

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

My subtitle if needed

Abbass Sleiman

January 23, 2024

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

Table of contents

1	Introduction	2
2	Data	2
3	Results	4
4	Discussion	4
4.1	First discussion point	4
4.2	Second discussion point	5
4.3	Third discussion point	5
4.4	Weaknesses and next steps	5
	Appendix	6
A	Additional data details	6
A.1	Diagnostics	6
	References	7

*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 ($1\mu\text{g/L}$) (*Parts Per Billion* 2023). In May of 2014, a study had showed that 13% of Torontonians homes exceeded Health Canada's standards for lead exposure of 10 ppb 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 Torontonians homes with lead exposure exceeding Health Canada's standards of 10 ppb. This paper finds that...

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

2 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 Torontonians 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?)

Table 1: Sample of Raw Lead Data

ID	Sample Number	Sample Date	Partial Postal Code	Lead Concentration (ppm)
1	1536645	2014-01-01	M4E	0.0078
2	1535456	2014-01-02	M4K	0.00011
3	1536641	2014-01-03	M1L	0.000092
4	1548101	2014-01-06	M3B	0.00019
5	1540991	2014-01-06	M6H	0.012
6	1535092	2014-01-07	M4J	0.0056

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 ($1mg/L$). For each water sample the following is included: the id (unique row identifier for the Open Data database), sample number (laboratory sample id), sample date (the date the sample was collected by the resident), partial postal code (only the first three digits of the resident’s postal code for privacy reasons), lead amount (the lead concentration in ppm from the sample), a sample of which can be seen in Table 1.

Table 2: Sample of Cleaned Lead Data

Sample Date	Partial Postal Code	Lead Concentration (ppb)
2015-01-02	M1N	2.4000
2015-01-02	M4V	0.5230
2015-01-02	M4J	11.1000
2015-01-02	M6H	0.6770
2015-01-02	M6R	0.0501
2015-01-02	M6J	0.2220

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. Moreover, 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). A sample of the cleaned data can be seen in Table 2 above.

Table 3: Summary Statistics by Postal Code

Partial Postal Code	Mean Lead Concentration (ppb)	SD of Lead Concentration (ppb)
M1B	0.8536438	2.0524238
M1C	0.9419778	3.5596075
M1E	0.0852074	0.0651574
M1G	0.1364583	0.1309917
M1H	0.1606364	0.1802339
M1J	0.4010842	0.9194901

This particular data set was chosen as it is derived from the exact same source that the original study conducted in 2014 - that deemed that 13% of Torontonians 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).

Talk more about it.

And also planes (). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

Talk way more about it.

3 Results

Our results are summarized in

4 Discussion

4.1 First discussion point

If my paper were 10 pages, then should be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

4.2 Second discussion point

4.3 Third discussion point

4.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

A.1 Diagnostics

References

- 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/>.