

Evaluating the factors pose the greatest threat to the lives and health of different groups of homeless people during 2017-2023*

Juliana Zhu

September 27, 2024

The paper evaluates the factors posing the greatest threats to the lives and health of various homeless people between 2017 and 2023. The dataset represents key determinants such as accident, drug toxicity and cancer that have impacted mortality rates among homeless people. By examining data, it emphasizes how these factors affect different groups of homeless people, including those based on age, gender, and year of death. The findings underscore the need for targeted interventions to reduce mortality of homeless and improve overall health conditions

1 Introduction

The homeless are one of the most vulnerable groups in society, and there are many factors that seriously affect their health and well-being. Discussing and studying the health and life threats of homeless can help urban planners consider the special needs of these people when designing and managing cities, thus creating a more inclusive and safer urban environment. At the same time, such research can prompt more social attention and support for homeless to promote the overall progress of society.

In this paper, I will analyze the homeless deaths by cause data from Open Data Toronto, focusing on the distribution and the relationships between the homeless deaths count, year of death, age groups, gender and cause of data. From the data, the main causes of death among the homeless from 2017 to 2023 are Accident, Cancer, Cardiovascular Disease, COVID-19, Drug toxicity, Homicide, Infection, Pneumonia, and Suicide. By analyzing the data on the

*Code and data are available at:

causes of death, we can better understand which factors are most harming the health and well-being of the homeless so that we can call on governments and society to take measures.

I use ‘R’ to do all the statistical analysis and data visualization (R Core Team (2023)) and use ‘tidyverse’ to plot the data (Wickham et al. (2019)). At the same time, I use ‘opendatatoronto’ to access and download Toronto City’s open data of homeless deaths by cause (Gelfand (2022)). Besides of that, I use ‘here’ package to simplify the file paths (Müller (2020)) and use ‘readr’ package to read the date from the file (Wickham, Hester, and Bryan (2024)).

The remainder of this paper is structured as follows: in Section 2 I present the data and the variables used for the analysis. Then I made the plots about relationships between other variables and the count of deaths and compare them to get the final finding.

2 Data

2.1 The background of the data

The dataset is from Toronto Public Health (TPH), starting in January 2017 to track the deaths of people experiencing homelessness to get a more accurate estimate of the number of deaths and their causes. For this data collection initiative, homelessness is defined as “the situation of an individual or family without stable, permanent, appropriate housing, or the immediate prospect, means and ability of acquiring it”.

2.2 Measurement of variables from the data

The data has 5 variables: “year_of_death”: the year that the death occurred; “cause_of_death”: the reasons of homeless deaths; “age_group”: the age of homeless people; “gender”: the gender of homeless people; “count”: the total number of homeless deaths.

2.3 Limitations of the data (from opendatatoronto)

-This data reflects only deaths that are reported to TPH by SSHA, community partners and the Coroner’s Office.

-Cause of death is unknown or pending in approximately 25% of the reported deaths.

-In 2019, TPH ceased reporting on location of death as this information is often unknown or unverified.

-The data does not identify Indigenous status as this is reported as unknown or missing in 70% of the reported cases. With this high a percentage of missing information, accurate conclusions

cannot be drawn, as such, it is standard epidemiological practice to suppress the release of data.

-Cause of death for transgender people not shown due to small counts. -Cause of death is unknown in approximately 25% of the reports. To protect privacy, causes of death with less than 2% of the cases are included in the Other category.

-Data are subject to change should TPH receive additional reports or further information on existing reports. Previously published data may change due to delays in reporting. The most recently published data should be considered the most complete.

2.4 Visualization of the data

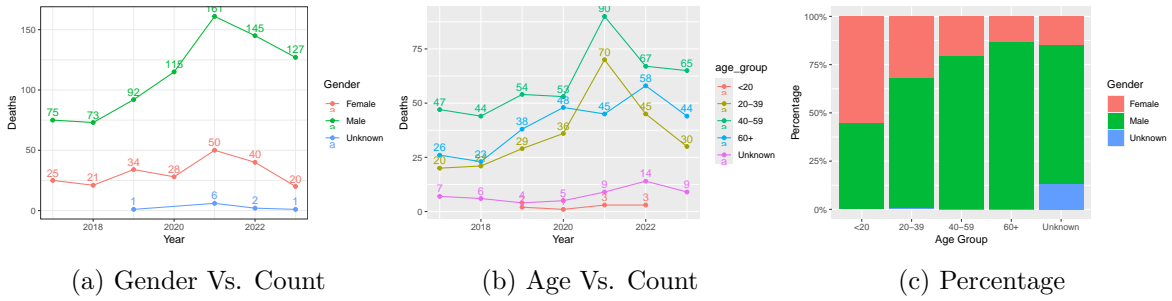


Figure 1: Number of deaths by year for different age groups and gender groups

Figure 1 presents the number of deaths among homeless people categorized by gender and age groups from 2017 to 2023, divided into three subplots. Figure 1a illustrates the number of deaths caused by being homeless recorded from 2017-2023, categorized by gender into Female, Male and the Unknown. According to the graph, the total death rate of homeless males shows a continuous upward trend, starting at 75 in 2017, peaking at 161 in 2021, and then slightly declining to 127 by 2023. In comparison, the death rate among homeless females has shown a steadier pattern. The number started at 25 in 2017, gradually rising to a peak of 50 in 2021, and ending with a total of 20 in 2023. The Unknown category remained consistently low throughout the 6-year period, starting at 1 in 2017, reaching a peak of 6 in 2021, and returning to 1 by 2023.

Figure 1b presents the number of deaths among homeless individuals categorized by age group from 2017 to 2023. According to the graph, the death rate among the 40-59 age group shows a continuous upward trend, starting at 47 in 2017, peaking at 90 in 2021, and then slightly declining to 65 by 2023. The death rate among the 60+ age group started at 26 in 2017, gradually rising to a peak of 58 in 2022, followed by a decrease to 44 in 2023. The 20-39 age group also reached its peak in 2021 with 70 deaths, before dropping sharply to 30 by 2023. The death rate for those under 20 shown in the graph remained consistently low, ending with

3 deaths in 2023. The Unknown age category showed low numbers throughout the period, peaking at 14 in 2022 before dropping to 9 in 2023.

?@fig-bar-3 presents the percentage distribution of deaths among homeless individuals categorized by both age group and gender. According to the chart, the male gender dominates across all age groups, with the highest representation in the 20-39, 40-59, and 60+ age groups, where males constitute nearly 75% or more of the total deaths. In the under 20 age group, males and females are more evenly represented. The Unknown category indicates some cases where gender identification may be incomplete or not recorded like I mentioned before in “Limitations of the data”.

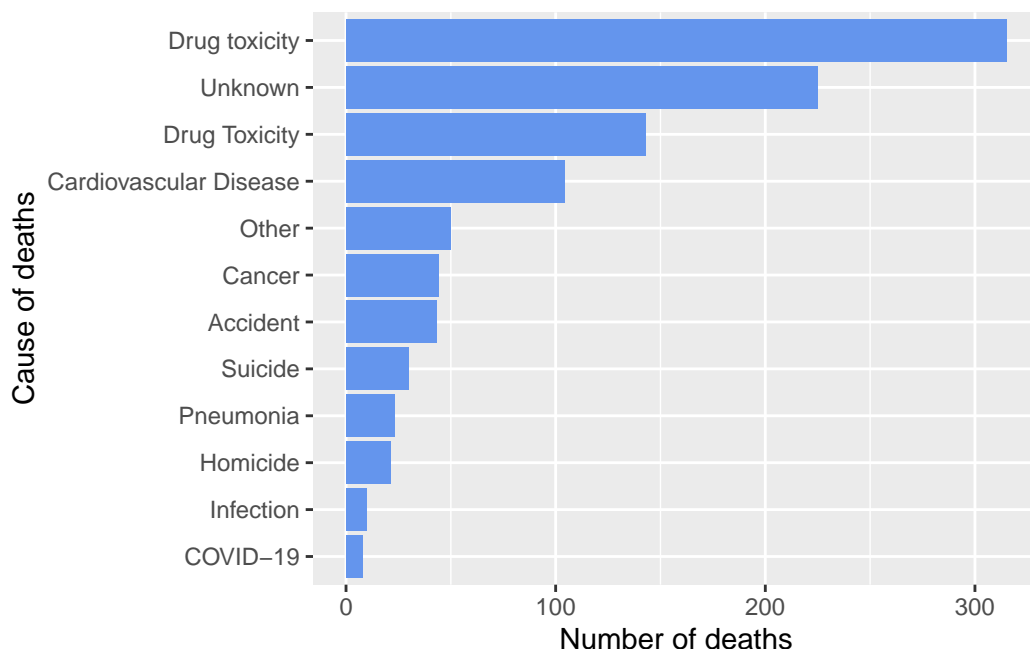


Figure 2: Number of deaths by different causes

Figure 2 presents the number of deaths among homeless individuals categorized by cause of death. According to the chart, drug toxicity is the leading cause, with over 300 deaths. As mentioned before in “Limitations of the data”, due to possible gaps in data collection or reporting, a significant number of deaths are categorized under unknown causes. Cardiovascular disease is the next most common cause, followed by cancer, accidents, and suicide. Pneumonia, homicide, and infection are less frequent causes, ending with COVID-19 resulting in the fewest deaths.

?@fig-bar-4 presents the percentage of deaths among homeless individuals categorized by cause of death and gender. According to the chart, the male gender ranks highest across nearly all causes of death, particularly in categories such as drug toxicity, cardiovascular disease, and cancer, where males have the majority of the deaths. In contrast, female homeless death caused

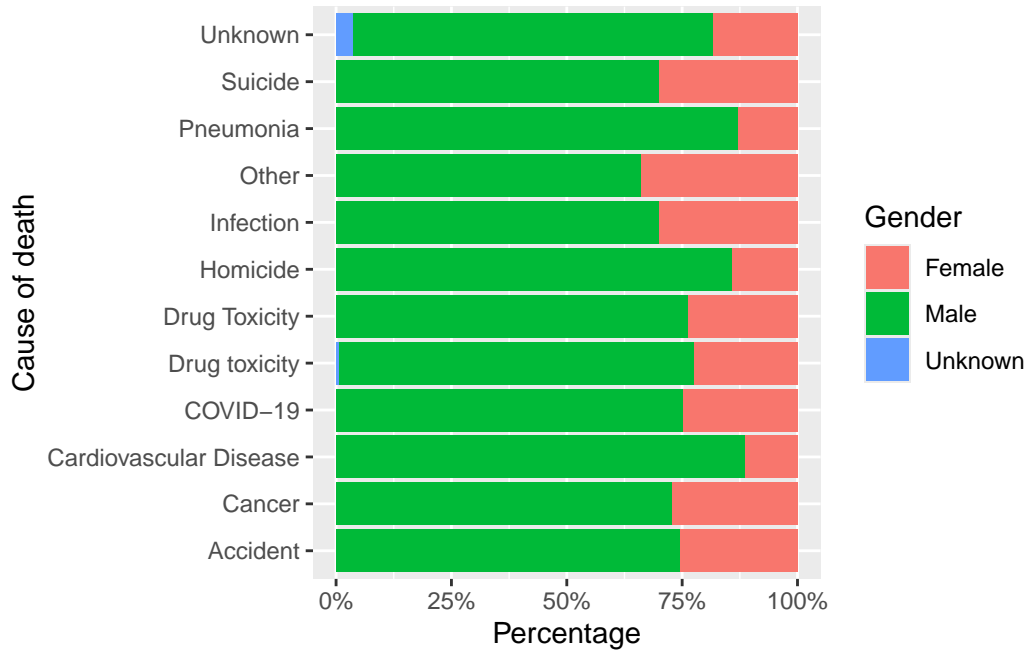


Figure 3: Percentage of deaths by causes for different genders

more by suicide and pneumonia, though still lower than that of males. The Unknown category appears in some causes, reflecting cases where gender identification might be incomplete or unrecorded, as discussed in “Limitations of the data.”

2.5 Conclusion

In conclusion, the analysis of homeless death data from 2017 to 2023, as provided by Open Data Toronto, has demonstrated the critical factors affecting the health and well-being of a very vulnerable community in society. The data reveals that drug toxicity is the leading cause of death among the homeless, with cardiovascular disease and cancer also contributing significantly to mortality rates. This analysis underscores the existing health challenges and risks faced by this community, calling for assistance from governments, organizations, and citizens. By raising awareness among individuals and fostering societal support, such as centers for drug rehabilitation and home shelters, the society could greatly help in reducing the unfortunates among the homeless.

References

- Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- R Core Team. 2023. *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 Grolemond, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>.