Exacerbated by the Pandemic: Analyzing the Impact of COVID-19 on Mortality Among Toronto's Homeless by Gender Disparities*

Dingning Li, Yiming Tang, Yingqi Pang

March 16, 2024

This study examines the impact of COVID-19 on homeless mortality in Toronto from January 2017 to June 2023, with a focus on gender disparities. Utilizing a Difference in Differences approach and data from Toronto Public Health, the research analyzed death counts before and after COVID-19, applying regression models with and without gender controls. Results indicate a general increase in deaths post-COVID, with a more pronounced rise in male mortality. The study underscores the pandemic's exacerbation of existing vulnerabilities in homeless populations, particularly among males, and highlights the need for targeted public health strategies.

Table of contents

*Code and data are available at https://github.com/Yiming1220/Impact-of-Covid-on-Gender-Difference-in-Homeless-Death.git

Introduction

Homelessness is a serious challenge in Toronto, Canada, drawing attention to the harsh realities faced by the most vulnerable groups. The health vulnerabilities and mortality rates of homeless individuals are alarmingly high when compared to the general population. Pre-existing burdens of mental illness, substance dependence, various medical conditions, infectious diseases, and physical trauma have been compounded since December 2019 by the advent of COVID-19, further endangering those in communal living situations.

This study aims to explore the impact of the COVID-19 on the difference of the death rate between female and male. Employing the "Difference in Differences" approach, I will analyze homeless death data in Toronto, spanning from January 2017 to June 2023. We assume two time periods: 2017 January to 2019 December is before covid; 2020 January to 2023 June is in the covid. Death count is the estimand. We will embark on an analysis to validate the hypothesis that the COVID-19 pandemic has led to an increase in the death count among the homeless population. We will present this analysis in two distinct parts: one without controlling for gender, to capture the overall trend (Figure 6), and another with gender control, to discern any differential impact between male and female homeless individuals (Figure 7). Firstly, we will inspect the aggregated data, where we anticipate an observable escalation in death counts following the outbreak, regardless of gender or age group differences. Subsequently, we will refine our analysis with a gender-based perspective. Here, we aim to unravel the nuances of how the pandemic may have disparately affected homeless males and females. This part of the analysis seeks to highlight the intersection of gender with the broader health crisis. In conjunction with these visual comparisons, we will also delve into a more granular analysis by presenting the results of a regression model. This model will quantify the impact of COVID-19 on death counts, first without the inclusion of gender, and then with gender considered. Through this comparative approach, we aim to discern whether the observed increase in death counts post-pandemic holds when we account for gender, as well as other control variables such as the cause of death and age (Table 2). The conclusions drawn will be critical, not just in understanding the impact of the pandemic but also in informing potential interventions and policy decisions aimed at mitigating the risks faced by homeless populations in such health crises.

The analysis will be conducted in R (R Core Team 2020) and will utilize an array of packages such as tidyverse (Wickham et al. 2019), readr (Wickham, Hester, and Bryan 2024), dplyr (Wickham et al. 2021), and opendatatoronto (Smith Year of Publication) for data manipulation and analysis. All visualizations in this report are generated through ggplot2 (Wickham 2016), with tables crafted using kableExtra (Zhu 2021).For importing and exporting 'SPSS', 'Stata', and 'SAS' files, the haven package (Wickham, Miller, and Smith 2023) will be used, alongside knitr (Xie 2021) for dynamic report generation in R.

Data

Data Collection

Data for this research were obtained from the City of Toronto Open Data Portal, facilitated by the script "01_downloaddata.R". The opendatatoronto package (Gelfand 2020) was instrumental in importing the data into the R environment for subsequent processing and analysis. The data, contributed and maintained by Toronto Public Health, is biannually refreshed, with the last update noted on September 29, 2023. Toronto Public Health initiated detailed records of homeless deaths from January 2017, aiming for a precise count and understanding of mortality causes. Data accumulation is a joint effort between Toronto Public Health, Shelter, Support, and Housing Administration (SSHA), various health and social service agencies, with corroborations by the Office of the Chief Coroner of Ontario (OCCO). For this data collection initiative, homelessness is defined as "a situation where an individual or family does not have stable, permanent, appropriate housing, or the immediate prospect, means, and ability to acquire it.

We utilize two datasets titled "Homeless deaths by month.csv" and "Homeless deaths by cause.csv". The dataset "Homeless deaths by cause.csv" comprises six variables: "_id", "Year of death", "Cause_of_death", "Age_group", "Gender", and "Count". The "_id" variable enumerates from 1 to 253, representing the recorded number of homeless individuals who died between 2017 and 2023. The "Year of death" spans from 2017 to 2023. "Cause_of_death" includes various categories such as "Accident", "Cancer", "Cardiovascular Disease", "COVID-19", "Drug Toxicity", "Homicide", "Other", "Pneumonia", "Suicide", and "Unknown/Pending". The "Age_group" is categorized as "< 20", "20-39", "40-59", "60+", and "Unknown". The "Gender" variable includes "Male", "Female", and "Unknown". The "Count" represents the number of deaths. The "Homeless deaths by month.csv" dataset contains four variables: "_id", "Year of death", "Month of death", and "Count". The "Month

Table 1: Preview of Homeless Deaths in Toronto

ID	Year Of Death	Cause Of Death	Age Group	Gender	Death Count
1	2017	Accident	40-59	Male	2
2	2017	Accident	60+	Male	3
3	2017	Cancer	60+	Female	1
4	2017	Cancer	40-59	Female	2
5	2017	Cancer	40-59	Male	2
6	2017	Cancer	60+	Male	4

of death" ranges from January to December. Given the onset of COVID-19 in December 2019, we will use this event as a pivotal point to compare the gender-specific death rates from two periods: 2017-2020 and 2020-2023. We will hold constant the variables "Age_group" and "Cause_of_death", considering them as controls in our analysis. The treatments will be defined as "X1 = time period" and "X2 = gender". We will encode the time periods as a binary variable, with "0" indicating the period from 2017-2020 and "1" representing 2020-2023. The dependent variable, denoted as Y, is the count of deaths.

Data Cleaning

In this data cleaning process, we are using R programming language and several of its packages: `readr` for reading CSV files, `dplyr` for data manipulation, and `kableExtra` for creating LaTeX-formatted tables. Firstly, we set up the paths to our input data file and where we want to save our cleaned data and output tables. After reading the CSV file into R with the correct UTF-8 encoding, we begin by renaming the columns for better readability and consistency in our dataset. The columns are renamed to "ID", "Year Of Death", "Cause Of Death", "Age Group", "Gender", and "Death Count". A preview of the first few rows of this cleaned dataset is then generated as a LaTeX table, which is saved to an output directory for tables. Next, we move on to the cleaning and preprocessing of the data. We define a function `process_age` to convert categorical age ranges to their numerical mean values, with the exception of "<20" being mapped to 15 and "60+" to 70. This function is then applied to the "Age Group" column to create a new "Age" column with numerical values. Where age values are missing, we replace them with the overall mean age. To maintain consistency in our analysis, we filter out records where the gender is listed as "Unknown." We then refine our dataset by selecting only the columns necessary for our Difference in Differences (DiD) analysis. These columns include "year", "cause", "gender", "count", "age", and "age_group". Finally, we create a binary treatment variable to distinguish between the pre- and post-COVID-19 periods based on the year of death. Records from before 2020 are labeled "Before", and those from 2020 onwards are labeled "After". The cleaned and processed data is then written out to a new CSV file, omitting row names to maintain a tidy format. This clean data is now ready for subsequent analysis.

Measurement

In our analysis, we distinguished between pre-pandemic and pandemic periods, a categorization not originally present in the dataset. The process of quantifying and measuring variables within the dataset is detailed in this section. We define the pre-pandemic period as before January 2020 and the pandemic period as from January 2020 onwards, in line with the global onset of COVID-19. Mortality counts among the homeless were aggregated from various sources and validated for accuracy. The data includes identifiers for the causes of death, which range from natural causes to accidental or violent circumstances. To ascertain the temporal impact of the pandemic, we classified the timing of each death relative to the emergence of COVID-19. For each record, we employed nominal scales to categorize causes of death and ordinal scales for variables indicating the level of medical intervention received prior to death, if any. The ratio scale was used for age, allowing for the calculation of relative differences and rates, such as mortality rate per age group. Interval scales were not explicitly utilized, as exact measurements of intervals between events were not central to our analysis.

Efforts to ensure the accuracy and consistency of these measurements included cross-referencing reported deaths with official health records and corroborating data with local health agencies and services. Challenges to data quality, such as underreporting or misclassification of causes of death, were mitigated through robust validation protocols and sensitivity analyses. Understanding the scale and nuances of the dataset is essential for interpreting the findings accurately and for replicating the study in future research. The comprehensive nature of our dataset, including various demographic details and cause-specific mortality counts, enables a multifaceted exploration of the impact of COVID-19 on the homeless population.

Data Analysis

The following charts allow us to observe different trends and changes related to the number of deaths:

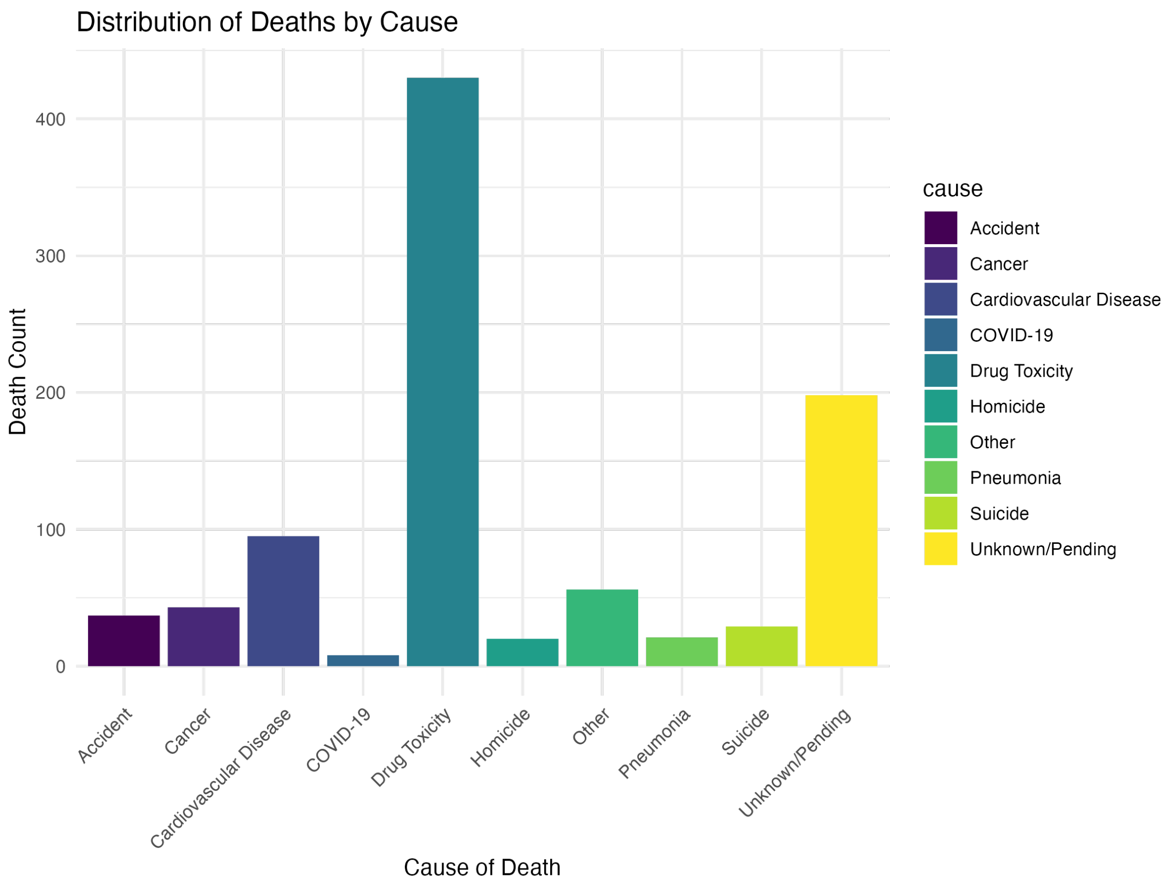

Figure 1: Distribution of Deaths by Cause

A histogram was created to display the distribution of deaths by cause, with categories including 'Accident,' 'Cancer,' 'Cardiovascular Disease,' 'COVID-19,' 'Drug Toxicity,' 'Homicide,' 'Other,' 'Pneumonia,' 'Suicide,' and 'Unknown/Pending' (Figure 1). The data indicate that deaths resulting from drug toxicity are the most prevalent.

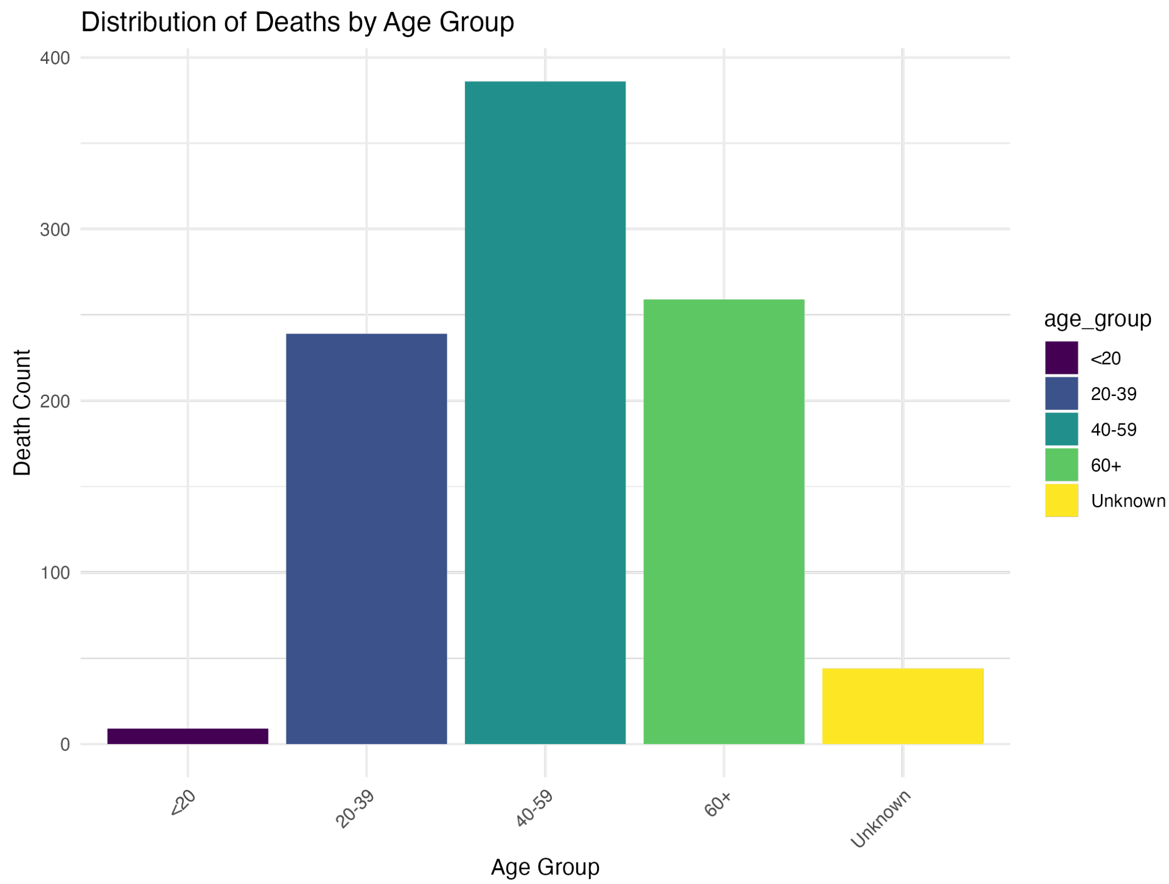

Figure 2: Distribution of Deaths by Age Group

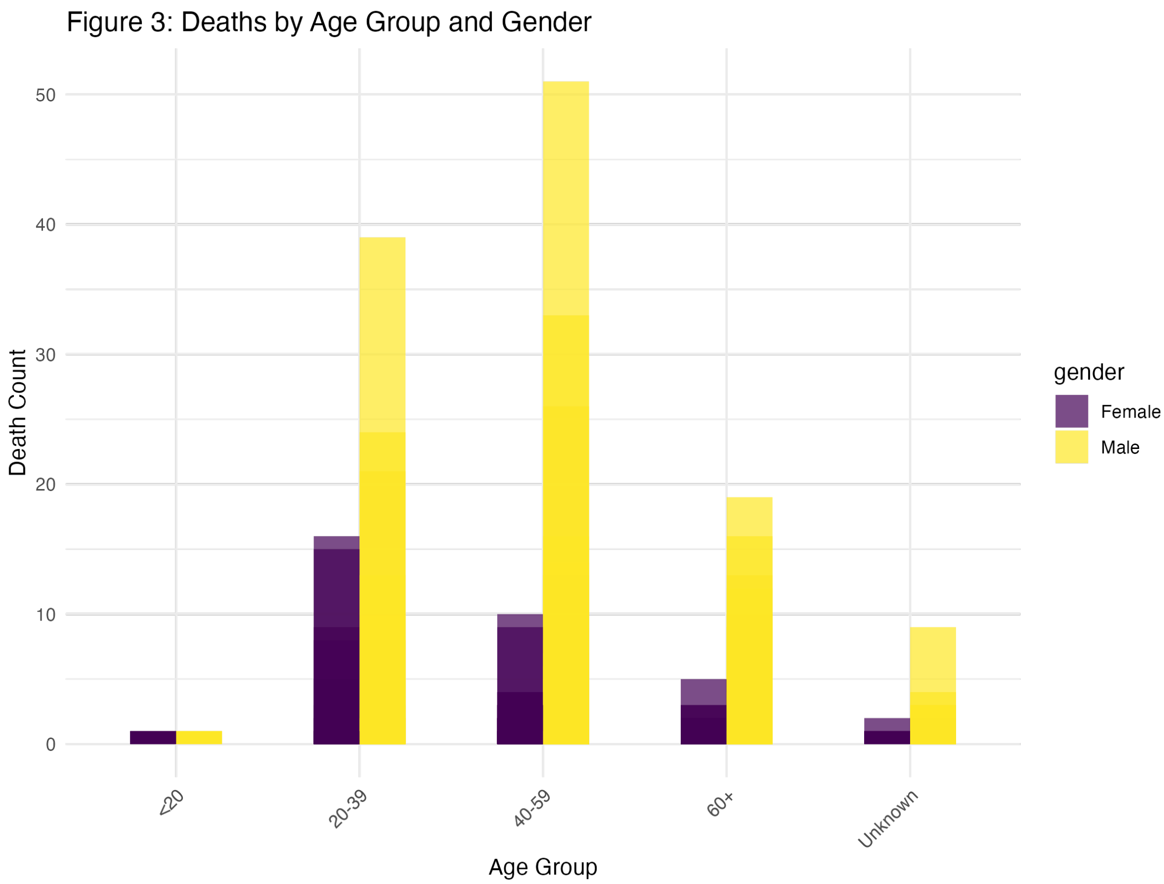

Figure 3: Deaths by Age Group and Gender

Age is a crucial factor when considering mortality rates, as vulnerability to death generally increases with age due to the natural decline in health. Therefore, our further analysis of death data across different age groups (Figure 2) indicates that the most deaths occur in the 40-59 age group, with significant numbers also in the 20-39 and 60+ age groups. In the homeless population, individuals of all ages, including older adults, middle-aged, and even younger ones, might face substantial risks due to hazardous living conditions, lack of healthcare, and potentially dangerous survival behaviors. We then stratify each age group by gender (Figure 3), and it becomes clear that in every age group, the number of male deaths is significantly higher compared to females.

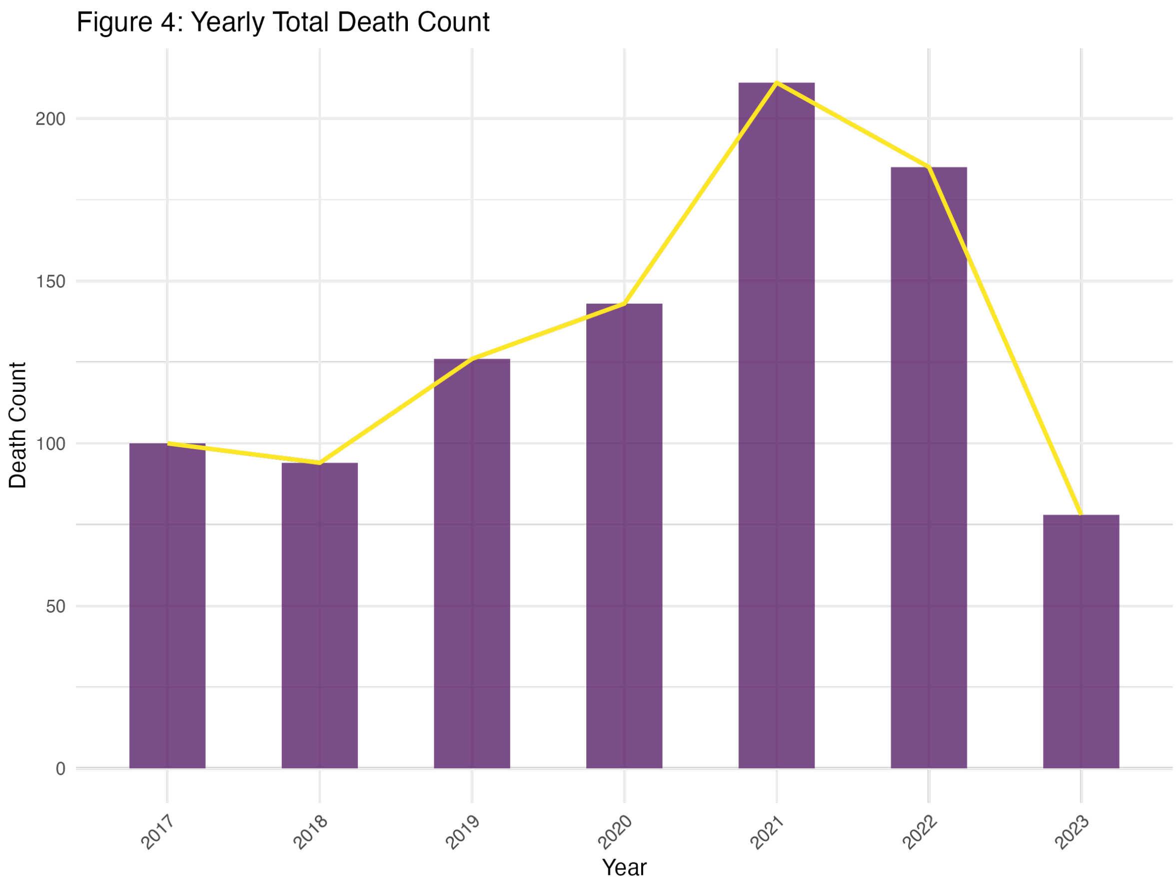

Figure 4: Yearly Total Death Count

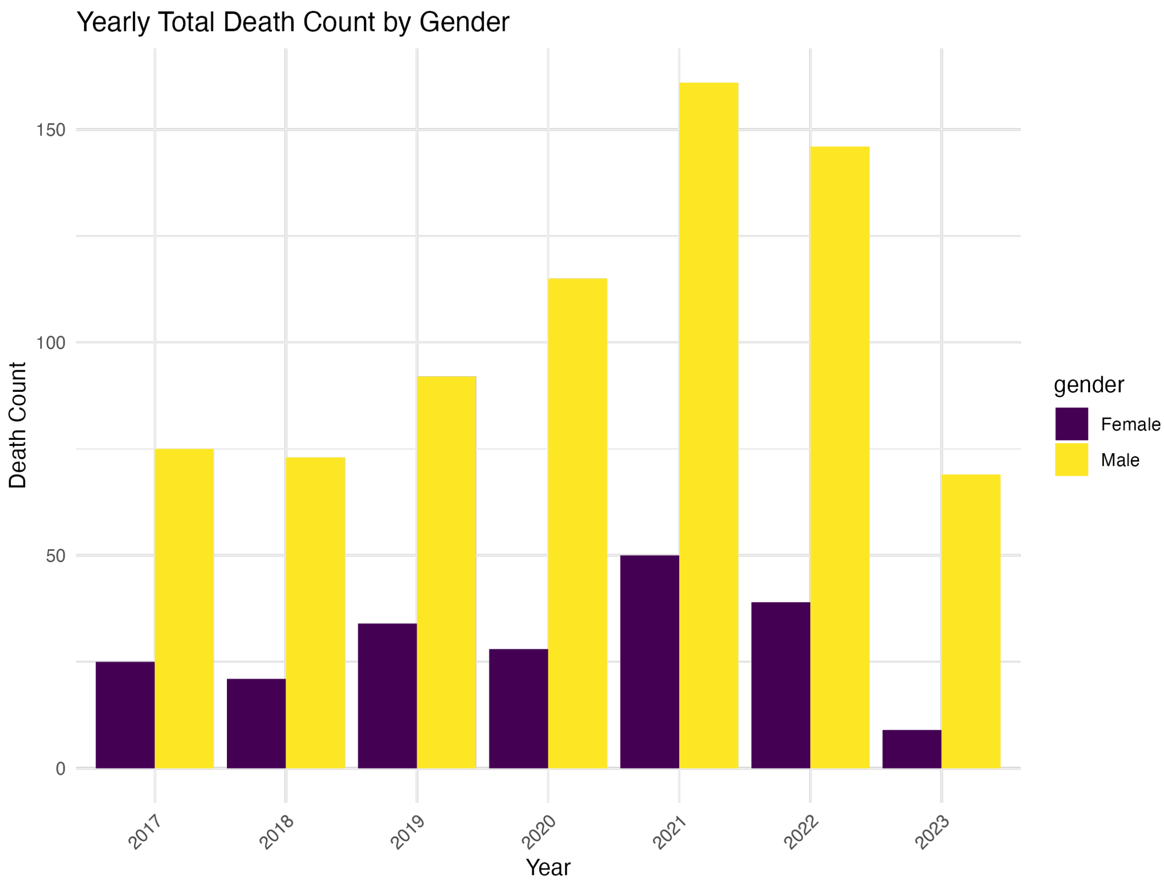

Figure 5: Yearly Total Death Count by Gender

The progression of years reflects changes over time that may influence mortality rates. The onset of COVID-19 marks an increase in the number of deaths among the homeless, attributable not only to the virus itself but also possibly to the compounded risks associated with the pandemic, such as reduced access to services or increased substance use. Figure 4 illustrates the trend in mortality numbers before and after the emergence of COVID-19. The COVID-19 period extends from December 2019 to June 2023. We notice a marginal decrease in numbers from 2017 to 2018, which predates the COVID period. From 2018 to 2021, there is an upward trend, likely due to the initial severe impact of COVID. After 2021, the trend decreases, possibly due to the end of the acute phase of COVID, with people gradually recovering, leading to a reduction in mortality rates. Hence, the link between the increase in mortality and COVID-19 is evident. We segment the annual number of deaths by gender (Figure 5), and it is noteworthy that male mortality consistently exceeds that of females every year.

The relationship between gender and death rates can be intricate. Men may face higher risks in certain contexts due to different social behaviors, health statuses, or utilization of services. The observed gender disparity in mortality rates invites further investigation into the potential impacts of COVID-19 on these differences, providing a rationale for this research report. To probe these relationships, statistical models can be employed to scrutinize the data while controlling for confounding variables, enabling us to assess the distinct impact of each factor. For instance, by incorporating both age and gender into our analysis, we can discern whether COVID-19's effect on mortality varies across different age groups or between genders and to what extent. By examining data from both before and after the pandemic's onset, we aim to distinguish the effects of the pandemic from other time-related factors that might influence mortality rates.

Model

Model Setup

$$Y_{it} = \alpha + \beta T_t + \gamma G_i + \delta(T_t \times G_i) + \mathbf{X}_{it}\theta + \epsilon_{it}$$

Model Justification

- $alpha$ is constant intercept term of the model, capturing the dependent variable(death counts) when all independent variables are 0.
- Y_{it} represents the dependent variable, which in this context is the number of deaths among the homeless population for individual i at time t,
- T_t is a binary indicator for the time period, with $T = 0$ for the pre-COVID-19 period and $T = 1$ for the post-COVID-19 period,
- G_i is a binary indicator for gender, with $G = 1$ for one gender (e.g., females) and $G = 0$ for the other (e.g., males),
- δ represents the DiD estimator, capturing the interaction effect between the time period and gender, indicating the differential impact of COVID-19 on death counts across genders,
- $\mathbf{X}_{it}$ is a vector of control variables for gender i at time t,
- θ is a vector of coefficients for the control variables,
- α, β, and γ are coefficients of time, gender and interaction of time and gender.
- ϵ_{it} is the error term, capturing the unobserved influences on the dependent variable.

The inclusion of the interaction term $(T_t \times G_i)$ is particularly important as it acknowledges that the relationship between time (and therefore the impact of COVID-19) and death counts may not be uniform across genders. This model is justifiable as it provides a comprehensive framework for examining the potential changes in homeless death counts over time while considering the interaction between time and gender, as well as the effect of other relevant factors. This model allows for a deeper understanding of the dynamics at play in this critical public health issue.

Results

Covid-19 Effect

after covid-19, death count of homeless increase. The bar graph illustrates the comparison of homeless death counts before and after the outbreak of COVID-19, not controlled by any other variables. The period prior to COVID-19 is represented by a significantly lower bar, indicating a lower death count among the homeless population. In contrast, the bar representing the post-COVID period is substantially higher, suggesting a marked increase in the number of deaths. The visualization of data comparison supports the hypothesis that the death count of homeless individuals has increased after the outbreak of the pandemic. It is worthy to note that this data does not differentiate by other variables, such as gender or aged group, implying that the increase in deaths is a general trend influencing the homeless population as a whole. This trend could be attributed to various pandemic-related factors, including increased vulnerability to the virus due to pre-existing health conditions, the strain of medical resources, and other indirect consequences of the global pandemic crisis.

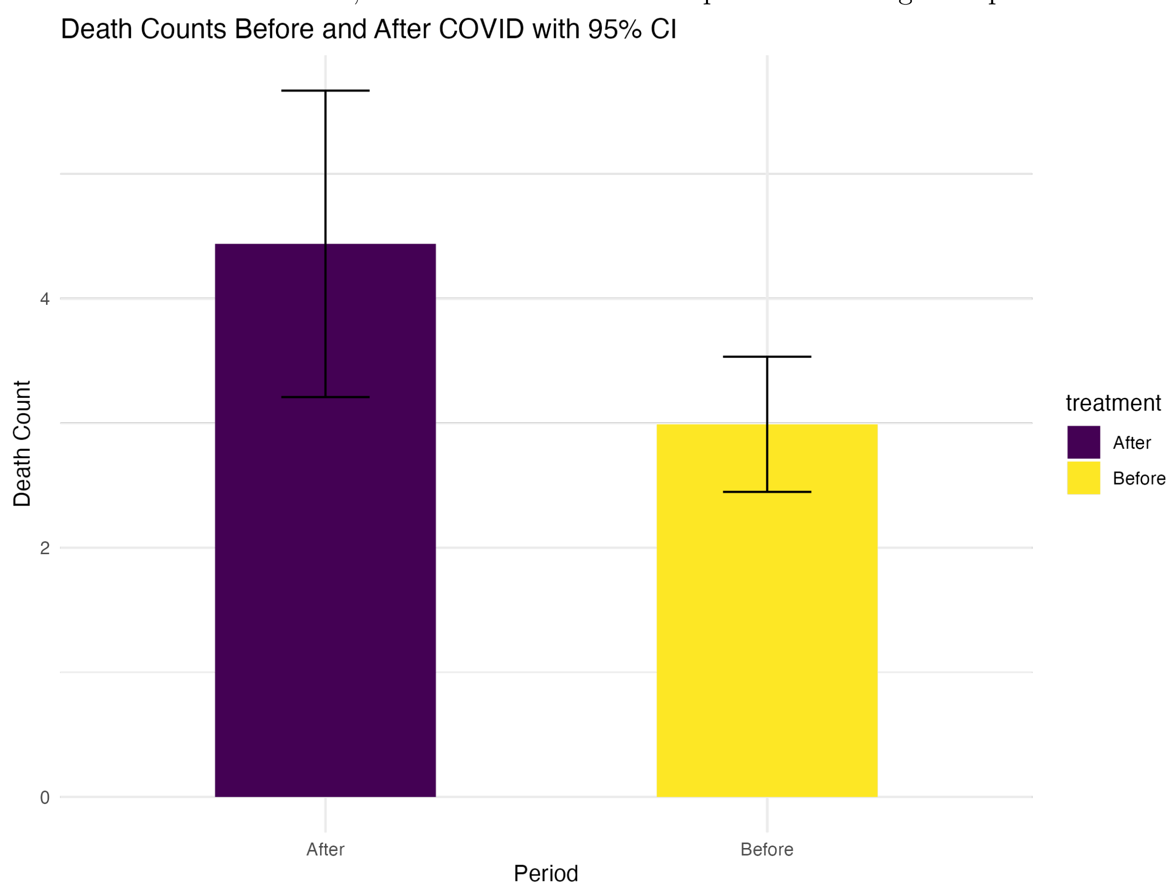

Figure 6: Death Counts Before and After COVID with 95% CI

Different Effect By Gender

The bar chart visualizes a gender-differentiated comparison of homeless death counts before and after the outbreak of the COVID-19 pandemic. In the post-pandemic period, the recorded death counts for both genders were markedly lower, with the male death count being higher than the female. The impact of the pandemic is evident in the post-pandemic period, where there is substantial rise in the death counts for both genders, This increase suggest that the pandemic has had a considerable effect on the homeless population, which is consistent with the analysis we discussed above, exacerbating the existing vulnerabilities faced by this group.

In post-pandemic period, the increase in death counts is significant among males, indicated by the pronounced height of the blue bar, which far exceeds the pre-pandemic figures. The rise in female death counts, represented by the orange bar, experience a less dramatic increase than that of their male counterparts. The chart clearly demonstrates that the pandemic's aftermath has led to a higher incidence of deaths among the homeless, with gender-specific data revealing that the male homeless population has experienced a more severe increase in mortality. This gender-controlled analysis underlines the heightened risks and challenges that the homeless group has encountered during the global pandemic, with the differential impact on the males and females pointing to underlying social, economic and health-related disparities.

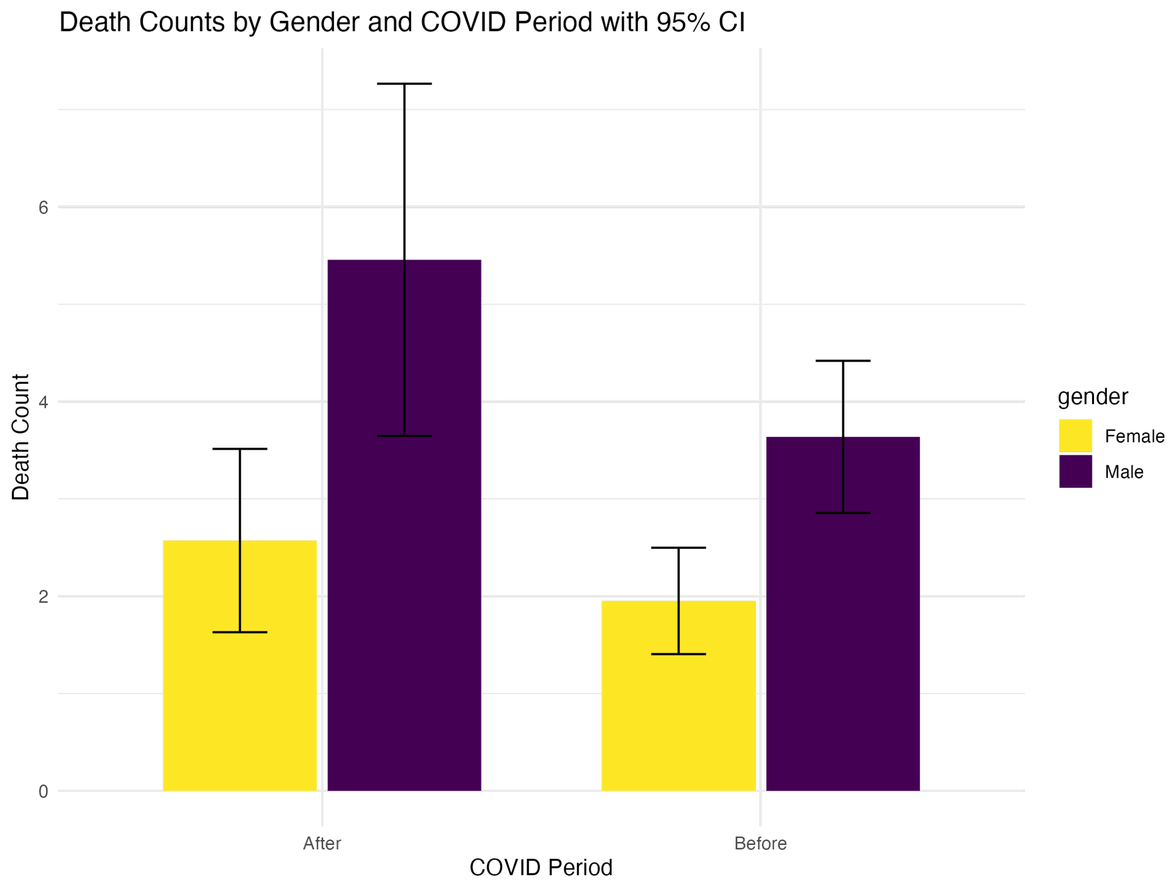

Figure 7: Death Counts by Gender and COVID Period with 95% CI

Difference in Difference (DiD) Results

The table presents the results of a regression analysis of the influence of COVID-19 on homeless death counts by gender. The regression model has been run twice: model 1 without controlling by gender, and model 2 with the control. Through the table and data included. We can gain a deeper understanding of the validity and significance of the data, which reveals aspects that visualization of data cannot show.

The coefficient of model 1 for the Post-COVID Period is -0.620, but it's not statistically significant since its p-value is larger than 0.1, suggesting that there is no significant change in the death counts post-COVID when not controlling for other variables. The coefficient for males is significantly positive (2.884) at the 5% levels, suggesting that male death counts are substantially higher than female death counts in the pre-COVID period. However, the interaction term between the post-COVID period and being male is not statistically significant (-1.199), including that the increase in death counts after pandemic is not significantly different for males compared to females.

The second model, which includes controls for the cause of death and age, shows a light substantial increase (0.040) in death counts during the post-COVID period compared to the pre-COVID period, but this increase is not statistically significant enough to provide strong evidence that the death counts have increased after COVID-19 once we account for these additional variables. However, male death counts are significantly higher (3.575) at less than 0.01 level, reinforcing the observations from the first model. The interaction term is negative (-1.807) and not statistically significant, implying that controlling for cause of death and age does not change the relationship between gender and death counts in the post-COVID period.

Table 2: Impact of COVID-19 on Homeless Death Counts by Gender

	Death Count	
	(1)	(2)
treatmentBefore	−0.620	0.040
	(1.212)	(1.074)
genderMale	2.884***	3.575***
	(1.016)	(0.898)
treatmentBefore:genderMale	−1.199	−1.807
	(1.526)	(1.343)
Constant	2.571***	−1.280
	(0.818)	(1.577)
Cause of Death Controlled	No	Yes
Age Controlled	No	Yes
Note:		*p<0.1; **p<0.05; ***p<0.01
	This table compares the difference in death counts between genders before and after the onset of COVID-19, with and without controls for cause of death and age.	

In summarizing the outcomes from the regression analysis, the data indicates a persistent association between genders and homeless death counts, implying that males are consistently more represented in these statistics. It is worthy to point out the outbreak of COVID-19 appears to have had little influence on the pre-existing gender difference in death counts among the homeless. This finding indicates that while COVID-19 has undoubtedly had a broad and profound impact on global health and mortality rates, but the analysis does not show that it also has significantly exacerbated the gender gap in mortality among the homeless population in Toronto. It is crucial to note that these results are determined by the accuracy and completeness of the reported death counts and the actual impacts may vary with additional context or data.

Discussion

Exploring the Significance of Our Findings

The analytical significance of examining the death counts among Toronto's homeless population during the pandemic are illustrations of how public health crises can exacerbate the vulnerabilities of already marginalized groups. Especially our findings underscore the association between gender and mortality within the homeless groups. These data and results provide an essential narrative that may inform public health strategies and services, and policymaking to better support the vulnerable homeless community during time of crisis.

However, our observation related to gender aspects, does not inherently indicate higher individual vulnerability for either gender group, it underscores the need for gender-specific data and targeted resources to address the unique challenges faced by both male and female homeless. In addition, these observations also highlight the importance of considering demographic proportions in public health research, A deeper understanding of the structure of the homeless people by gender could reveal insights into differential access to resources, health care, and exposure to risk factors that contribute to mortality. Further research that accounts for the percentage distribution of gender within the homeless population would provide a more detailed understanding of the gender dynamics and adjust supportive interventions more effectively.

Limitations

While our research provides insights into the death counts of homeless group before and after pandemic, particularly in relation to gender, it is essential to acknowledge several limitations that could influence the results drawn from our analysis.

Firstly, the integrity of our conclusions is related to the quality and accuracy of the data used in the study. If the data on death counts is not comprehensive or if there are inaccuracies in recording the data, this could lead to misinterpretations of the true impact of the pandemic on different genders within the homeless population. Incomplete data might fail to capture all cases of mortality among homeless, potentially underestimating the actual number of deaths.

Secondly, the regression analysis does not account for all possible confounding variables that affect death counts. This is also a limitation of the source data, which does not include all the variables that could affect the death count. Factors such as socio-economic status, access to healthcare, and geographic location are just a few examples of variables that were not controlled for but could significantly affect the outcomes. For example, whether the ratio of male to female among homeless community also counts for a variable in death. The absence of these variables means that we might not fully understand the relationships.

Finally, the local context of the pandemic in Toronto, including specific policy responses, the capacity of the healthcare systems of different regions, and public health measures implemented, had a significant influence on death counts among the homeless. While our findings focus on Toronto, they may not be directly applicable to other cities or regions due to differences in regional policies, healthcare systems, and public health strategies. It is crucial to note that our results apply only to the specific context of Toronto during the pandemic.

Ethical Concerns and Bias

The research of death counts of homeless community before and after pandemic could raise potential biases and ethical concerns that might happened and need to be addressed to ensure the applicability of the research. Ensuring that research methods and data interpretation do not include stereotypes or misrepresentations is crucial.

The comparison of pre and post-pandemic period reveals the increase in homeless death counts, ethical concerns cannot be overlooked. The research should extend beyond data representation, advocating for systematic policy change to address the fundamental causes of homelessness, such as poverty, lack of affordable housing, insufficient mental health support, to avoid increase in homeless death counts. While presenting data on the higher mortality rates among male homeless, it is critical not to ignore the struggles faced by female homeless. Additionally, researchers have a duty to handle sensitive information with the respects for the deceased and their families, ensuring that privacy and the dignity of individuals is maintained.

Regarding bias, the potential for data collection to skew the reality must be acknowledged. We must consider if deaths among the homeless accurately reported and attributed, and data fully capture the scope of issue. There is a concern that homeless individuals who die without interaction with medical systems may be not collected and lead to an underestimated of the true death counts. This potential underrepresentation calls for a critical examination of the methodologies employed in data collection and necessitates a vigilant approach to interpreting the results to avoid misinforming policy or public perception.

Future Research Directions

Future research directions stemming from our study topic should aim to deeper understandings of the dynamics of homeliness in relation to health crises. Prospective studies could focus on the constant effects of pandemic and investigate the success of interventions. Particularly, understanding the gender difference in mortality rates, and qualitative research, such as interviews which illuminate the personal experiences of the homeless during such crisis, should be paid more attention. By widening a range of socio-economic variables, future studies could provide more comprehensive insights that are critical to the development of specific policies and support systems to effectively assist the homeless.

Conclusion

In conclusion, this study has quantitatively examined the impact of the COVID-19 pandemic on mortality among Toronto's homeless population, with a focus on gender disparities. The analysis, which applied a Difference in Differences approach, suggests an overall increase in deaths post-pandemic, with males experiencing a notably higher mortality rate. Although the increase in death counts after COVID-19 was not statistically significant when accounting for confounders such as cause of death and age, the gender gap in mortality persisted. These results highlight the importance of targeted public health interventions and policies to address the needs of homeless populations during health crises. These results reinforce the call for comprehensive public health strategies and targeted support systems to mitigate the pandemic's impact on the homeless population.

Reference

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Smith, Phillip L. F. Year of Publication. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.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*. Vol. 4. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. https://readr.tidyverse.org.

Wickham, Hadley, Evan Miller, and Danny Smith. 2023. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. https://haven.tidyverse.org.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. http://haozhu233.github.io/kableExtra/.