Testing the Waters: A Dive into Elevated Lead Levels in Toronto Homes*

My subtitle if needed

Abbass Sleiman

January 24, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

Table of contents

*Code and data used in this analysis can be found at: https://github.com/AbbassSleiman/Lead-Concentrations.

1 Introduction

Lead exposure is a serious concern for many, capable of seriously harming children's health, as well as causing a multitude of effects including damage to the brain, slowed growth, decreased IQ, and various others (*Health Effects of Lead Exposure* 2022). Of the various sources of lead exposure, lead in drinking water is a relevant concern for many of us given the average person's daily reliance on tap water. Thus, given the dangers of lead exposure, in conjunction with the fact that lead cannot be seen, smelled, nor tasted, means that getting one's water tested and having knowledge of its lead concentration is incredibly useful, and possibly impactful (*Lead in Drinking Water* 2023).

Water produced by Toronto's drinking water treatment plans is not contaminated by lead. Instead, lead either contaminates the water through water service pipes, solder, or leaded-brass fixtures in objects like valves or faucets, depending on when the particular house was built. In 2011, the Toronto City Council approved a lead in water mitigation strategy that aimed to reduce lead in drinking water. In 2014, the city had begun to add phosphate to the drinking water treatment process which forms a protective coating in all pipes and plumbing fixtures, effectively aiding in the reduction of lead contamination in water (*Lead & Drinking Water* 2024).

Lead concentration in water is typically measured in parts per billion (ppb), which is a unit of measurement describing small concentrations in water whereby 1 ppb is equivalent to 1 microgram per litre ($1\mu g/L$) (*Parts Per Billion* 2023). In May of 2014, a study had showed that 13% of Torontonian homes exceeded Health Canada's standards for lead exposure of 10 ppb (the limit at the time) after analyzing 15,000 water samples provided to the city by homeowners between 2008-2014 through the Residential Lead Testing Program (*High Lead Levels Found in Some Toronto Drinking Water* 2014). This paper utilizes data from the Residential Lead Testing Program that includes 9,302 water samples provided by households between 2015-2024 and seeks to evaluate whether the implementation of phosphate into the drinking water treatment in 2014 has made any impact on the portion of Torontonian homes with lead exposure exceeding Health Canada's past standards of 10 ppb, as well as the updated standard maximum of 5 ppb by Health Canada (Canada 2019). This paper finds that…

The remainder of this paper is structured as follows. Section 2….

2 Data

2.1 Raw Data

The data used in this paper is derived from Open Data Toronto and is read into this paper through the opendatatoronto library (Gelfand 2022). The particular data set used to analyze the lead concentrations in water samples in Torontonian homes is Non Regulated Lead Sample (Toronto 2024). All the data analysis was done through R (R Core Team 2022) with the aid of the following packages: tidyverse (**citetidyverse?**), here (**citehere?**), dplyr (**citedplyr?**), tibble (**citetibble?**), janitor (**citejanitor?**), ggplot2 (**citeggplot2?**), and knitr (**citeknitr?**). (maybe readr?)

The data used is published by Toronto Water and features data from Toronto's Residential Lead Testing Program, providing information on various houses' lead concentrations based on water samples that the households themselves provide. The data is refreshed daily and the particular data used in this paper is up-to-date as of January 22, 2024. The raw data set features the lead concentration in parts per million (ppm) of 12,810 water samples where 1 ppm is equivalent to 1000 ppb or 1 milligram per litre ($1mg/L$). The data set also includes the date that each sample was collected, as well as the partial postal code (only the first three digits of the resident's postal code for privacy reasons).

2.2 Cleaned Data

Some of the data points had missing attributes whereby a "NA" was put in place of the true value. Such entries were removed entirely in the data cleaning process to simplify the analysis procedure. Moreover, the raw data set includes samples collected as early as January 1 2014 and as late as January 2 2024. As this paper is concerned with the after-effects of the phosphate addition to the drinking water treatment process in 2014, all entries in 2014 were also eliminated in the cleaning process to ensure that the data analysis is conducted only on water samples taken after the policy was put into effect. Furthermore, the cleaned data features only the columns for the date, partial postal code, and lead concentration (in ppb as it is the more commonly used unit of measurement). Some lead concentration entries in the raw data were also deemed to be outliers and were subsequently removed in the data cleaning process. In the context of this paper, a lead concentration outlier is defined to be any value exceeding (and including) 100 ppb, 20 times Health Canada's standard of 5 ppb (Canada 2019), and as such given that in Canada the concentration of lead in water is generally below the maximum (Canada 2021), it is reasonable to assume that values above 100 ppb are clear outliers or simply errors in data collection. A sample of the cleaned data can be seen in Table 1 and a scatter plot showcasing every observation, by date of collection, can be seen in Figure 1.

Table 1: Sample of Cleaned Lead Data

Sample Date	Partial Postal Code	Lead Concentration (ppb)
2015-01-02	M1N	2.40
2015-01-02	M4V	0.52
2015-01-02	M4J	11.10
2015-01-02	M6H	0.68
2015-01-02	M6R	0.05
2015-01-02	M6J	0.22

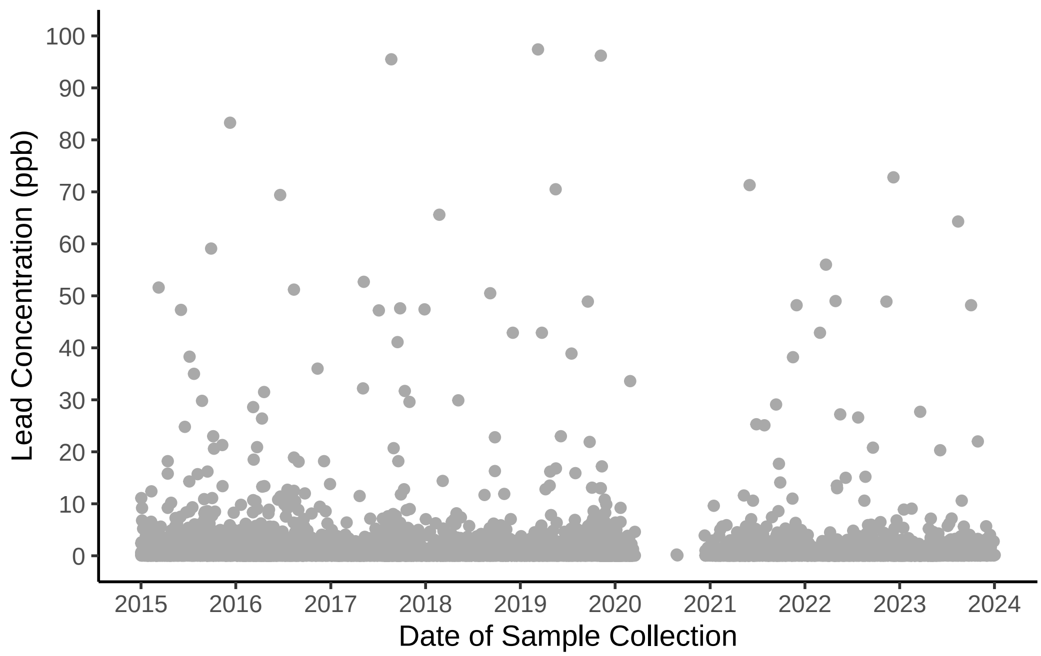

Figure 1: Scatter plot of lead concentration by sample collection date

2.3 Basic Summary of Data

Though Figure 1 provided some insight into the attributes of the data, namely that the vast majority of the water samples lie below the previous lead concentration limit of 10 ppb, and that there is a clear gap in data in the year 2020, more information is required to get a better grasp of the full picture. To gain a clearer insight into the intricacies of the data, it is important to first make note of the number of observations in each year data was collected in (2015-2024). This information is laid out in Table 2, enabling us to see that we have access to much fewer data points in the year 2020 (likely as a result of the COVID-19 pandemic), as well as illuminating the fact that there is but a single observation in the year 2024. Another important aspect to discuss is the overall mean and standard deviation of lead concentrations (ppb) that the entire data set yields. As showcased in Table 3, we can clearly see that, on average, households tend to have lead concentrations well below the maximum limit (approximately 1.04 ppb compared to the limit of 5 ppb). However, taking into account the fairly large standard deviation of approximately 4.05 ppb, we cannot make any reasonable conclusions as of yet. A deeper analysis will follow in Section 3.

Table 2: Number of Observations by Year

Year of Sample Collection	Number of Observations
2015	972
2016	1451
2017	1260
2018	976
2019	1953
2020	370
2021	844
2022	743
2023	692
2024	1

Table 3: Mean and Standard Deviation of Aggregate Lead Concentrations (ppb)

Mean Lead Concentration (ppb)	SD of Lead Concentration (ppb)
1	4

2.4 Discussion of Data Selection

This particular data set was chosen as it is derived from the exact same source that the original study conducted in 2014 - which deemed that 13% of Torontonian households exceeded the maximum acceptable limit of 10 ppb - used (*High Lead Levels Found in Some Toronto Drinking Water* 2014). Thus, in an attempt to mitigate potential biases, the paper makes use of data that was collected in the exact same manner but over the time period of interest (2015 and onward). Moreover, this data set contains a large number of observations spaced out over a number of years, allowing us to discuss findings with lesser worry on its validity as a result of a lack of observations, as well as allowing us to examine possible trends in the data over time.

3 Results

3.1 Examining the Portion of Households Exceeding the Lead Concentration Limit

We are primarily interested in whether the portion of households that exceed the lead concentration limit of 10 ppb has changed from the past portion of 13% (*High Lead Levels Found in Some Toronto Drinking Water* 2014). However, it is also important to examine whether there is a possibly significant portion of households that feature a water lead concentration that exceeds the more recent limit of 5 ppb. Table 4 summarizes the portion of households from the data set that fall under various ranges of lead concentrations.

Table 4: Portion of Households that Fall Under (Fix Name)

Lead Concentration (ppb)	Portion of Households
<5	97.14
5-10	1.59
10-20	0.62
>20	0.66

Through the use of Table 4, we can see that the vast majority of water samples (98.73%) contained a lead concentration below the previous limit of 10 ppb. Even more so, approximately 97.14% of water samples are below the new limit of 5 ppb. In essence, there is clear evidence that since the previous study conducted in 2014 there has been fairly significant progress with regards to the portion of households that exceed, not only the previous limit of 10 ppb, but even the newer limit of 5 ppb, as we see that a small minority of households (2.87%) exceed the new limit, and an even smaller portion (1.28%) of households exceed the past limit of 10 ppb compared to the past portion of 13%.

3.2 Investigating the Relationship Between Time and Lead Concentration

The addition of phosphate to the water cleaning process was done in 2014, however based on the analysis done thus far it is difficult to evaluate whether its impact on reducing lead concentrations was instantaneous or done over longer periods of time. A deeper understanding of this may be of use to cities or nations struggling with reducing lead exposure by water who may require solutions that can act fast. As such, one metric we can use to see the effect of the phosphate addition over time is to simply compare the means of water samples collected in a particular period of time and compare it to subsequent periods. In our particular case, we can employ Figure 2 to see the change in the mean lead concentrations of water samples across various years from 2015 to 2024.

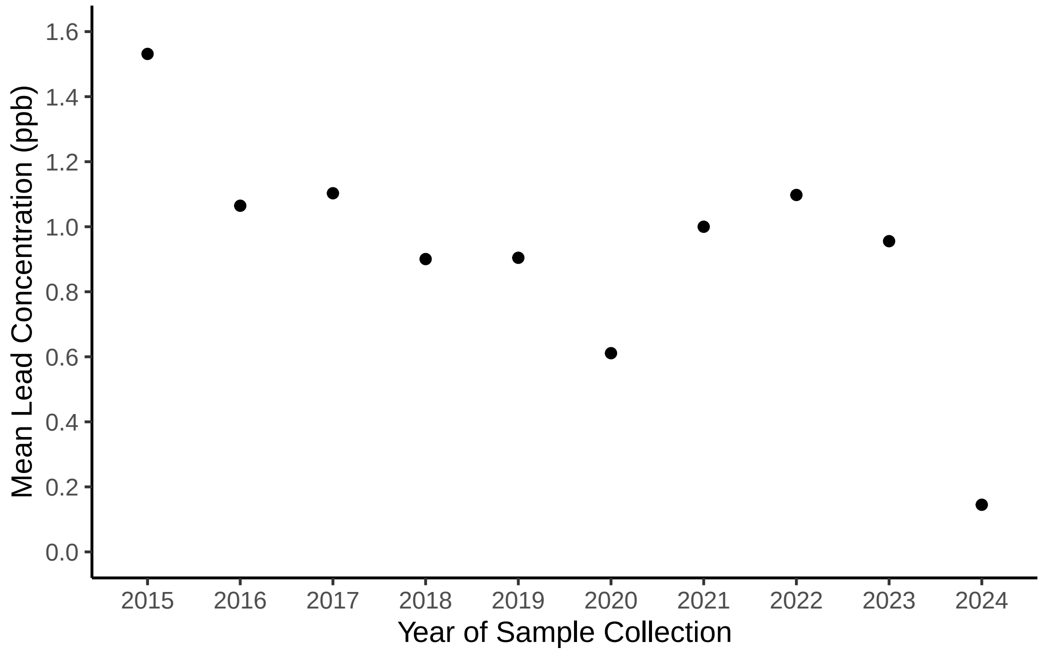

Figure 2: Mean Lead Concentration (ppb) by Year of Sample Collection

Figure 2 offers some evidence that the improvement in lead concentrations could have been gradual as a negative correlation between mean lead concentration and time is visible. It is important to note, however, that the data point for the year 2024 is not entirely trustworthy given that it is based off only one observation. Using just the mean as a metric is not enough to conclude whether the improvement was gradual, however, as the portion of households exceeding the lead concentration limit could have either remained constant or even possibly increased over time. Thus, in an attempt to confront this issue, we can compare the portion of households that exceed a lead concentration of 10 ppb, as well as the portion of households that exceed a lead concentration of 5 ppb, over time using Figure 3 and Figure 4 respectively. Do note that as a result of having only one observation for the year 2024, it is omitted from both graphs as examining the portion of observations that fall under any category in that year will yield either 100% or 0%, and as such, does not aid in our discussion.

Though both figures appear to showcase a slight rise between 2019-2022, we can see an overall fairly consistent decline over time in the portion of households that exceed a lead concentration of 10 ppb as well as 5 ppb. The year with the greatest portion of households exceeding a lead concentration of 10 ppb is 2015, which also happens to be the year with the greatest portion of houses exceeding a concentration of 5 ppb. Similarly, 2020 features the lowest portion of households exceeding a lead concentration of 5 ppb as well as 10 ppb.

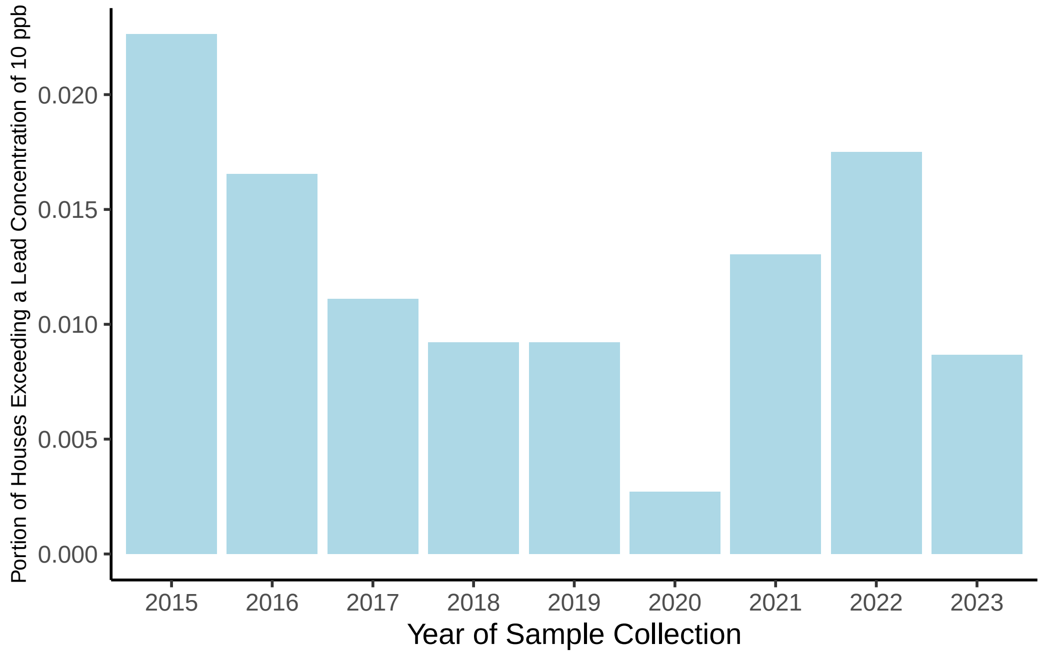

Figure 3: Portion of Households Exceeding a Lead Concentration of 10 ppb by Year

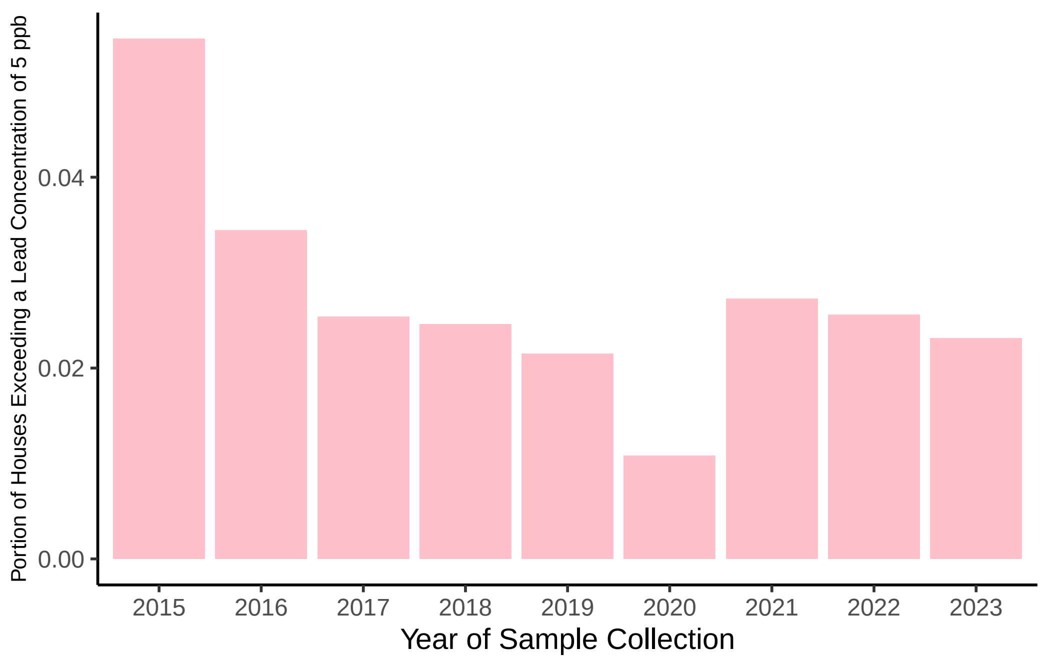

Figure 4: Portion of Households Exceeding a Lead Concentration of 5 ppb by Year

3.3 Exploring the Relationship Between Location and Lead Concentration

We can further expand our discussion by observing potential trends in lead concentration by location by evaluating the mean lead concentration in each (partial) postal code. However, due to the sheer number of varying postal codes available in the data set, Figure 5 simplifies the information slightly by grouping all entries from the same general geographic location by merging values from postal codes with the same first two characters. Though we may lose some underlying information about the lead concentrations in each individual partial postal code, by grouping entries together in this way we end up with a larger number of data points for each observation. Thus, it is more likely that the means that we calculate will better reflect the means of the populations of interest relative to the means we would calculate for each individual partial postal code as some partial postal codes in the data feature a minimal number of observations compared to others.

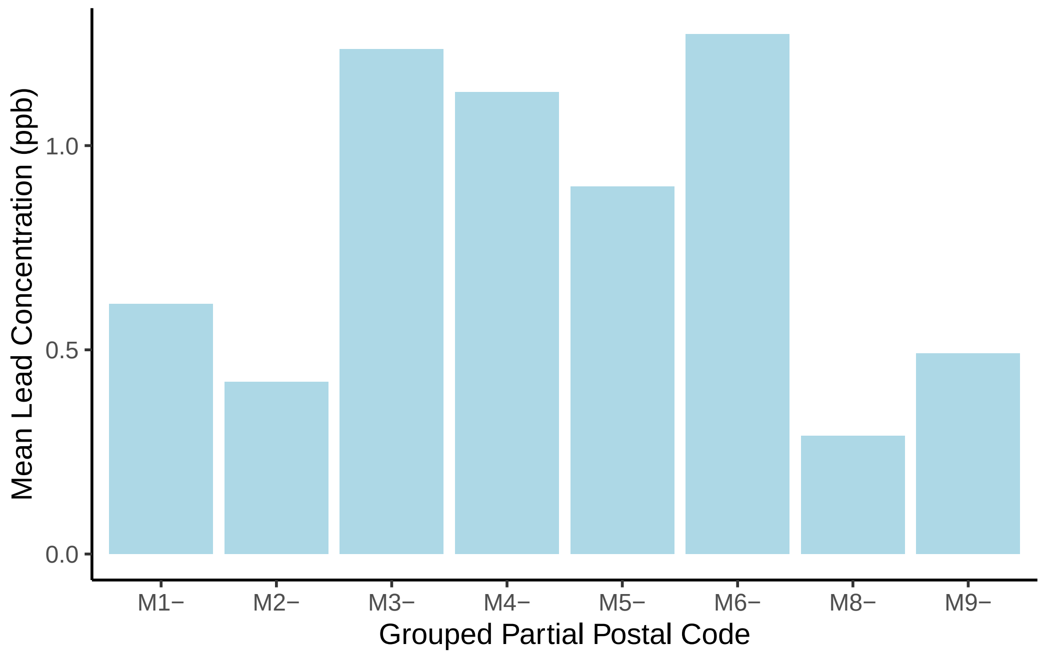

Figure 5: Mean Lead Concentration (ppb) by Partial Postal Code

Table 5: Sample of Mean Lead Concentration (ppb) by Partial Postal Code

Partial Postal Code	Mean Lead Concentration (ppb)	SD of Lead Concentration (ppb)
M1B	0.85	2.05
M1C	0.94	3.56
M1E	0.09	0.07
M1G	0.14	0.13

Table 5: Sample of Mean Lead Concentration (ppb) by Partial Postal Code

Partial Postal Code	Mean Lead Concentration (ppb)	SD of Lead Concentration (ppb)
M1H	0.16	0.18
M1J	0.40	0.92

4 Discussion

4.1 First discussion point

4.2 Second discussion point

4.3 Third discussion point

4.4 Weaknesses and next steps

Appendix

A Additional data details

A.1 Diagnostics

References

Canada, Government of. 2019. *Guidelines for Canadian Drinking Water Quality: Guideline Technical Document – Lead*. https://www.canada.ca/en/health-canada/services/publications/healthy-living/guidelines-canadian-drinking-water-quality-guideline-technical-document-lead.html.

———. 2021. *Lead Information Package - Some Commonly Asked Questions about Lead and Human Health*. https://www.canada.ca/en/health-canada/services/environmental-workplace-health/environmental-contaminants/lead/lead-information-package-some-commonly-asked-questions-about-lead-human-health.html.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Health Effects of Lead Exposure. 2022. Centers for Disease Control; Prevention. https://www.cdc.gov/nceh/lead/prevention/health-effects.htm.

High Lead Levels Found in Some Toronto Drinking Water. 2014. CBC News. https://www.cbc.ca/news/canada/toronto/high-lead-levels-found-in-some-toronto-drinking-water-1.2648775.

Lead & Drinking Water. 2024. City of Toronto. https://www.toronto.ca/services-payments/water-environment/tap-water-in-toronto/lead-drinking-water/.

Lead in Drinking Water. 2023. Centers for Disease Control; Prevention. https://www.cdc.gov/nceh/lead/prevention/sources/water.htm.

Parts Per Billion. 2023. GreenFacts. https://www.greenfacts.org/glossary/pqrs/parts-per-billion.htm.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Toronto, Open Data. 2024. *Non Regulated Lead Sample*. https://open.toronto.ca/dataset/non-regulated-lead-sample/.