling on their own (non-sentinel sites). The testing kit comes with written sampling instructions, including the recommendation that the sampled tap should not be flushed before sample collection. Pre-stagnation flushing was however part of sampling instruction for the City of Flint until December 2015 and MDEQ waited until January 2016 to amend its water testing rules for Michigan (Milman, 2016b). It is noteworthy that the written instructions still ask residents to pour water into sample bottles "gently", which is part of controversial practices since it reduces the amount of lead and other material that is dislodged from pipes by a strong flow of water (Milman, 2016a).

Starting on 2/16/2016 samples were also collected bi-weekly at >600 sentinel sites chosen by the EPA and MDEQ across the city to determine the general health of the distribution system and to track

Please cite this article as: Goovaerts, P., The drinking water contamination crisis in Flint: Modeling temporal trends of lead level since returning to Detroit water system, Sci Total Environ (2016), http://dx.doi.org/10.1016/j.scitotenv.2016.09.207

changes in lead concentrations over time (Flint Safe Drinking Water Task Force, 2016). Sentinel teams including a member of DEQ, a licensed plumber and a community member visited the houses selected to be part of the sentinel network. A plumbing inspection was conducted to verify the service line material entering the home and residents were shown how to draw samples of their water in a scientifically accurate manner. In June 2016, after five rounds of sentinel sampling, a new sentinel program called "Extended Sentinel Site Program" started with a focus on the highest-risk areas. Criteria for selecting this smaller set of 160 homes (Calley, 2016) included: 1) have known LSLs, 2) have service lines the State paid to replace under Mayor Karen Weaver's Fast Start Program, 3) have copper and galvanized service lines found to have high lead levels during the original sentinel program, and 4) are from areas where elevated blood lead levels were suspected to be higher. Encouraging results were reported by the State of Michigan in August 2016, "*Despite the need for further action, all recent data shows an encouraging trend of improvement. Especially encouraging is the fact that 93 percent of samples from round three of the Extended Sentinel Site program are also at or below the lead action level. This marks the third sampling round in a row in which the Extended Sentinel Site data meets the LCR action level criteria*". In July 2016 Governor Snyder (2016) even proposed to use 10 μg/L instead of 15 μg/L as action level to offer more protection for residents than federal rules (LCR) provide.

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 results have been posted periodically at http://www.michigan.gov/flintwater. To the author's knowledge, these data have undergone little processing so far besides the creation of a few dot maps highlighting locations where WLL exceeds 15 μg/L, supplemented by the computation and comparison of summary statistics (e.g., 90th percentile, percentages above specific thresholds) between sentinel sampling rounds. In particular, results from the sentinel and voluntary (referred to as "non-sentinel" hereafter) programs have not been compared.

The main objectives of this study were to: i) create a reliable space-time database of WLL in Flint matching each observation with a tax parcel unit where housing characteristics (e.g., presence of lead SL, built year) are available, ii) identify by exploratory data analysis and Generalized Estimating Equations (GEE) a few easily accessible variables that influence the likelihood of WLL above 15 μg/L and use this information to correct for potential sampling bias, iii) estimate the timing and rate of change in the percentage of water samples above 15 μg/L using joinpoint regression (Kim et al., 2000; Goovaerts, 2013), and iv) explore the impact of the type of sampling program (sentinel vs non-sentinel) and geographical location on the results. In addition to describing the first application of joinpoint regression to the modeling of time series of WLL data, this paper presents the first comprehensive statistical study of how lead level in Flint drinking water has changed since corrosive water from the Flint River stopped flowing through the city distribution network.

2. Data sources and methods

2.1. Datasets

24,755 WLL measurements recorded over the period 9/3/2015–9/15/2016 were downloaded from http://www.michigan.gov/flintwater (residential testing results). Data were then allocated to an individual tax parcel unit on the basis of their postal address. Data with incomplete address (416 samples), collected outside the city limits (1353 samples), or that failed to geocode (80 samples) were discarded. Since the EPA Lead and Copper rule focuses on residences, all 3931 samples collected on tax parcels classified as non-residential (e.g., industrial, commercial, utility) were excluded. This also ensured a greater uniformity between sentinel and non-sentinel sites since 98.8% of sentinel samples were collected on residential parcels while this percentage was much smaller (79.8%) for non-sentinel sites. Most of the remaining data (18,760 = 98.9%) were collected after Flint returned to DWSD water supply (10/16/2015), which is the focus of the present analysis.

Online data are stored in ten different Excel files: one file for each round of sentinel sampling (Table 1) and a master file that combines all data without any information on their origin (i.e., sentinel vs non-sentinel sites). The identification of sentinel data in the master file was straightforward for sampling rounds S5 and X1X4 (extended sentinel site program) since both the master file and the sentinel files include a sample ID number. Because sentinel data files for sampling rounds S1 to S4 do not include this ID number nor a street number, sentinel data in the master file were identified based on their sampling date, street name, and lead and copper levels in water. Although sentinel sites might have been sampled on other occasions (e.g., prior to the start of the sentinel program), only the data included in sentinel Excel files were labeled as sentinel in the subsequent analysis. Sentinel data represent 20.2% of WLL data (3798 observations) and 7.3% of tax parcels sampled (811 out of 11,152). Note that a tax parcel can include multiple sentinel or non-sentinel sites (e.g., individual units in an apartment complex).

A digital map of Flint's lead water pipe was obtained from the GIS center at the University of Michigan, Flint. The shape file includes for each of the 56,039 tax parcel units the type of service line (SL) classified into three categories: lead, other, or unknown. These data were gathered by combining information from 45,000 index cards with 240 parcel maps from the city water department, showing city blocks divided into little squares of property, with a code that indicated which type of service line was running into each house (Gold, 2016). The reliability of this dataset was here investigated by comparing the digital data with the composition of supply line recorded during the sentinel team visit

Table 1
Datasets available for the time trend analysis: master file and nine sentinel files for each sampling round. Statistics include the number of data available (samples on residential parcels), the sampling period, the percentage of WLL above 15 μg/L, the 90th percentile (P90), and the composition of service line (SL) that was only reported for sentinel data (3 main categories besides plastic, unknown, and other).

	Data (n)	Sampling period	%WLL > 15 μg/L	P90 (μg/L)	Composition of SL		
					Lead	Galvanized	Copper
Master file	18,760	10/20/2015–9/15/2016	8.98	13.0			
Sentinel files							
Round S1	607	2/16/2016–2/29/2016	9.56	14.0	5.93	20.76	68.53
Round S2	607	2/24/2016–3/13/2016	8.40	13.0	9.39	19.93	67.05
Round S3	652	3/15/2016–3/24/2016	8.13	12.0	11.66	19.48	64.11
Round S4	640	3/29/2016–4/5/2016	7.19	10.0	13.59	17.50	64.22
Round S5	617	4/13/2016–4/15/2016	6.48	10.0	13.94	15.24	65.64
Round X1	169	5/23/2016–6/7/2016	7.10	12.0	44.97	9.47	44.97
Round X2	179	6/14/2016–6/30/2016	9.50	15.0	49.16	7.82	42.46
Round X3	166	7/19/2016–7/22/2016	6.63	12.0	46.99	8.43	43.98
Round X4	161	8/15/2016–8/22/2016	9.94	15.0	43.48	9.32	45.96

Please cite this article as: Goovaerts, P., The drinking water contamination crisis in Flint: Modeling temporal trends of lead level since returning to Detroit water system, Sci Total Environ (2016), http://dx.doi.org/10.1016/j.scitotenv.2016.09.207

Table 2
Contingency table comparing for 811 sampling sites the composition of supply line recorded during the sentinel team visit (on-site data) with the information retrieved from the digital map (digital data).

	Digital data			Total
	Lead	Other	Unknown	
On-site data				
Lead	74	8	35	117
Other	12	562	84	658
Unknown	7	0	29	36
Total	93	570	148	811

(811 on-site data). Digital data predicted accurately whether the supply line was lead, other or unknown in 82% of cases (Table 2). Most discrepancies were caused by a lack of information (i.e., unknown material) rather than lead SLs being mistakenly classified as non-lead, and vice versa: 6.84% of sentinel lead lines were misclassified as other material while only 1.82% of non-lead lines were classified as lead SL, resulting in an underestimation of the percentage of LSL by digital data (93 vs 117 sites). The strength of the relationship between both types of data was quantified using the Kappa statistic (Cohen, 1960), with a value $\kappa = 0.56$. According to Fleiss (1981) a value between 0.4 and 0.6 indicates a fair to good agreement between the two SL classifiers.

The following data were also derived from the 2016 Parcels GIS layer provided by Flint GIS: parcel classification use (e.g., residential, industrial, commercial, utility), and the year the house was built (Year_built). The attribute "Year_built" was missing for 20,372 parcels and was estimated by ordinary kriging (Goovaerts, 1997) with a mean absolute error of prediction of 6.43 years. Each parcel was then allocated to one of three categories of built year: <1940, 1940–1959, and >1959. Lead in drinking water mainly comes from metal pipes installed in older homes and from solder at pipe connections installed prior to 1986 (Lee et al., 1989; Cartier et al., 2011; Clark et al., 2015). Because <2% of samples were collected in houses built after 1986, this year could not be considered as a threshold year. Instead, the somewhat arbitrary years of 1940 and 1960 were used for the discretization of this continuous variable. Although year of construction and presence of lead SL lines are expected to provide overlapping information, the analysis of their contingency table yielded a negative Kappa statistic (−0.1408), indicating a poor agreement between these two categorical variables (Landis and Koch, 1977).

Percentage of habitants living below the poverty line in 2014 was obtained for each census tract (CT) in order to characterize the socio-economic status of sampled homes.

There are many other variables known to influence lead in drinking water. For example, longer water age (i.e., water travel time between the treatment plant and home plumbing system) can decrease the effectiveness of corrosion control; increasing leaching and water lead levels (US EPA, 2002b; Wang et al., 2014). Temperature is another major factor affecting lead release (Schock, 1990; Deshommes et al., 2013); for example, Cartier et al. (2011) reported a 5% rise in dissolved lead per 1 °C increase in water temperature between 10 and 23 °C. The main objective of the present study was however to rely on easily available data layers to correct for potential sampling bias in space and time during the interpretation of temporal trends. A detailed analysis of the causes for elevated lead levels is beyond the scope of the current paper.

2.2. Generalized estimating equations

The analysis started with the coding of each WLL data $z(\mathbf{u}_\alpha;t)$ into an indicator of being greater or not than the threshold $z_c = 15$ µg/L:

$$i(\mathbf{u}_\alpha;t) = \begin{cases} 1 \text{ if } z(\mathbf{u}_\alpha;t) > z_c \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

where $\mathbf{u}_\alpha = (x_\alpha, y_\alpha)$ are the geographical coordinates of the tax parcel centroids and t represents the sampling date. A regression model was then fitted to predict the probability of exceeding 15 µg/L on the basis of four covariates: three categorical variables (type of service line, sentinel vs non-sentinel sites, and year of construction), and one continuous variable (time since source water switch). Because sentinel sites 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., house built prior to 1940), compared to the odds of the outcome occurring in the absence of that event (e.g., house built after 1940). The impact of geographical location on the results was also explored by fitting a regression model within each of Flint nine wards.

2.3. Joinpoint regression

While GEE tests only for the existence of a linear temporal trend, a more detailed modeling of these trends can be achieved using joinpoint regression. Let $\{r(t), t = 1, \ldots, T\}$ be the percentages or rates of WLL above 15 µg/L recorded at T different time periods (e.g., weeks). Each observation $r(t)$ is computed as the ratio $d(t)/n(t)$, where $n(t)$ is the total number of water tests at time t while $d(t)$ is the number above 15 µg/L. A more general expression for the rate $r(t)$ is the following weighted mean of $n(t)$ indicators defined in Eq. 1:

$$r(t) = \sum_{\alpha=1}^{n(t)} \frac{l_{\alpha t}}{n(t)} i(\mathbf{u}_\alpha;t) \quad \text{with} \quad \sum_{\alpha=1}^{n(t)} \frac{l_{\alpha t}}{n(t)} = 1 \qquad t = 1, T \tag{2}$$

The weights $l_{\alpha t}$ can be used to correct for sampling bias across space and time. For example, temporal fluctuations in the sampled proportion of homes with lead service lines (SL), $p_{lead}(t)$, are accounted for by setting the weight $l_{\alpha t}$ to $P_{lead} / p_{lead}(t)$ or $(1 - P_{lead}) / (1 - p_{lead}(t))$, depending on whether the housing unit at $\mathbf{u}_\alpha$ has a lead SL or not. P_{lead} is here the time-invariant proportion of 51,045 residential tax parcel units with lead SL. By using P_{lead} as reference term one can correct for temporal variability and compute a rate that is representative of the housing stock in Flint at all times.

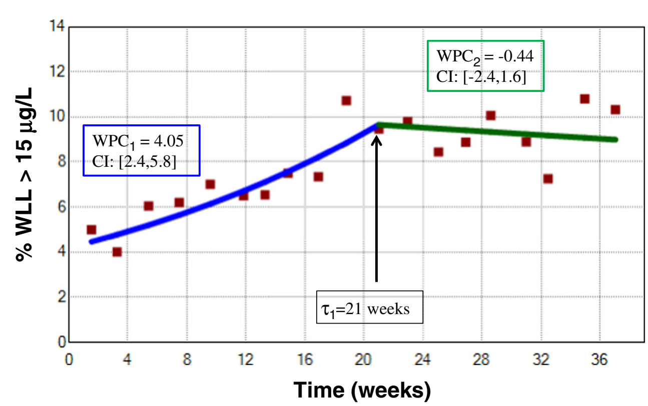


Fig. 1. Hypothetical time series of the percentages of WLL data above 15 µg/L that were measured over a 38 week time period The segmented regression model (solid line) includes one joinpoint (τ) that corresponds to the time of a statistically significant change in temporal trend: week #21. The estimates and 95% confidence intervals of the weekly percent change (WPC) are computed for each segment.

The log-linear version of the joinpoint regression model (Kim et al., 2000) takes the form:

$$Log(r(t)) = \mu(t) + \varepsilon(t) \qquad t = 1, \ldots, T \tag{3}$$

where $\varepsilon(t)$ is the residual for the t-th time, and the regression mean $\mu(t)$ is defined as a succession of $(k + 1)$ linear segments over the time interval $[a,b]$: $[a,\tau_1]$ … $[\tau_k,\tau_{k+1}]$ … $(\tau_k,b]$. The parameter τ_k is the timing (joinpoint) for a statistically significant change in the slopes β_k and β_{k+1} of two successive segments.

For example, Fig. 1 shows a hypothetical time series of percentages of WLL data above 15 μg/L recorded biweekly over a 38 week period ($T = 19$). The observed time series was fitted with a regression model that includes one joinpoint (τ_1 = week #21) using the public-domain Joinpoint Regression Program 4.3.1 April 2016 developed at the US National Cancer Institute, NCI (http://surveillance.cancer.gov/joinpoint/).

The unknowns in the segmented regression model (Eq. 3) include the number K and values τ_k of the joinpoints, as well as the slopes β_k of the different linear segments. They are estimated using a two-step procedure: 1) a grid search method (Lerman, 1980) is conducted over the set of possible joinpoints, and 2) at each step of the search the regression parameters and their standard errors are estimated by weighted least-square regression using the following criterion (Kim et al., 2000):

$$Q = \sum_{t=1}^{T} w(t)\left(\log(r(t)) - \mu(t)\right)^2 \tag{4}$$

The weights account for the fact that the variance of the residuals $\varepsilon(t)$ varies with time (heteroscedasticity) as the number of water tests can fluctuate greatly from one week to the next. The weights are the reciprocal of the variance which was here computed as $n(t) / [r(t) \times (1 - r(t))]$ according to a binomial distribution for the residuals. In addition to being heteroscedastic, the random errors in the regression model could be autocorrelated when the observations are serially correlated. The autocorrelation parameter was here computed automatically by the Joinpoint Regression Program using the procedure outlined in Kim et al. (2000).

The number K of joinpoints is estimated through an iterative procedure that tests whether models of increasing complexity (i.e., including more joinpoints) provide a significantly better goodness-of-fit than simpler models (Kim et al., 2009). The tests of significance are based on the Monte Carlo Permutation procedure described in Kim et al. (2000). To reduce the number of solutions and the computational time, a maximum number of joinpoints is typically specified (i.e., $K_{max} = 3$ here). To keep joinpoints from getting too close together or too close to either end of the time series, a minimum number of observations between joinpoints is also required and was set to 4 in the present application. This minimum number allows the computation of the standard error of the slope parameters and the associated p-values.

Temporal trend models fitted by joinpoint regression are characterized by: i) the number and timing of statistically significant changes (i.e., joinpoints τ_k), and ii) the rate of change computed for each linear segment $[\tau_k,\tau_{k+1}]$ (so called "weekly" percent change, WPC). Following Clegg et al. (2009) this statistic is derived from the slopes β_k of the regression model as:

$$\mathrm{WPC}_{k+1} = 100 \times \left(\exp\{\beta_{k+1}\} - 1\right) \tag{5}$$

A $(1 - \alpha)$ confidence interval (CI) can be computed and if it contains zero, then there is no evidence to reject the null hypothesis that the true rate of change is zero at the significance level of α. For the example of Fig. 1, the WPC is particularly large for the first segment: the percentage of data above 15 μg/L increased by 4.05% (relative rate) every week until week #21 before it started declining at a non-significant rate since the 95% CI includes 0.

3. Results and discussion

3.1. Time series of WLL data

Fig. 2 shows the time series of two statistics (90th percentile and % WLL data > 15 μg/L) computed from all sentinel and non-sentinel WLL data collected on residential parcels over the period 9/3/2015–9/15/2016. To facilitate visualization daily time series are displayed before (A,B) and after (C,D) smoothing using 15-day moving windows. The horizontal axis indicates the number of weeks since Flint returned to DWSD water supply; negative numbers correspond to the time when water from the Flint River was still used. Despite masking all statistics computed from < 10 observations, daily plots still display great temporal variability caused by huge fluctuations in the number of samples collected over time: from $n = 11$ to $n = 884$ per day (daily average = 123). Twenty to eighty water samples were collected weekly on residential parcels for the first three months after Flint returned to the DWSD as its source of drinking water. This number jumped to 1911 on week #14 (1/18/2016–1/24/2016) and hit a maximum of 2181 on week #16 before steadily decreasing (Fig. 2E, solid line). On week #18 sentinel data started being collected bi-weekly with a maximum of 640 on week #22 (Fig. 2E, dashed line).

Smoothing the time series using 15-day moving windows facilitates the visualization of temporal trends before and after the return to Detroit Water Systems, as well as the comparison of results obtained for sentinel vs non-sentinel sites (Fig. 2 C,D). Temporal trends until week #14 need however to be interpreted cautiously because of high variability caused by sparse sampling (Fig. 2E). Following the return to the Detroit Water System, the percentage of WLL data above 15 μg/L decreased until week #10 (late December 2015) before increasing again, which coincides with the time the sampling protocol was modified to stop pre-stagnation flushing and use normal flow rates during sampling (Milman, 2016a,b). This turn-around was confirmed when the number of voluntary samples increased substantially on week #14. In the following month both statistics increased steadily and exceeded the LCR action level on week #18 (2/15/2016–2/21/2016) which coincides with the start of the sentinel site program.

Although the sentinel site program aims to determine the general health of the distribution system, results of rounds S1–S5 displayed a trend opposite to the one observed for data collected by concerned citizens on a voluntary basis (non-sentinel data). Indeed, the 90th percentile computed from non-sentinel data kept increasing before reaching a plateau above 20 μg/L, which is more than twice the level derived from sentinel data on the same weeks. Expressed in terms of percentage of WLL data above 15 μg/L, the difference is 14% (non-sentinel data) to 7% (sentinel data). As expected, the start of the extended sentinel site program targeting high-risk areas on week #32 led to an increase in WLLs although the smaller number of sampling sites (160 instead of 600) caused an increase in temporal fluctuations. Around the same time, the WLL measured at non-sentinel sites started declining to fall below the LCR level around early August (week #42) before moving back above that level early September (week #46). The pattern observed in August was more in agreement with what is expected: sentinel sites targeting high-risk areas showed higher WLLs than non-sentinel sites.

Fig. 3 shows smoothed time series of percentage of WLL data above four other WLL thresholds ranging from 1 to 50 μg/L. Temporal trends for 1 μg/L differ from the other thresholds in that: 1) percentages increased only moderately after week #10 and never went above the levels observed in September and October 2015 when Flint River was still used as water source, and 2) percentages were generally greater at sentinel sites relative to non-sentinel sites, in particular during the extended sentinel site program. Another finding is that the percentage of WLL data above 10 μg/L has exceeded 10% at both sentinel and non-sentinel sites for most days since week #14 (1/18/2016–1/24/2016), which emphasizes the challenge of using 10 μg/L instead of 15 μg/L as

Please cite this article as: Goovaerts, P., The drinking water contamination crisis in Flint: Modeling temporal trends of lead level since returning to Detroit water system, Sci Total Environ (2016), http://dx.doi.org/10.1016/j.scitotenv.2016.09.207

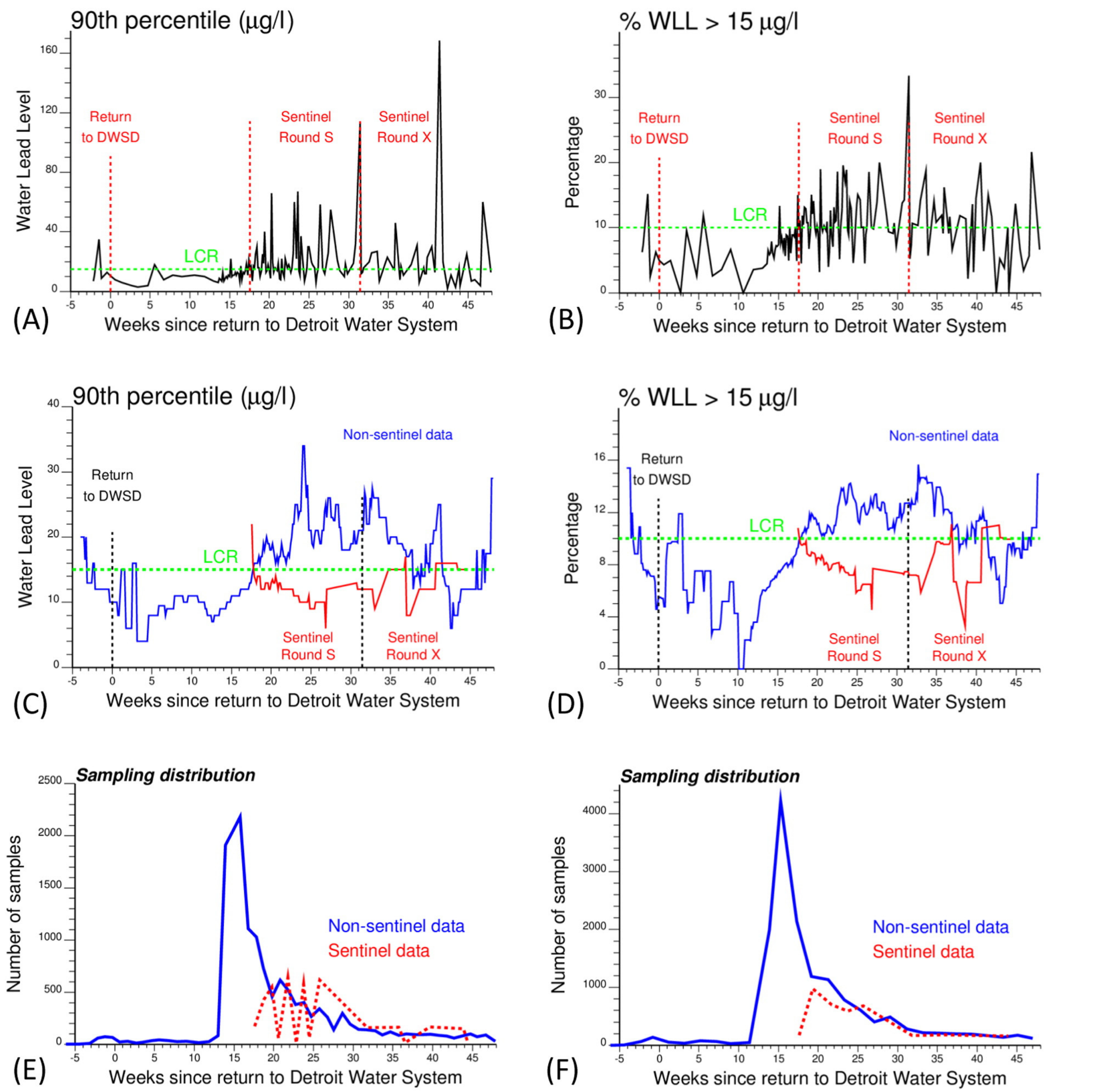


Fig. 2. Time series of two statistics (90th percentile and % WLL data >15 μg/L) computed from data collected during three sampling campaigns: voluntary (non-sentinel), sentinel and extended sentinel sites programs. The sampling period (9/3/2015–9/15/2016) covers 53 weeks, including five weeks prior to Flint returning to Detroit Water System. For easier visualization daily time series are displayed before (A,B) and after (C,D) smoothing using 15-day moving windows. Bottom graphs show the number of WLL data collected at sentinel and non-sentinel sites aggregated over 1 and 2 week non-overlapping windows.

action level for Flint. Last, the proportion of samples with WLL > 50 μg/L is non-negligible (4–6% at non-sentinel sites).

Any further interpretation of these plots requires however to account for the spatial and temporal distributions of sampled sites in relation to factors susceptible to impact WLL, such as housing characteristics.

3.2. Spatial sampling

Fig. 4A shows the location of all 18,760 WLL data collected on residential tax parcels (Fig. 4B) that were used for the creation of plots in Figs. 2 and 3. Sentinel data represent 20.2% of samples (3798 observations) and are depicted by red dots while polygons delineate Flint wards. The sampling density and the proportion of sentinel versus non-sentinel sites appear to vary spatially. This visual finding is confirmed by statistics listed in Table 3. The percentage of residential parcels that were sampled ranges from 10.34% (Ward 3) to 37.57% (Ward 7), while the percentage of water samples collected during the sentinel sampling program (S1–S5) fluctuates between 9.27% (Ward 4) and 24.25% (Ward 2). This uneven sampling density could be partially explained by the perceived risk for WLL to exceed 15 μg/L (targeted sampling). For example, Wards 3 and 4 have only 6.83–7.59% of samples above 15 μg/L, while this percentage is 9.29% for Ward 7 and 10.46% for Ward 2 (Table 3).

Housing characteristics are another expected driver for preferential geographical sampling. The two housing characteristics used as covariates in modeling (Section 3.4) are mapped in Fig. 2C–D and their ward-level statistics are summarized in Table 3. Ward 5 has a much larger percentage of pre-1940 homes (85.69%) with a higher frequency of lead service lines (9.95%). Two other wards with a greater percentage of

Please cite this article as: Goovaerts, P., The drinking water contamination crisis in Flint: Modeling temporal trends of lead level since returning to Detroit water system, Sci Total Environ (2016), http://dx.doi.org/10.1016/j.scitotenv.2016.09.207

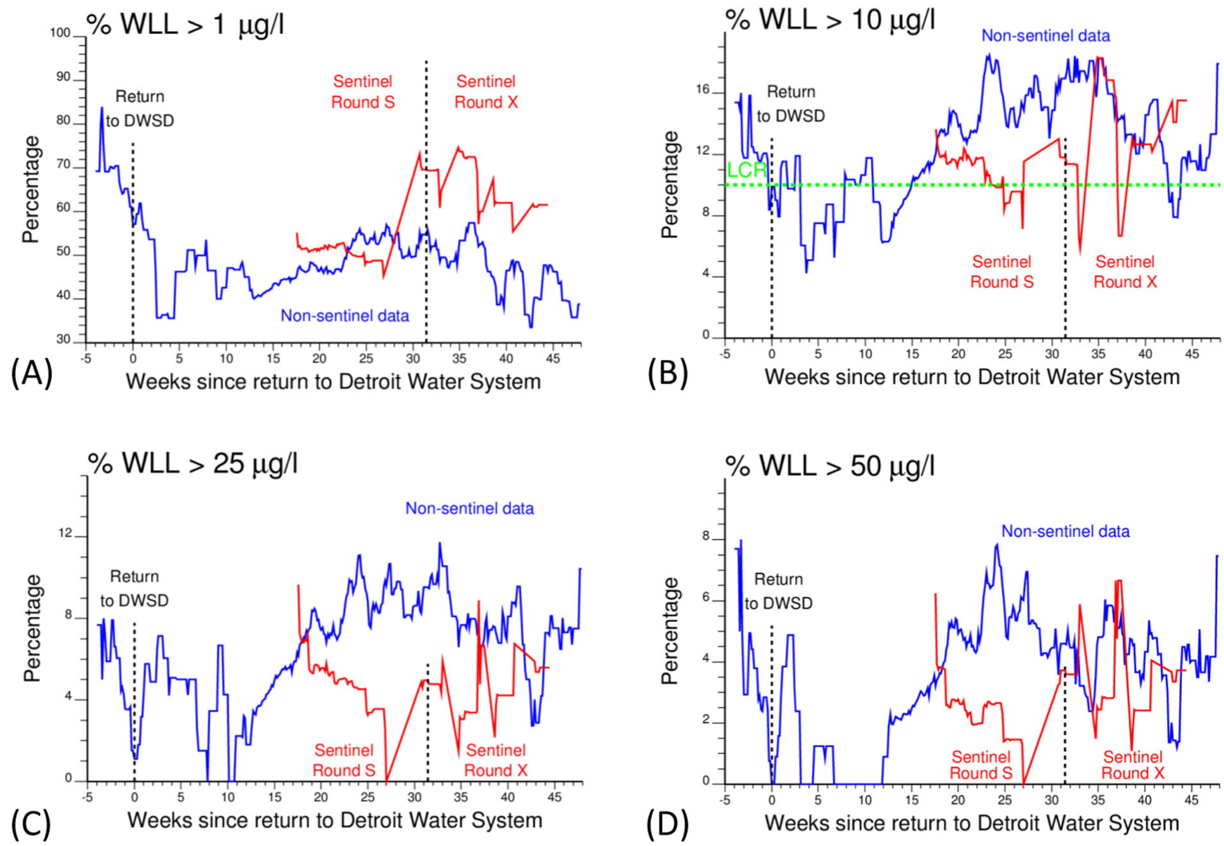


Fig. 3. Time series of the percentage of WLL data recorded above four different thresholds (1, 10, 25 and 50 µg/L) during three sampling campaigns: voluntary (non-sentinel), sentinel and extended sentinel site programs. The sampling period (9/3/2015–9/15/2016) covers 53 weeks, including five weeks prior to Flint returning to Detroit Water System. To facilitate visualization daily time series have been smoothed using 15-day moving windows.

older houses are Wards 6 (53.53%) and 7 (44.63%), which along with Ward 5 were found to have the highest percentages of elevated blood lead levels in children (Hanna-Attisha et al., 2016). On the other end, Ward 1 with only 10.88% of sentinel samples includes more recent houses (17.55% of pre-1940 houses), and these homes tend to have fewer lead SL (3.27%), resulting in the smallest percentage of WLL data above 15 µg/L (4.80%). In general, the 90th percentile and percentage of WLLs above 15 µg/L vary greatly among Wards; e.g., P90 is only 6 µg/L in Ward 1 (4.80% > 15 µg/L) but reaches 19 µg/L in Ward 6 (11.71% > 15 µg/L).

It is noteworthy that the three wards with the smallest percentages of parcels sampled (Wards 1, 3, 5) are also the wards with the largest percentages of habitants living below the poverty level (44.6–48.4%). On the other end, sampling density is the largest in the three wards (Wards 7, 8, s9) with the lowest poverty level (27.3–29.3%). One could argue that the low sampling density in Ward 1 is mainly due to the presence of fewer lead SLs, as explained above. The case of Ward 5 is more problematic since it was sampled less frequently than Ward 1 despite having the largest proportions of pre-1940 houses and lead SLs. Another ward-level statistic following a trend similar to poverty level is the difference between results of non-sentinel and sentinel (S1–S5 rounds) campaigns, as measured by the ratio of their percentages of WLL data above 15 µg/L. Table 3 (last row) indicates that this ratio exceeds one in all but one wards, which agrees with the visual interpretation of temporal trends (Figs. 2–3) in Section 3.1. The only ward with a ratio below 1, corresponding to a larger percentage of WLL data recorded above 15 µg/L at sentinel sites, is Ward 7 which has the second lowest percentage of habitants in poverty (27.96%). The highest ratios (4.32 and 1.69) are found in Wards 1 and 5 with poverty levels of 44.55% and 48.38%, respectively. These empirical findings do not imply causation but may warrant further investigation.

3.3. Temporal sampling

The heterogeneity of the housing stock in Flint, coupled with the uneven spatial sampling, stresses the need to incorporate this information when analyzing temporal trends of elevated water lead levels. The characteristics of sampled houses have also evolved over time. These are summarized on a bi-weekly basis for both non-sentinel and sentinel campaigns in Fig. 5. To facilitate comparison with WLL temporal trends, the daily time series displayed in Fig. 2A,B were also aggregated using 14-day non-overlapping windows (Fig. 5A,B). This level of aggregation is consistent with the sampling intervals for sentinel sites and allows smoothing out periodic fluctuations in the number of samples, compare Fig. 2F to 2E.

Fig. 5C shows that starting week #7 the percentage of non-sentinel samples collected in houses with lead SL has been fluctuating around Flint average for residential parcels (7.32%, Table 4) depicted by an horizontal dashed line. Sentinel program has been targeting a larger proportion of houses with lead SL, in particular in the extended sentinel site program. This percentage is now close to the level of 50% required under LCR. Percentages for sentinel sites were computed from both digital (solid line) and on-site (dashed line) LSL data. Although less accurate and underestimating the existence of LSLs (Table 2), digital data capture temporal trends in housing characteristics which will be used for bias adjustment in Section 3.5.

The percentage of samples collected in pre-1940 houses through the voluntary program follows a trend similar to the one displayed by LSL

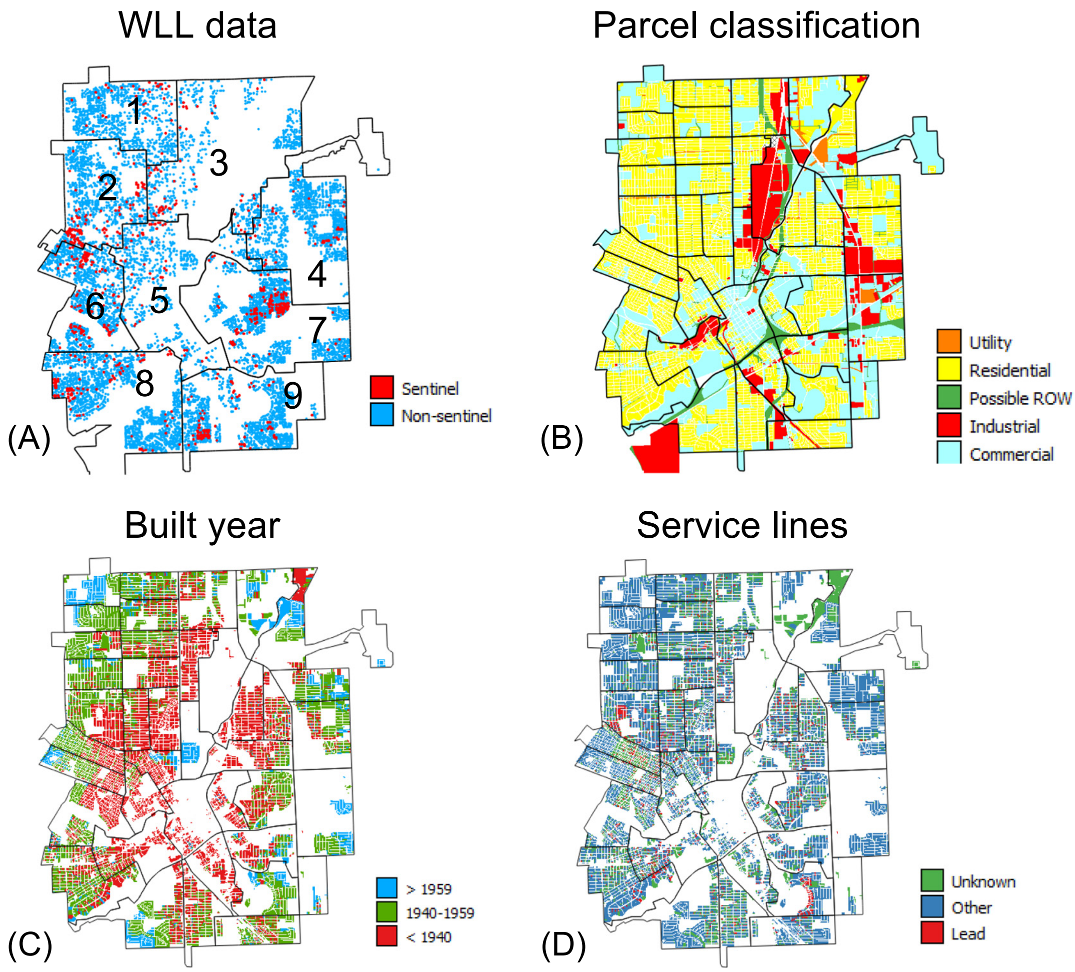


Fig. 4. Spatial distribution of sampling sites (A, red dots indicate sentinel sites) and key data layers available for each tax parcel unit: B) parcel classification use and possible ROW (right of way), C) year built (residential parcels), and D) composition of service lines. The boundaries of 9 wards and 40 census tracts are overlaid on maps A and B–D, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Fig. 5D), except that on average it has been 12% lower (35.92% vs 47.46%) than the global percentage in Flint (Table 4). The opposite trend is exhibited by sentinel sites with a steady decline during the sentinel rounds S1 to S5, leading to even fewer pre-1940 houses being sampled relative to the non-sentinel program. In both cases, one culprit is the under sampling of Ward 5 which has the highest percentage of pre-1940 homes (85.69%). A similar pattern is observed for poverty level: fluctuations around a level that is half Flint average poverty level for non-sentinel sites and steady sampling of less impoverished neighborhood for the sentinel program (Fig. 5E).

3.4. Housing characteristics and water lead level

Table 4 illustrates for each sampling campaign the impact of housing characteristics and poverty level on the percentage of WLL data above 15 μg/L. Data collected under the voluntary program follow expected

Table 3
Statistics on the residential parcels and WLL data collected within each ward in Flint. Data collected under the Extended Sentinel Site Program were excluded from the computation of summary statistics since high-risk areas were sampled preferentially.

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
% lead SL	3.27	8.34	8.56	5.48	9.95	9.51	6.46	6.88	7.07
% built year < 1940	17.55	33.89	64.55	41.83	85.69	53.53	44.63	39.40	40.59
% living below poverty line	44.55	36.33	47.42	41.50	48.38	29.47	27.96	27.29	29.26
% parcels sampled	15.96	16.84	10.34	25.73	11.85	26.30	37.57	31.88	30.57
Number of samples	1517	1996	1164	1887	1262	2294	3287	3010	2343
% sentinel data (S rounds)	10.88	24.25	18.64	9.27	12.28	20.75	19.23	14.45	16.39
% sentinel data (X rounds)	1.05	3.76	2.66	0.74	2.14	5.10	4.69	3.26	6.10
90th percentile (μg/L)	6.0	17.0	10.0	9.0	17.0	19.0	13.0	16.0	11.0
% data > 15 μg/L	4.80	10.46	7.59	6.83	10.36	11.71	9.29	10.20	7.77
$\frac{\%\text{non-sentinel data}>15\ \mu g/L}{\%\text{sentinel data}>15\ \mu g/L}$	4.32	1.04	1.46	1.55	1.69	1.42	0.80	1.31	1.60

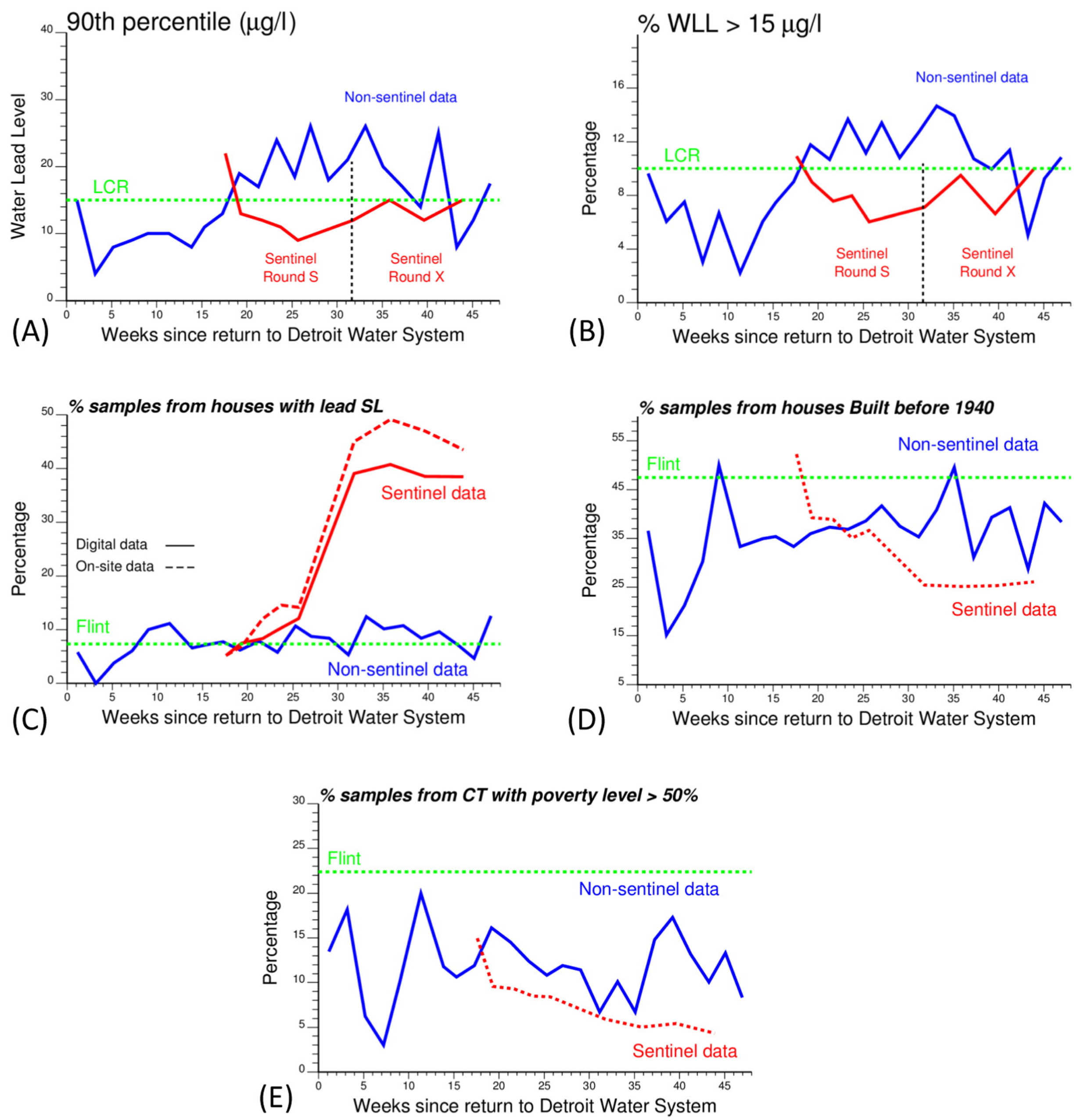


Fig. 5. (A,B) Time series of two statistics (90th percentile and % WLL data > 15 μg/L) computed from data collected during three sampling campaigns: voluntary (non-sentinel), sentinel and extended sentinel site programs. Results were aggregated within 14-day non-overlapping windows. Housing characteristics (presence of lead SL, pre-1940 construction year) and census-tract (CT) poverty levels are plotted using the same time scale for each type of sampling sites.

trends: higher percentages in presence of lead SL (14.04%) and in dwellings that were built prior to 1940 (12.32%). The impact of poverty level is mixed as one of the most impoverished Ward (Ward #1, poverty level = 44.55%) had also the fewest pre-1940 houses (17.55%); see Table 3. Trends exhibited by sentinel data are less intuitive although smaller sample sizes caused some percentages to be less reliable (e.g., only 35 samples collected during the extended sentinel site program are located in census tracts with poverty levels above 50%). Nevertheless surprising rates, such as 10.82% of samples collected in post-1960 houses during the sentinel program (rounds S1–S5) exceeded 15 μg/L, were computed from 194 water tests. Similarly, the rate of 5.48% of WLL above 15 μg/L for houses located in the poorest neighborhoods was based on 292 water samples.

GEE regression was used to test the statistical significance of these differences, as well as the existence of a linear temporal trend (Table 5). The analysis was extended to four additional thresholds ($z_c = 1$, 10, 25 and 50 μg/L) and results are expressed in terms of odds ratio (OR), which represent the odds that the outcome (exceedance of a threshold) will occur given a particular exposure (e.g., lead SL), compared to the odds of the outcome occurring in the absence of that exposure (e.g., non-lead or unknown SL). For the temporal trend ORs can be interpreted as the change in odds for a one unit change in this covariate (i.e., week), holding all other covariates constant. For example, the odds of exceeding 50 μg/L increased on average by 1.1% weekly (relative rate) since the return to DWSD.

For $z_c = 1$ μg/L, all three categorical covariates (presence of lead SL, construction year, poverty level) have odds ratios that are highly statistically significant ($\alpha = 0.01$). The presence of lead SL triples the likelihood of measuring WLL above 1 μg/L compared to other types of SL. Similar odds were observed for pre-1940 houses compared to post-1960 houses. The increase in odds is 50% for SLs of unknown composition and houses built from 1940 to 1960. As the threshold increases,

Table 4
Characteristics of the complete and sampled sets (non-sentinel, sentinel and extended sentinel sites) of residential tax parcel units: composition of service lines (on-site data are used for sentinel sites), built year, and poverty level. The last three columns report the percentage of data above 15 μg/L in each category of sampled data.

		Samples			WLL data > 15 μg/L		
	Parcels	Non-sentinel	Sentinel	Extended sentinel	Non-sentinel	Sentinel	Extended sentinel
Total number	51,045	14,962	3123	675	9.23	7.94	8.30
Service lines (%)							
Lead	7.32	7.43	10.95	46.22	14.04	11.70	7.37
Other	69.19	75.47	84.60	53.04	8.34	7.38	9.22
Unknown	23.49	17.10	4.45	0.74	11.06	9.35	0.0
Built year (%)							
<1940	47.46	35.92	38.52	25.48	12.32	7.40	10.47
1940–1959	41.56	49.18	55.27	61.48	8.03	8.00	8.19
≥1960	10.98	14.90	6.21	13.04	5.74	10.82	4.55
Poverty level (%)							
<25%	16.05	29.57	31.12	45.93	8.81	8.95	11.61
25–50%	61.57	58.53	59.53	48.88	9.62	7.80	5.76
>50%	22.39	11.90	9.35	5.19	8.37	5.48	2.86

the predictive power of the presence of lead lines decreases substantially and the odds ratio for type of SL becomes non-significant for WLL thresholds of 25 and 50 μg/L (Table 5). The impact of built year decreases to a much lesser extent than the type of SL to become the dominant and only significant covariate at higher thresholds. This result suggests that higher lead levels likely originate from lead fixtures and pipes present within old houses (premise plumbing) as opposed to lead service lines bringing water from street main water breaks to the property.

Poverty level impacts significantly only the odds of exceeding 1 μg/L and 10 μg/L ($\alpha = 0.01$) and 15 μg/L ($\alpha = 0.05$). The fact that significant odds ratios are all < 1, as well as the decrease in odds as the poverty level increases, confirm the mixed effect inferred from Table 4. Once again, this might be an artifact of Ward #1 that combines both the lowest 90th percentile and the third highest poverty level (Table 3). The temporal trend is the only covariate that is never statistically significant at $\alpha = 0.01$ which is expected given the non-linear shape of temporal trends displayed in Figs. 2 and 3.

3.5. Sentinel versus non-sentinel sites

Besides diverging spatial and temporal distributions (Sections 3.2 and 3.3), the type of sampling site also influences the impact of housing characteristics on WLL; see discussion of Table 4. Its overall impact on WLLs was tested for thresholds ranging from 1 to 50 μg/L through the incorporation of an indicator term (presence/absence of sentinel sites) in the regression model. No distinction was made between sentinel and extended sentinel sites since the model adjusts for the presence of LSL and the X1X4 sites are largely a subset of the original set of sentinel sites (S1–S5).

Table 5
Odds ratio for covariates of the GEE regression models fitted to five different WLL thresholds using all the data collected between 10/16/2015 and 9/15/2016.

Effects	WLL thresholds				
	1 μg/L	10 μg/L	15 μg/L	25 μg/L	50 μg/L
SL: lead vs others	2.929**	1.687**	1.582**	1.207	1.039
SL: unknown vs others	1.449**	1.148*	1.125	1.082	0.985
Built year: <1940 vs 1960–2016	3.143**	2.171**	2.106**	2.195**	2.232**
Built year: 1940–1959 vs 1960–2016	1.318**	1.426**	1.377**	1.348*	1.398*
Poverty: 25–50% vs < 25%	0.877**	0.968	1.021	1.040	1.056
Poverty: >50% vs < 25%	0.600**	0.760**	0.786*	0.817	0.835
Δt source switch (week)	1.001	1.001	1.002	1.006	1.011*
Sentinel site: No vs Yes	0.879*	1.331**	1.480**	1.712**	2.055**

* Significantly different from 1 at $\alpha = 0.05$.
** Significantly different from 1 at $\alpha = 0.01$.

The odds ratios in Table 5 indicate that the probability of exceeding 10 μg/L and higher is significantly larger at non-sentinel sites and this discrepancy increases with the WLL threshold. This result confirms the interpretation of time series in Figs. 2 and 3, even after accounting for housing characteristics. The opposite trend is observed for 1 μg/L (OR = 0.879), which is in agreement with the time series of Fig. 3A where except for weeks #23 to #27 (3/21/2016 to 4/23/2016) the percentage of WLLs above 1 μg/L was always greater at sentinel sites relative to non-sentinel sites.

Given the inclusion of main housing characteristics (i.e., age of the house and presence of lead service lines) in the regression models, the significant contribution of the type of sampling site suggests the potential influence of the sampling protocol on the results. According to sampling instructions a first-draw water sample should be collected from a cold water faucet in the kitchen or bathroom following a minimum period of 6 h when no water was used in the home (e.g., early morning or after return from work). The instruction that the sampled tap should not be flushed before sample collection could however be interpreted as no pre-stagnation flushing or no pre-draw flushing. A first-draw collected with pre-stagnation flushing (typically 5 min.) corresponds to water from the main (no lead component) which then stagnates close to the tap. This would thus represent the premise plumbing contribution to water lead levels, which is expected to be larger in older homes. A first-draw sample collected without pre-stagnation flushing (correct procedure, LCR) corresponds to water from various sources depending on water use in the house prior to stagnation (e.g., service lines, plumbing, main contribution to WLL) which then stagnates close to the tap (premise plumbing contribution to WLL). This would lead to higher concentrations relative to the first case. Pre-draw flushing would lead to sample water from the main, resulting in smaller readings. The last scenario would be to run the tap briefly before taking a sample (i.e., second-draw samples). In that case, water could come from premise plumbing or even lead service lines depending on how long the tap was ran and length of pipes. At this stage and given the information currently available, it is unclear whether lead in drinking water has been under-estimated at sentinel sites or over-estimated at non-sentinel sites. What is clear is that both sampling campaigns generated results that are statistically different and, as illustrated in Figs. 2 and 3, these discrepancies could influence whether there is compliance or not with the Lead and Copper Rule.

3.6. Modeling temporal trends using joinpoint regression

According to GEE the likelihood of exceeding WLL thresholds $z_c = 1$ to 50 μg/L has increased (odds ratio > 1), albeit not significantly at $\alpha = 0.01$, over the 48 week sampling period. These temporal trends are however clearly non-linear (Figs. 2–3). A multi-segmented model

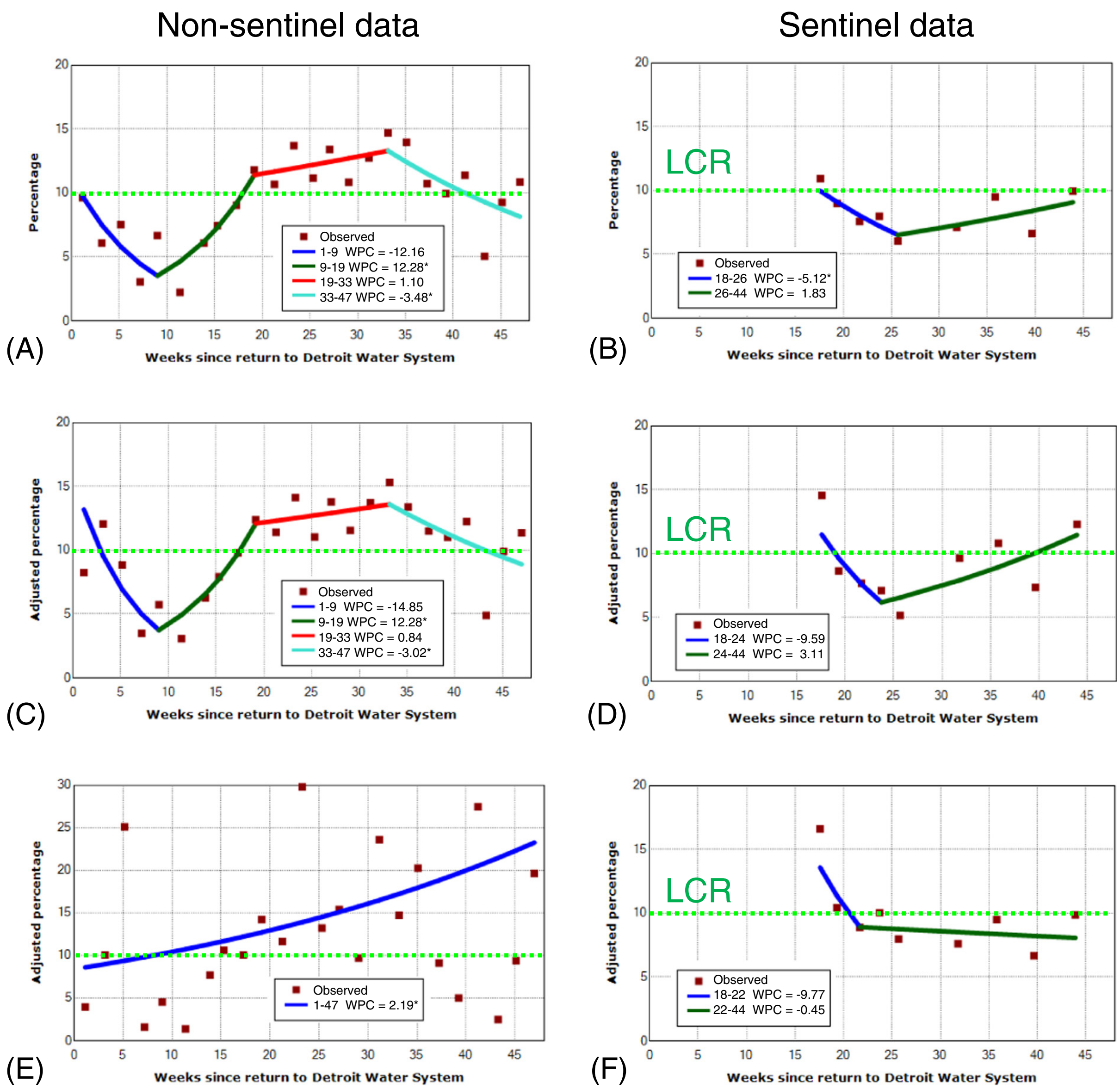


Fig. 6. Joinpoint regression models fitted to the time series of the percentage of WLL data recorded above 15 μg/L at non-sentinel and sentinel sites: (A,B) original time series, (C,D) percentages adjusted to reflect the proportions of houses with lead SL and construction year categories in Flint, and (E,F) percentages adjusted to meet the LCR target of 50% of houses with lead SL. The symbol * denotes weekly percent change (WPC) significantly different from zero at $\alpha = 0.05$.

(joinpoint regression) is thus more appropriate than the linear trend used in GEE regression.

Joinpoint regression was first used to model the bi-weekly time series of Fig. 5A. Sentinel and non-sentinel sites were modeled separately because of their distinct temporal trends (Fig. 6A,B). At non-sentinel sites the percentage of WLL data above 15 μg/L displayed four successive trends since the return to Detroit Water System: a decline until week #9 (12/14/2015–12/20/2015, when pre-stagnation flushing was removed from sampling instructions), followed by a significant increase until week #19 (2/22/2016–2/28/2016) when it crossed the LCR action level of 10% (green horizontal dashed line) and started to rise more slowly, before finally declining around week #33 (5/30/2016–6/5/2016). The much lower percentage (5%) recorded mid-August (weeks #43–44) must be interpreted with caution since it was computed from 139 samples. The shorter time series formed by the nine sentinel sampling rounds was modeled by a decreasing trend until the start of the extended sentinel site program that saw an increase in the percentage of WLL data above 15 μg/L. This later trend is however clearly an artifact of this new sampling program targeting high-risk areas, in particular houses with lead SL (Table 1).

To correct for biased sampling in both space and time, original time series were adjusted using Eq. (2) and two types of reference proportions for LSLs and construction years: 1) the global proportions displayed by residential parcels in the City of Flint (Table 4, 2nd column), and 2) the sample proportions observed during sentinel round X2, which was the closest to the 50% of lead SLs recommended by the Lead and Copper Rule (49.16%, Table 1). The impact of the first correction was minimal for the time series of non-sentinel data (Fig. 6C) since Flint housing stock is well represented in the voluntary program; major changes included larger adjusted percentages in the first few weeks to correct for the under-sampling of houses with lead SLs (Fig. 5C) and in the last few weeks leading to a reduced rate of decline (WPC). The impact of this correction was more noticeable for the sentinel sampling program which has targeted fewer and fewer pre-1940 houses over-time (Fig. 5D). Because older houses tend to have higher percentages of WLL above 15 μg/L, in particular in the extended sentinel site program (Table 4), the adjusted time series is higher and exceeds the LCR action level of 10% towards the end (Fig. 6D).

The second type of correction increased the weight assigned to the lead SLs covariate compared to construction year. For non-sentinel data (Fig. 6E) the time series is much more erratic because the percentages computed from <10% of houses with LSLs (Fig. 5C) now weigh as much as the other 90% of houses sampled during the voluntary program. Since this correction is based on proportions derived from the extended sentinel site program, it had little impact on results for rounds X 1X4 (Fig. 6F); it however increased the rates for the sentinel sampling program (rounds S1–S5) where only 6 to 14% of sampled houses had LSLs (Table 1). Under this scenario, the percentage of WLLs above 15 μg/L has been declining since the start of the sentinel sampling program.

3.7. Spatial variability in temporal trends

Although the LCR threshold of 10% applies to Flint public water systems as a whole, it is worth looking at temporal trends in different parts of the city, in particular in the light of differences between ward-level

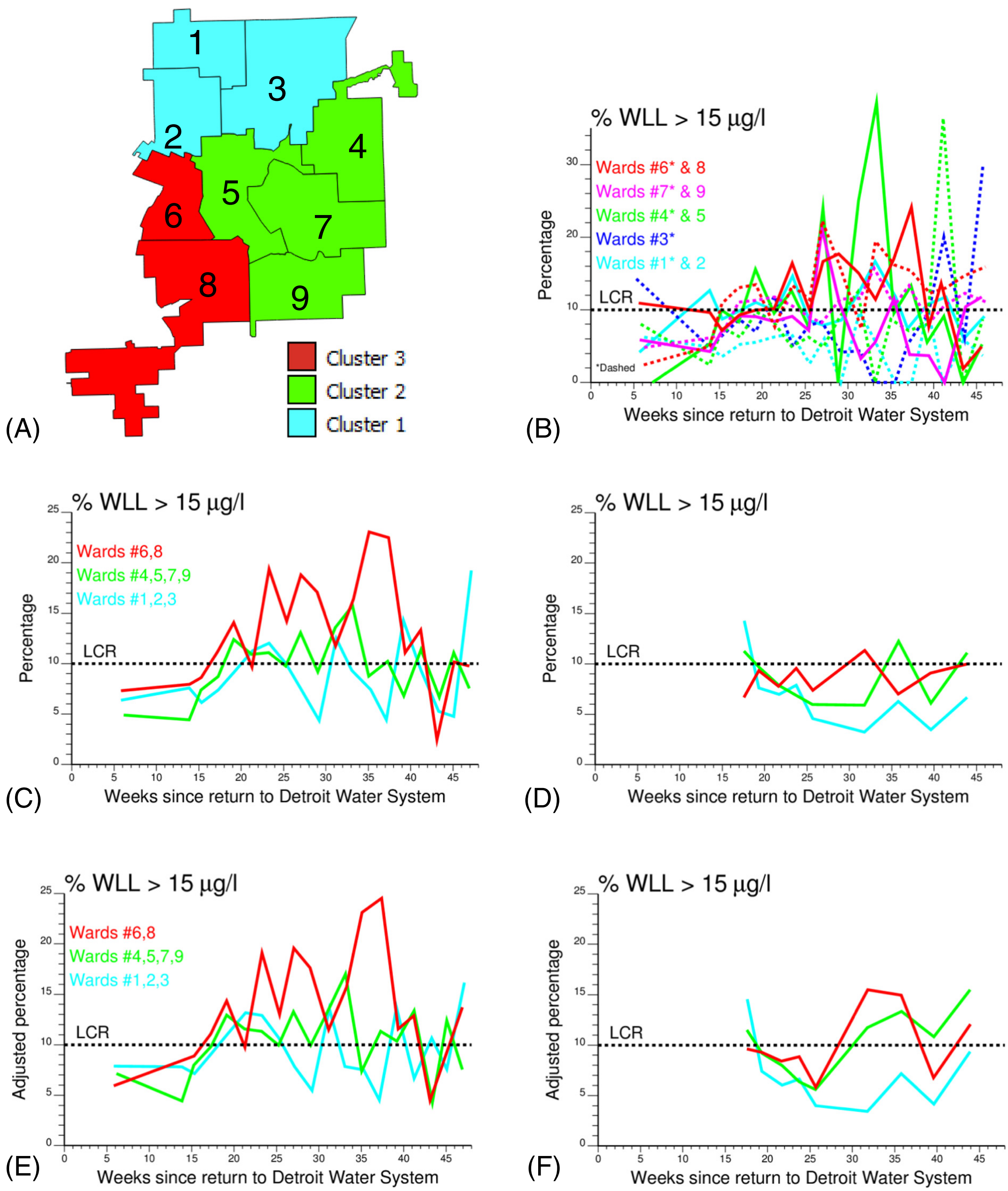


Fig. 7. Time series of the percentage of WLL data above 15 μg/L recorded within each ward (A, B). To facilitate visualization of trends and differences between voluntary (non-sentinel) and sentinel sites wards were aggregated within three clusters on the basis of odds ratio of the temporal trend in GEE model fitted within each ward. The original time series (C, D) were also adjusted to reflect the proportions of houses with lead SL and construction year categories in each spatial cluster (E, F).

Please cite this article as: Goovaerts, P., The drinking water contamination crisis in Flint: Modeling temporal trends of lead level since returning to Detroit water system, Sci Total Environ (2016), http://dx.doi.org/10.1016/j.scitotenv.2016.09.207

statistics reported in Table 3 and discussed in Section 3.2. Small sample sizes however make the time series very erratic even after aggregating all data collected before week #9 and combining both sentinel and non-sentinel data (Fig. 7B). To facilitate the visualization of temporal trends, results needed to be aggregated within geographical units of larger size.

GEE regression conducted within each ward (results not shown) highlighted the spatial variability in temporal trends for the action level threshold of 15 μg/L. Mapping the odds ratio for this regression term revealed a clear spatial pattern with lower rate of increase (OR < 1.01) in the Northern part of the city (Wards 1, 2 & 3), while OR above 1.02 were found in Wards 6 and 8 in South-West (Fig. 7A). The other wards formed a third cluster (OR = 1.01–1.02). These empirical clusters were used to aggregate the individual time series. Larger sample sizes also allowed plotting temporal trends for non-sentinel and sentinel sites separately. All three time series for non-sentinel sites are still highly variable but a similar pattern emerges: an initial increase, followed by a plateau and a recent decline (Fig. 7C). The important fact is that the action level of 10% (LCR) was exceeded to a greater extent and for a longer period of time in Cluster #3 compared to Clusters #1 and 2. It is noteworthy that all three time series tend to converge towards the end of the sampling period, which suggests a possible homogenization of water quality across the city. The shapes of temporal trends differ the most for sentinel data, from a slow decline for Cluster #1 (Fig. 7D, blue curve) to a flat trend for Cluster #3 (Fig. 7D, red curve). The ranking of clusters in terms of average percentages above 15 μg/L is however similar for both types of sampling sites.

Adjusting for ward-specific percentages of houses in three construction year categories and lead SLs had a negligible impact on the time series of non-sentinel data (Fig. 7E), which confirms results observed city-wide in Fig. 6. The impact of this correction was much greater for the sentinel sampling program, in particular for Cluster #2 which now displays an increasing trend with adjusted percentages above 10% for the entire extended sampling program. This result highlights the need to look beyond city-wide statistics and investigate water quality at the ward level.

4. Conclusions

The mechanisms leading to water crises, such as in Flint or Washington DC 15 years ago, are now well understood. Each crisis was triggered by changes in water chemistry leading to the destabilization of lead-bearing mineral scales that coated lead service lines and the corrosion of lead-bearing solder, pipes, faucets, and fixtures. These lead and iron particles were then released into drinking water and their presence was either overlooked or hidden through faulty testing and inadequate monitoring procedures. What is less clear is how long it will take for the protective scale layer to rebuild inside the pipes and for the water quality to get back to levels suitable for human usage. A first step is obviously for Flint drinking water to be in compliance with the Lead and Copper Rule (LCR) that states that no more than 10% of samples collected from high-risk homes can exceed the action level of 15 μg/L for lead and 1.3 mg/L for copper.

The large database of >24,000 water lead samples collected since Flint returned to its pre-crisis source of drinking water offered a unique opportunity to explore the spatial and temporal variability of WLL in a city water supply network following a major intervention. The collection of such a large number of data over an eleven month period was made possible by the involvement of concerned citizens who decided to acquire a testing kit and conduct sampling on their own. This type of crowd sourcing was supplemented by a State-controlled monitoring program that started four months later and proceeded in two phases: 1) a sentinel program to determine the general health of the distribution system and to track changes in lead concentrations over time, and 2) an extended sentinel site program to target specifically the highest-risk areas necessary to verify LCR compliance. In all cases, water samples were collected by homeowners following written instructions provided with the testing kits or on-site training given by sentinel teams.

A key finding of the present study was that even after adjusting for housing characteristics samples collected at sentinel sites exceeded water lead levels of 15 μg/L and higher at a statistically significant lower rate than samples collected on a voluntary basis by homeowners (non-sentinel sites). Given that residents at sentinel sites were trained to draw water samples in a scientifically accurate manner, one hypothesis is that discrepancies between the two types of sampling programs could have been caused by differences in the implementation of the sampling protocol. Information available in this study however did not allow to support or reject this hypothesis conclusively. This issue needs to be carefully investigated as it influences conclusions regarding the improvement of water quality since the return to the pre-crisis source of drinking water; in particular whether Flint is in compliance with LCR or not.

The exploratory data analysis revealed some counter-intuitive relationships for sentinel data, such as lower percentages of WLL data above 15 μg/L found for houses with lead SLs (extended sentinel site program) or for pre-1940 houses (sentinel program). Such patterns were not observed for data collected during the voluntary sampling program. Poverty level is another covariate whose influence on water lead levels differed between sentinel and non-sentinel sampling programs, with some of the largest differences between the two datasets taking place in wards with the greatest percentage of habitants living below the poverty line. Although these could simply be sampling fluctuations caused by the smaller size of the sentinel dataset, any "abnormal" behavior should warrant further investigation.

The study of non-sentinel data had its own caveats, such as relying on potentially inaccurate digital data for assessing the presence of lead service lines. The comparison of digital and on-site SL data at sentinel sites however demonstrated the reliability of this information: most discrepancies were caused by a lack of information (i.e., unknown material) rather than lead SLs being mistakenly classified as non-lead, and vice versa. Moreover, non-sentinel data represent 80% of the database and inform on a much longer period than the sentinel program. These data will thus be critical to assess past exposure to lead in drinking water, which is another reason to investigate discrepancies between both sampling programs as they could lead to vastly different estimates.

GEE regression conducted for WLL thresholds ranging from 1 to 50 μg/L provided important information on the relative influence of housing characteristics. The built year (e.g. pre vs post-1940) turned out to be the main risk factor, whereas the presence of lead service lines is mostly important for smaller thresholds (e.g. 1 μg/L). This result suggests that higher lead levels in these first-draw samples likely originate from lead fixtures and pipes present within old houses (premise plumbing) as opposed to lead service lines bringing water from street main water breaks to the property. The fact that fewer pre-1940 houses were sampled by the sentinel program compared to non-sentinel sites and Flint housing stock in general could explain the smaller percentage of WLLs above 10 μg/L and higher recorded at sentinel sites while the opposite trend was observed for 1 μg/L. Priority should thus be given to expanding the pool of pre-1940 houses in future sentinel sampling programs.

The characterization of lead levels in Flint drinking water is an ongoing matter as hundreds of new data are being posted online every week. These additional data are needed to confirm the recent decline in lead levels observed for non-sentinel sites. Encouraging news is that similar findings were shared by the Virginia Tech Team during a recent press conference (FlintWaterStudy.org, 2016). Based on an independent sampling of 162 houses conducted in July 2016, they found that compared to March 2016 (month included in the present analysis) the number of non-detectable lead samples increased from 37 to 45% while the 90th percentile declined from 22.5 to 13.9 μg/L, which is below the LCR action level but still above the target of 10 μg/L set by the State. In addition, this set of 162 houses, like the set of non-sentinel

Please cite this article as: Goovaerts, P., The drinking water contamination crisis in Flint: Modeling temporal trends of lead level since returning to Detroit water system, Sci Total Environ (2016), http://dx.doi.org/10.1016/j.scitotenv.2016.09.207

sites analyzed in this paper, is not an approved LCR sampling pool as <50% of lead SLs was sampled. One should thus expect higher concentrations to be recorded in high-risk areas.

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. The constructive comments of six anonymous reviewers are also gratefully acknowledged. 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

Agency for Toxic Substances and Disease Registry, 2010. Case Studies in Environmental Medicine (CSEM) Lead Toxicity. Course: WB 1105 Original Date August 15. US Department of Health and Human Services, Public Health Service, Atlanta.

Calley, B., 2016. Understanding recent flint water test results. 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., 2011. Investigating dissolved lead at the tap using various sampling protocols. J. Am. Water Works Assoc. 103, 55–67.

Clark, B.N., Masters, S.V., Edwards, M.A., 2015. Lead release to drinking water from galvanized steel pipe coatings. Environ. Eng. Sci. 32 (8), 713–721. http://dx.doi.org/10.1089/ees.2015.0073.

Clegg, L.X., Hankey, B.F., Tiwari, R., Feuer, E.J., Edwards, B.K., 2009. Estimating average annual percent change in trend analysis. Stat. Med. 28, 3670–3682.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20 (1), 37–46.

Del Toral, M.A., Porter, A., Schock, M.R., 2013. Detection and evaluation of elevated lead release from service lines: a field study. Environ. Sci. Technol. 47 (16), 9300–9307.

Deshommes, E., Prévost, M., Levallois, P., Lemieux, F., Nour, S., 2013. Application of lead monitoring results to predict 0–7 year old children's exposure at the tap. Water Res. 47 (7), 2409–2420.

Edwards, M., Dudi, A., 2004. Role of chlorine and chloramine in corrosion of lead-bearing plumbing materials. J. Am. Water Works Assoc. 96 (10), 69–81.

Edwards, M., Triantafyllidou, S., Best, D., 2009. Elevated blood lead levels in young children due to lead-contaminated drinking water: Washington, DC, 2001–2004. Environ. Sci. Technol. 43 (5), 1618–1623.

Fleiss, J.L., 1981. Statistical Methods for Rates and Proportions. 2nd ed. John Wiley, New York.

Flint Safe Drinking Water Task Force, February 2016. Recommendations on MDEQ's Draft Sentinel Site Selection. 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, 2016. Report. March Retrieved from https://www.michigan.gov/documents/snyder/FWATF_FINAL_REPORT_21March2016_517805_7.pdf (on June 16, 2016).

FlintWaterStudy.org, 2015. Lead results from tap water sampling in Flint, MI during the Flint Water Crisis. Retrieved from http://flintwaterstudy.org/2015/12/complete-dataset-lead-results-in-tap-water-for-271-flint-samples/ on August 28, 2016.

FlintWaterStudy.org, 2016. Flint Water Press Conference. Available at http://flintwaterstudy.org/2016/08/flint-water-press-conference-august-11-2016/.

Gold, R., January 28, 2016. Retrieved from https://news.umflint.edu/2016/01/28/10668/.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.

Goovaerts, P., 2013. Analysis of geographical disparities in temporal trends of health outcomes using space-time joinpoint regression. Int. J. Appl. Earth Obs. Geoinf. 22, 75–85.

Hanna-Attisha, M., LaChance, J., Sadler, R.C., 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. Am. J. Public Health 106, 283–290.

Hoekstra, E.J., Hayes, C.R., Aertgeerts, R., Becker, A., Jung, M., Postawa, A., Russell, L., Witczak, S., 2009. Guidance on Sampling and Monitoring for Lead in Drinking Water. European Commission, Joint Research Centre, Institute for Health and Consumer Protection, Luxembourg.

Kim, H.J., Fay, M.P., Feuer, E.J., Midthune, D.N., 2000. Permutation tests for joinpoint regression with applications to cancer rates. Stat. Med. 19, 335–351 (correction: 2001;20:655).

Kim, H.J., Yu, B., Feuer, E.J., 2009. Selecting the number of change-points in segmented line regression. Stat. Sin. 19 (2), 597–609.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174.

Lee, R.G., William, C.B., David, W.C., 1989. Lead at the tap: sources and control. J. Am. Water Works Assoc. 81 (7), 52–62.

Lerman, P.M., 1980. Fitting segmented regression models by grid search. Appl. Stat. 29, 77–84.

Liang, K.Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.

Milman, O., 2016, Jnuary 25a. Flint Rewrites Water Testing Directions Blamed in Lead Pollution Crisis. The Guardian Retrieved from http://www.theguardian.com.

Milman, O., 2016b. Michigan Removes ‘Pre-flushing’ Practice From State Water Testing Rules. The Guardian January 27. Retrieved from http://www.theguardian.com.

Milman, O., 2016, June 2c. Tests on Flint Water Targeted Homes Far From Network of Lead Pipes. The Guardian Retrieved from http://www.theguardian.com.

Milman, O., Glenza, J., 2016, June 2. At Least 33 US Cities Used Water Testing ‘Cheats’ Over Lead Concerns. The Guardian Retrieved from http://www.theguardian.com.

Renner, R., 2009. Out of plumb: when water treatment causes lead contamination. Environ. Health Perspect. 117 (12), A542–A547.

SAS Institute Inc., 2011. SAS/STAT 9.3 User's guide. SAS Institute Inc., Cary, NC.

Schock, M.R., 1990. Causes of temporal variability of lead in domestic plumbing systems. Environ. Monit. Assess. 15, 59–82.

Snyder, R., July 8, 2016. Extended Water Testing in Flint Continues showing Positive Signs. Press release. Available at https://votesmart.org/candidate/public-statements/124011/rick-snyder.

State of Michigan, August 5, 2016. Flint Water System Test Results Show Continued Improvement: Additional Actions to Follow. Retrieved from http://www.michigan.gov/som/0,4669,7-192-47796-390808–,00.html.

US Environmental Protection Agency, Office of Water, Lead and Copper Rule 40 CFR Part 141 Subpart I, 1991a. Available at https://www.epa.gov/dwreginfo/lead-and-copper-rule (Accessed July 18, 2016).

US Environmental Protection Agency, 2002a. Lead and Copper Monitoring and Reporting Guidance for Public Water Systems. Office of Water. Available at https://www.epa.gov/dwreginfo/lead-and-copper-rule-compliance-help-public-water-systems Accessed August 30, 2016.

US Environmental Protection Agency, 2002b. Effect of water age on distribution system water quality. Office of Ground Water & Drinking Water. Available at https://www.epa.gov/sites/production/files/2015-09/documents/2007_05_18_disinfection_tcr_whitepaper_tcr_waterdistribution.pdf Accessed August 30, 2016.

US Environmental Protection Agency, Office of Ground Water & Drinking Water, 2016r. Memorandum: clarification of recommended tap sampling procedures for purposes of the Lead and Copper Rule. Available at https://www.epa.gov/dwreginfo/memo-clarifying-recommended-tap-sampling-procedures-lead-and-copper-rule Accessed August 30, 2016.

Wang, Z., Devine, H., Zhang, W., Waldroup, K., 2014. 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. 140 (9), A4014004.

Please cite this article as: Goovaerts, P., The drinking water contamination crisis in Flint: Modeling temporal trends of lead level since returning to Detroit water system, Sci Total Environ (2016), http://dx.doi.org/10.1016/j.scitotenv.2016.09.207