Investigating the lead concentration in Toronto water pipe*

Yuchao Niu

January 25, 2024

Lead exposure can have severe health implications especially for the vulnerable groups of children and pregnant and breastfeeding women. This report investigates the risk of lead exposure through tap drinking water in the city of Toronto by analyzing an open dataset and found the risk level to be negligible. The report also found winter months to correlate with lower likelihood of lead exposure through tap drinking water. The report supported that the corrosion control plan implemented by the city is effective.

1 Introduction

Lead is a naturally occurring element present in the earth's crust. Lead compounds have been widely used in our society including in paint, pipes, gasoline, batteries, and cosmetics. Lead can also be introduced into the environment through industrial activities such as through mining and spark-ignition engine aircraft. Despite some of their beneficial uses, lead can be potentially toxic to humans and animals at elevated levels affecting nearly every organ in the body with the greatest risk to the nervous system. Children are most susceptible to the effects of lead, where even low levels of lead in the blood can lead to behavior and learning problems, slowed growth, lower IQ and anemia (Reference). Lead can also accumulate in our bodies. Pregnant and breastfeeding women previously exposed to lead may have it accumulated in their bodies and released into the bloodstream causing harms to their fetus or infants.

Exposure to lead through tap drinking water is one major cause of lead poisoning and has led to past public health crisis [flinch citation]. Lead was commonly used in plumbing materials such as water pipes and solder and can leach into the water supply. New regulations after 1990s have restricted the use of lead in plumbing, however older buildings still face the risk of lead

*Code and data are available at:https://github.com/MelanieNiu/Investigating-the-risk-of-lead-exposure-through-tap-drinking-water-in-Toronto-

leaching. The Canadian Drinking Water Quality Guidelines establish a maximum allowable concentration of 0.005 ppb for total lead in drinking water, measured directly at the tap.

Toronto Residential Lead Testing Program supported by the city of Toronto to test lead concentration in tap drinking water provides data in this aspect. Utilizing this dataset, my report conducted statistical analysis to gain insight of the risk of lead exposure through Toronto tap drinking water from 2014 to the present and provide support that the corrosion control plan initiated by the city since 2014 has effectively reduced the risk of lead exposure through tap drinking water. Section 2.1 the Toronto Residential Lead Testing Program dataset "Non Regulated Lead Sample" and its measurement is introduced. In Section 2.2 each variable of the dataset is explained and visualized. In section 3, seasonal and regional factors to potential impact lead concentration in tap drinking water are explored. While regional factors show no clear relationship with lead concentration in tap drinking water in Toronto, winter months show the lowest number of drinking water samples with lead concentration exceeding the allowed concentration set out by public health guidelines.

2 Data

2.1 Data Source and Measurement

The data utilized in this report are retrieved from Open Data Toronto Portal. The dataset is titled "Non Regulated Lead Sample". The data are being collected by the City's Residential Lead Testing Program which provides free testing to Toronto residents who have concerns about potential lead contamination in their drinking water. Residents have the option to collect and return a water sample kit at any of the six Toronto Public Health facilities. Detailed instructions are provided on how to procure an accurate water sample. The lead test results are shared with the requesting residents when available.

One inherent limitation of this dataset is that sampling is conducted based on requests of residents other than a systematic collection method. Therefore not all water pipes of the city are sampled consistently. Since residents who are more concerned about lead toxicity or residents who live in households installed with older lead water pipes are more likely to request sampling, a self-selection bias may be present and hence limit the usefulness of this dataset. In addition, the water samples are collected by individual. requesting resident. Even though with detailed instructions are provided, the City's Residential Lead Testing Program has no complete control over the exact sample collection method and location. Privacy of the requesting residents is preserved; no personal identifier or exact location of sampling is included in the dataset for privacy reasons.

This dataset contains 13000 observations and 5 columns containing variables of ID, Sample Number, Sample Date, Partial Postal Code of the sampling location and Lead Amount detected in the drinking water sample in ppm. The dataset was last updated on January 24, 2024 and has captured data from 2014 till present.

The data were cleaned, analyzed, and visualized using the open-source statistical programming language R (R Core Team 2022) Functionalities from additional packages tidyverse (Wickham et al. 2019), lubridate (Grolemund and Wickham 2011) and ggplot (**ggplot?**) were also utilized.

3 Results

References

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

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

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.