

IN DETAIL

The murky tale of Flint's deceptive water data

When children in Flint, Michigan showed signs of lead poisoning, residents rightly suspected their tap water was to blame. Authorities denied the fact for months, but the official water test data was misleading – so citizens fought back with statistics of their own. By **Robert Langkjær-Bain**



The water crisis in Flint, Michigan is a human tragedy. In the past three years, thousands of adults and children have been exposed to poisonous lead in their drinking water. The toxic heavy metal is especially harmful to children, causing behaviour and learning problems, slowed growth and iron deficiency. The effects can be irreversible. But for months, officials insisted the water was safe – pointing to test results that proved it.

However, independent investigators and state prosecutors say that data was manipulated, and charges have been brought against the officials thought responsible.

The story of the Flint crisis offers a cautionary tale of how flawed and inadequate testing creates misleading data. It also illustrates how citizens and scientists can work together to help right wrongs.

Sick city

Flint is the second poorest city of its size in the United States and has spent six of the past 15 years in a state of financial emergency. One of the cost-cutting measures taken by emergency managers was to stop buying water, sourced from Lake Huron, from the Detroit Water and Sewerage Department. Instead, Flint would use the Flint River for its water supply while waiting for a new pipeline to Lake Huron to be opened. The move was expected to save roughly \$5 million over a period of two years.

The Flint River supply was switched on in April 2014. Not long after, problems arose.

Flint resident and mother of four LeeAnne Walters noticed that the water coming out of her taps was orange. More worryingly, her family's hair was falling out, her preschool sons had broken out in rashes and one of them had stopped growing.

The orange colour was from iron, but the family's symptoms pointed to a far more dangerous contaminant: lead.

Lead is still widespread in the USA's ageing infrastructure, so water companies routinely add chemicals to water to prevent pipes from corroding and leaching lead into drinking water. They also regularly monitor levels of lead in water from a sample of homes. While there is no completely safe amount of lead consumption, the limit allowed by the Lead and Copper Rule (LCR) of 1991 is 15 parts per billion (ppb). If this is exceeded in more than 10% of homes tested (or if the 90th percentile value of the total sample is above 15 ppb), action is required. And to make sure problems are caught, sampling for lead in water is supposed to target the "worst-case" homes – those in areas served by lead pipes.

Officials at the Michigan Department of Environmental Quality (MDEQ) insisted that lead levels in Flint's water were below thresholds. But when a city employee tested two samples of water from Walters's home, one had lead concentrations six times higher than the 15 ppb "action" level, at 104 ppb; the other was 25 times higher at 397 ppb.

Walters was advised to stop drinking the water immediately, but the authorities still insisted it was not their problem, instead blaming the pipes in Walters's own home.

The true origin of the lead contamination only became clear months later – and it certainly was not from the plastic



Robert Langkjær-Bain is a freelance journalist. He was previously editor of *Lux* magazine and deputy editor of *Research* magazine

pipes on the Walters's property. The water of the Flint River is naturally high in chloride, which corrodes pipes. And in Flint's hurry to switch to the new supply, no corrosion-controlling chemicals were added (a failure for which state officials are now facing felony charges).

But long before this was known, Walters was intent on proving hers was not an isolated case. "I decided we needed to get to the science if anyone was ever going to believe us," Walters later told a congressional hearing (youtu.be/t6687jtuyv8). "I started researching and educating myself about water." To get to the bottom of the story, she says, she had to become a "water warrior".

Walters contacted the federal Environmental Protection Agency (EPA) and had an expert test her water supply. He was shocked by the amount of lead and made his concerns clear in a preliminary report dated June 2015 (bit.ly/2laMPP9). But when the findings were made public, an MDEQ spokesman told National Public Radio that the report was the work of a "rogue employee" (n.pr/2laJAHv). The advice to Flint citizens from that same spokesman? "Relax" (bit.ly/2laVv7X).

A desperate Walters also turned to Marc Edwards, an environmental engineering expert at Virginia Tech University. When she sent him a water sample from her tap, Edwards says it was the worst he had ever seen, with lead levels of 13 200 ppb – high enough to qualify as hazardous waste.

To Edwards, this was all depressingly familiar. He had spent much of the last decade investigating water contamination in Washington, DC, where high levels of lead were found in 2001. The lack of accountability among those responsible in Washington made a repeat of the scandal inevitable, he believed. And here it was, in Flint.

People power

In 2015, when Edwards became involved in Flint, the MDEQ was still claiming everything was fine – their testing showed that the crucial 90th percentile reading for lead in their own samples of tap water was just below the limit of 15 ppb. Walters and Edwards were suspicious and thought there must be something wrong with the official data. They resolved to collect samples of their own.

A research team of scientists and citizens gave out water sampling kits to hundreds of residents and explained how to take samples and send them to Virginia Tech for analysis.

Asking citizens to collect water samples themselves is not unusual – it is the same method used by the authorities. But Virginia Tech's scientists had no access to information on where lead pipes were located, so they could not target worst-case homes. They had to make do with a self-selecting sample of concerned citizens who had volunteered to take part. ▶

This is a cautionary tale of how flawed and inadequate testing creates misleading data

► The citizens of Flint were determined to ensure they could not be accused of manipulating the samples. The research team noted (bit.ly/2laLkXj) how participants “developed and implemented procedures on their own to minimize the likelihood of someone tampering with them, or to counter accusations that they tampered with the samples. ... For example, they developed a way of having each homeowner seal and sign the kit, so that no one but the homeowner could have opened the bottle before we check the seal.”

Of the 300 water-testing kits distributed, 271 samples were returned – a response rate many times higher than the city managed in its own testing. They were stunned by what they found: 45 samples exceeded 15 ppb, representing almost 17% of all samples, far above the 10% threshold for action. The 90th percentile value was 26.8 ppb, and the highest individual sample came in at 158 ppb. How much worse would the results have been had the researchers identified the worst-case homes? And why had the city’s own testing not picked up these levels of lead?

Sampling errors

Thanks to the tenacity of Flint’s citizen scientists, the Virginia Tech researchers, and activists and reporters who helped scrutinise the data and pressure the authorities for answers, we now know that there was a long list of failures in how the city and the MDEQ collected, analysed and handled their data.

To begin with, they did not target the worst-case homes. Despite claiming that all the homes they sampled had lead service lines, city officials later admitted that they did not know which homes in Flint were served by lead pipes and which were not (bit.ly/2laPOXU).

The city, on the advice of the MDEQ, also employed sampling procedures that were likely to underestimate the amount of lead in the water supply. In instructions sent to households (bit.ly/2lyqW9W), residents were told: “Flush

the cold water for at least 5 minutes. Let the water sit for at least 6 hours before you plan to collect the sample.” But the EPA recommends against the first part of that instruction. “Pre-stagnation flushing” – as it is known – “may potentially lower” lead levels as flushing “removes water that may have been in contact with the lead service line for extended periods” (bit.ly/2lyvJZ3). Indeed, running water “until it becomes cold” is one of several ways that MDEQ advises home owners to reduce their risk of exposure to lead in water.

Miguel Del Toral, the “rogue” EPA scientist who tested Walters’s water, had voiced his concern about this sampling practice: he wrote in his preliminary report that pre-flushing “has been shown to result in the minimization of lead capture”. In fact, Del Toral co-authored a 2013 study that compared water samples taken at the same sites in Chicago but under different testing conditions (bit.ly/2m4FHVO). Under the “pre-flushing” (PF) condition, homeowners were told to flush their taps for five minutes prior to the six-hour stagnation period, while under the “normal household usage” (NHU) condition there was no PF requirement. The study found that lead results under NHU conditions were “numerically higher overall than the corresponding PF values for most sites, but the differences were not statistically significant”. However, it noted that normal household usage might involve activities like showering and washing dishes, which could clear lead from pipes in a similar way to pre-flushing. Thus, the researchers said, “it stands to reason that if the NHU activities were not undertaken, and a larger sample set were used, the NHU samples would yield results that were statistically higher than the corresponding PF samples”.

A more recent example of the impact of pre-flushing comes from New York, where changes to the protocols used to test water in public schools revealed a greater number of outlets with lead concentrations above the specified “action” level (nyti.ms/2kU3ZN6). Previously, the city had allowed two hours of pre-flushing the night before testing and found that only 1% of outlets in schools were above 15 ppb; after retesting without flushing, nine times as many outlets exceeded that level (based on results from a third of city schools; on.ny.gov/2kUerE7).

In his 2015 report, Del Toral noted that although pre-flushing “is not specifically prohibited by the LCR, it negates the intent of the rule”, which requires that samples are drawn from worst-case homes to ensure that measurements are representative of those most likely to be at risk of lead contamination. The inclusion of pre-flushed samples therefore undermines the statistical validity of the results, which are meant to be representative of real-world worst cases.

The problems did not end there. Sample bottles provided by the city had openings so small that they could only be filled from a gently running tap. This practice is also thought to reduce the amount of lead in samples, as slow-flowing water is likely to dislodge smaller amounts of material from pipes than fast-flowing water. Again, the EPA notes that: “To best approximate flows from taps in actual day-to-day use, samples should be collected from taps opened fully.”

Adjustments were also made to the number of samples collected. The LCR sets minimums on how many “sites” must

BELOW Marc Edwards and LeeAnne Walters, pictured during a House Oversight and Government Reform Committee hearing on 3 February 2016





iStock.com / LindaParton

be tested, based on size of population. For a city of more than 100 000 people, the regulations require 100 samples. For the first period of monitoring after switching to Flint River water (covering July to December 2014) the city reported 100 samples. But in the second monitoring period (January to June 2015) the city struggled to gather as many. Although Flint's population was over 100 000 at the time of the 2010 Census, population estimates since put it below that threshold, so it was decided that a minimum of 60 samples would suffice.

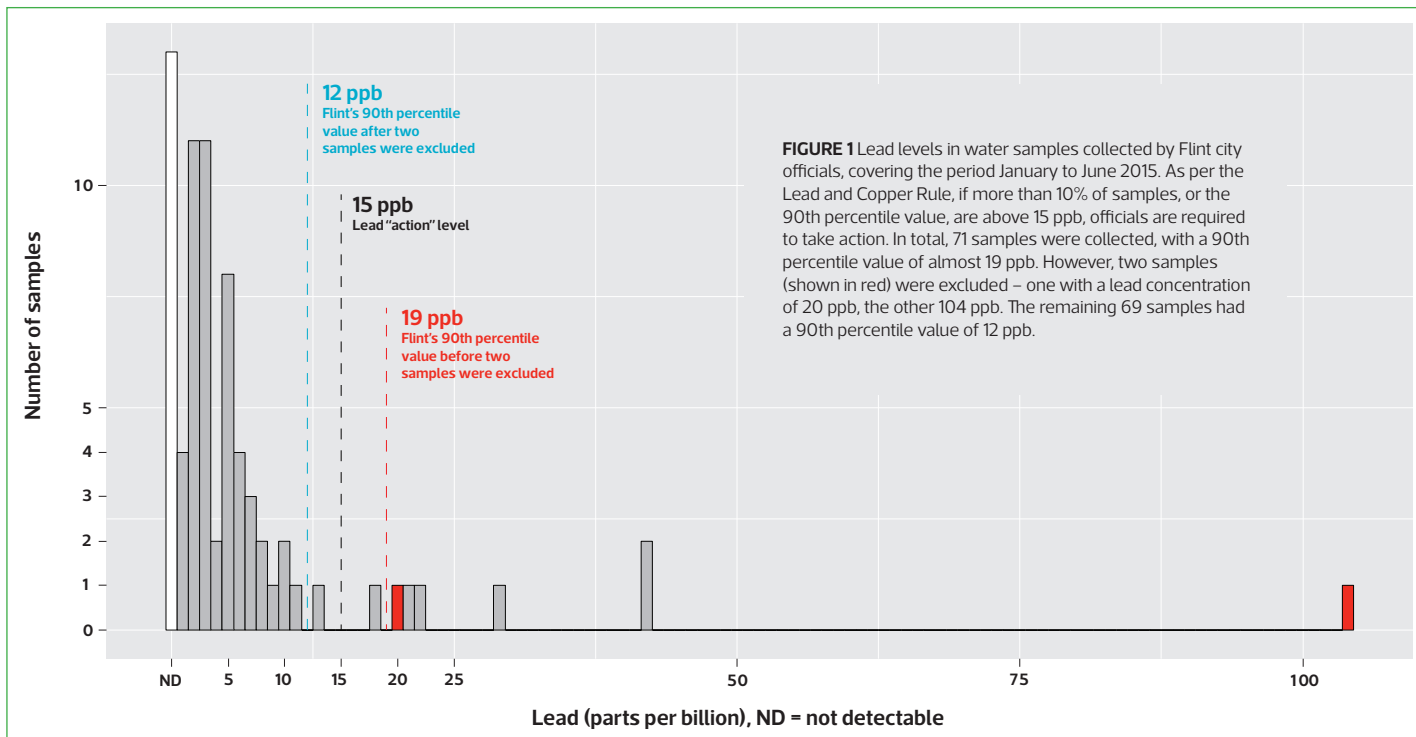
Seventy-one were eventually collected, but then two samples went missing. In the city's final report for the second monitoring period, only 69 samples were included. A Freedom of Information request later revealed that the MDEQ told the city's water quality supervisor to remove two of the samples that had come in over the "action" level (bit.ly/2IGCo6h). Why?

The MDEQ explained in an August 2015 email (quoted at bit.ly/2laD1Ey) that one of the samples – from LeeAnne Walters's home, measuring 104 ppb – was removed because the house was fitted with a filter, which meant it would not qualify as one of the worst-case homes. This explanation ignored the fact that Walters's house had one of the highest lead levels discovered (not to mention the fact that other steps to get a worst-case sample were not taken). The second sample, at 20 ppb, was thrown out because it was from a business rather than a home, said the MDEQ (bit.ly/2IGCo6h).

"There are a lot of statistical methods looking at whether an outlier should be deleted," says Barry Nussbaum, the president of the American Statistical Association. "I don't endorse any of them." Nussbaum knows plenty about

Citizens and rogues: Three times "unofficial" data proved its value

- **1715** – For the astronomer Edmond Halley, the total solar eclipse of 1715 was a once-in-a-lifetime opportunity to test his Newtonian predictions of exactly when and where it would fall. Halley had the Royal Observatory in Greenwich at his disposal, but he could not track the movement of the eclipse's shadow from just one location – he needed help. As he wrote afterwards: "I caused a small map of England, describing the track and bounds [of the eclipse] to be dispersed all over the Kingdom, with a request to the curious to observe what they could about it." Twenty-six observers in locations as far flung as Exeter, Anglesey and Dublin took measurements and sent them to Halley, allowing him to confirm that his original prediction had been pretty much spot on, give or take 20 miles.
- **2008** – In the build-up to the 2008 Olympic Games in Beijing, concerns were raised about air pollution, but no reliable data was available. So, US Embassy staff in the city installed an air quality monitor on their roof, which automatically tweeted readings every hour, making the data public on the internet for anyone to use. At first, the Chinese government complained the data was illegal, but it quickly backed down and revamped its air pollution policies. There are now monitors in place at US embassies and consulates in several other cities, and non-governmental organisations have taken similar steps in other countries.
- **2015** – How many people are killed each year by police in the United States? With only limited official data available, the *Guardian* and the *Washington Post* took it upon themselves to start keeping tabs, using crowdsourced data. A total of 1146 deaths were counted in 2015 – a figure that is shocking, even more so in the light of the failure of the authorities to comprehensively record and report these deaths. However, the Department of Justice recently issued new rules on the reporting of deaths occurring during interactions with law enforcement (see page 5, *Significance*, February 2017).



► environmental statistics – he served as chief statistician at the EPA until early last year (although he was not involved with the Flint case). He says: “Typically you do omit outliers if there is some measurement problem among the very high or low observations. But if it is a properly measured value, you must investigate further. You must look at those outliers and ask, what’s going on here? To just delete it because it’s an outlier is completely wrong.”

But outliers should not be a statistician’s only concern, says Nussbaum. There may be “inliers” too – samples that are affected by measurement problems but which go unnoticed because they happen to fall within the expected range of values. Perhaps some of the Flint samples with lead levels below 15 ppb were from homes with filters, or other businesses, or from those without lead service lines.

In any case, the fact that water samples from a home with a filter still showed dangerously high levels of lead should be cause for concern and further investigation. As for the business sample, a report by Michigan Radio paraphrased water quality expert and former EPA official Elin Betanzo as saying that “if a site is sampled – even if it’s not technically supposed to be in the sampling pool – the results should still be used to determine if the water supply is safe”. High levels of lead in drinking water are a concern, no matter where they are found. The same is true when finding a high lead value from a home with a filter, which should lower lead.

The exclusion of these two samples nudged the 90th percentile reading for the monitoring period below the all-important limit of 15 ppb (see Figure 1). Their inclusion would have forced the authorities to take a series of costly steps:

warning the public, putting corrosion control in place, and potentially replacing pipes.

Right and wrong

Eighteen months after the water supply was switched, the authorities admitted that the citizen scientists were right: contaminated water was poisoning Flint’s residents. In October 2015, the city switched back to the old Detroit Water supply from Lake Huron and, a short while later, a state of emergency was declared.

Hundreds of millions of dollars of state and federal funds have since poured in to Flint to replace pipes, provide bottled water and support those affected. The city’s new mayor has brought Marc Edwards on board to oversee water testing, and the water system now seems to be recovering: lead levels for the second half of 2016 fell back below the “action” level, with fewer than 10% of the 368 samples exceeding 15 ppb. The reported 90th percentile value was 12 ppb.

The problems with sampling that were highlighted in Flint also brought attention to similar issues elsewhere. An investigation by the *Guardian* newspaper found that 33 US cities were using questionable practices in their sampling that might lower the amount of lead detected (bit.ly/2laKnlc). The utilities involved have said they will change their methods, and the EPA is in the process of revising its rules on water testing and sampling.

It is hard to imagine how the Flint water activists could have been more thoroughly vindicated. But that will be of little comfort to those whose loved ones, friends and neighbours were harmed by months of lead exposure. *Significance* invited both the MDEQ and the EPA to comment on what went wrong in Flint and

what has been done to prevent it happening again, but neither responded to our questions. (In the EPA's case, communications were on hold on the orders of President Donald Trump.)

However, a report by the Flint Water Advisory Task Force, commissioned by Michigan state Governor Rick Snyder, lists a catalogue of failures and places primary responsibility for the crisis on the shoulders of the MDEQ. "MDEQ misinterpreted the LCR and misapplied its requirements," reads page 28 of the report (bit.ly/2lyFtLY). "As a result, lead-in-water levels were under-reported and many residents' exposure to high lead levels was prolonged for months." Moreover, it says MDEQ's guidance to Flint officials on the use of pre-flushing and small-mouthed bottles to collect samples, "while possibly technically permissible, was not designed to detect risks to public health".

The aftermath

The state of Michigan has brought 43 criminal charges against 13 state and local officials, including water engineers and water quality supervisors, as well as two of the emergency managers who oversaw the switch to the Flint River water supply. Charges include misconduct in office, tampering with evidence, wilful neglect of duty and various counts of conspiracy. Meanwhile, residents have launched a class action lawsuit against 14 people, including Governor Snyder himself.

Flint's new mayor Karen Weaver said in December that those in charge at the time could have prevented the disaster if they had only listened and acted. "But they didn't."

Amid the human tragedy, Marc Edwards is optimistic about the impact of the Flint story on science's relationship with the public.

He reckons that he and his team have spent close to \$300 000 on pursuing the truth about Flint – much of it thanks to a grant Edwards received from the MacArthur Foundation in 2007. But they have managed to recoup almost all of it thanks to an online fundraising campaign. It began with cheques for \$50 and \$100, then one day Edwards received a cheque for \$70 000. Worth most of all, though, were the handful of letters from residents of Flint containing one-dollar bills. "This is the second poorest city [of its size] in the US," Edwards says. "And the fact that we created this impression of science as a public good, when the relationship between science and society is in such dangerous disrepair ... it's truly priceless."

The Task Force report concluded that the Flint crisis was a story of official failure, but also one "of something that did work: the critical role played by engaged Flint citizens, by individuals both inside and outside of government who had the expertise and willingness to question and challenge government leadership".

The public's ability to prove that official data was mishandled certainly gives cause for optimism. It is the fact that they needed to do so in the first place that gives cause for fear. ■

Explore the Flint data at flintwaterstudy.org

Reference

1. Hanna-Attisha, M., LaChance, J., Sadler, R. C. and Champney Schnepp, A. (2016) Elevated blood lead levels in children associated with the Flint drinking water crisis: A spatial analysis of risk and public health response. *American Journal of Public Health*, 106(2), 283–290. bit.ly/1QpONEc



Mona Hanna-Attisha testifies at a Flint water crisis hearing, 10 February 2016

Getty Images/Gabriella Demczuk

Lead in the blood

Citizen science helped make the case that lead levels in Flint's water were dangerously high, but it was medical science that helped demonstrate an association between the change in water supply and raised levels of lead in the blood of Flint's children.

Michigan State University's Mona Hanna-Attisha and colleagues compared blood tests taken in 2013 to those taken in 2015 to see whether a greater number of children were exhibiting elevated blood lead levels (EBLL) since the Flint River came on tap. The reference level for EBLL was 5 micrograms of lead per decilitre of blood.

Across Flint, the researchers found a statistically significant increase – the incidence of EBLL had doubled from 2.4% in 2013 to 4.9% in 2015. In neighbourhoods with higher concentrations of lead in water, the incidence of EBLL increased from 4% to 10.6%.

Hanna-Attisha's analysis was based on blood tests processed by the Hurley Medical Center. She had asked for access to a broader set of data held by the Michigan Department of Health and Human Services (MDHHS), but those data were never made available, according to the Flint Water Advisory Task Force.

The Hurley results were published in November 2015 in the *American Journal of Public Health*,¹ but Hanna-Attisha took the extraordinary step of releasing her findings to the public before the peer-review process was completed. "State officials called my science faulty and accused me of creating hysteria," she wrote in a recent *New York Times* op-ed (nyti.ms/2ko4z5e). Indeed, the Task Force quotes an MDHHS memo, written a few days after the press conference, asking staff to "make a strong statement with a demonstration of proof that the blood levels seen are not out of the ordinary". However, analysis by the MDHHS's own epidemiologists subsequently led the department to agree publicly with Hanna-Attisha's analysis.