

An Analytical Exploration of Homelessness and Suicide in Toronto*

Hailey Jang

2024-03-14

This paper presents a study focused on addressing the critical issue of suicide among Toronto’s homeless population, a significant public health challenge. We developed a risk stratification model using sociodemographic data to identify homeless individuals at high risk for suicide, employing a generalized linear model that incorporates factors like sex and age. The findings, derived from data collected by Toronto Public Health since January 2017, offer insights into the patterns and predictors of suicide attempts within this vulnerable group. The importance of this research lies in its potential to guide interventions, enabling more effective prevention strategies for suicide among the homeless, particularly in settings where immediate access to mental health professionals is limited.

Table of contents

1	Introduction	2
2	Data	3
2.1	Data source	3
2.1.1	Data measurment	3
3	Model	4
3.1	Model set-up	4
3.1.1	Model justification	6
4	Results	6
5	Discussion	6

*Code and data are available at: <https://github.com/Hailey-Jang/Suicide-and-Homelessness.git>

1 Introduction

Homelessness represents a significant public health crisis, with the intersection of inadequate housing and mental health issues exacerbating the risk of suicide among this population. Individuals experiencing homelessness face a 2- to 6-fold increased risk of suicide compared to the general population (Sinyor et al. (2017)). This alarming statistic not only highlights the vulnerability of homeless individuals but also underscores the urgent need for targeted interventions and preventive measures. This paper delves into the development of a risk stratification model designed to identify homeless individuals at an elevated risk of suicide. By integrating sociodemographic data within a generalized linear model, focusing on sex and age as primary factors, our research aims to provide a predictive tool for healthcare providers and social services. The motivation behind this study stems from the critical gap in existing research regarding effective, data-driven strategies for suicide prevention in the homeless community.

Our research methodology involved the analysis of comprehensive data collected by Toronto Public Health, starting from January 2017. This dataset not only provided a foundation for our model but also enriched our understanding of the prevalence and causes of suicide among the homeless. The outcomes of our study revealed significant patterns and predictors that can enhance the effectiveness of suicide prevention strategies, offering a new perspective on addressing this public health challenge. The significance of our findings extends beyond academic interest; they offer practical insights for policymakers, healthcare providers, and social workers who are directly involved in supporting the homeless population. By identifying high-risk individuals, our model can inform targeted interventions, potentially saving lives and allocating resources more efficiently.

The structure of the paper is organized as follows: The data section delves into the broader context of the dataset, emphasizing the crucial aspects of measurement relevant to our study. The results section is comprehensive, encompassing various illustrative tables and graphs alongside detailed statistical analyses. Subsequently, the model section elucidates the setup and justification of our chosen modelling approach, ensuring clarity and transparency of our methodology. Lastly, the discussion articulates the study's contributions to our understanding of homelessness and suicide and outlines directions for future research, ensuring a thorough contemplation of the study's broader implications.

Table 1: Summary of Suicides by Age Group

	Age_group	Total Counts
[H]	40-59	376
	60+	255
	20-39	225
	<20	8

2 Data

2.1 Data source

The study was performed using data from the City of Toronto’s **opendatatoronto** database portal, accessed through the ‘opendatatoronto’ package and processed using the statistical programming environment **R**. The **tidyverse** package facilitated the data and the **kableExtra** package was utilized for enhancing table presentations. Specific to this study, the LaTeX package was implemented in the R markdown setting to ensure stable positioning of figures and tables.

2.1.1 Data measurment

Initiated in January 2017, Toronto Public Health (TPH) embarked on a systematic record-keeping of homeless mortality to gain a clearer understanding of the prevalence and causative trends of these incidents. The dataset comprises variables such as the year of death, cause of death, age group, gender, and number of deaths. It classifies individuals into age categories spanning 20 years, starting from 20 years to 60+ years, with gender recorded as either Male or Female. The dataset enumerates the deaths annually from 2017 through 2023, categorizing the causes into distinct classifications like Accident, Drug Toxicity”, Suicide, among others.

Focusing on suicide-related fatalities, the dataset was refined to exclude entries marked as Unknown or empty. This filtration led to the construction of two specialized datasets: one delineating the yearly suicide death toll segregated by gender and another by age group. This reorganization necessitated aggregating individual counts from each report into a consolidated figure for these subgroups, facilitating a focused examination of suicide trends across different demographics over the years.

Table Table 1 presents the refined data, specifically spotlighting instances of suicide by age groups, and depicting the temporal and numerical specifics of each case.

Table Table 2 presents the refined data, specifically spotlighting instances of suicide by gender, and depicting the temporal and numerical specifics of each case.

Table 2

Year	sex	count
2017	Male	69
2018	Male	65
2019	Male	87
2020	Male	107
2021	Male	153
2022	Male	129
2023	Male	64
2017	Female	21

Summary of suicides Rates by Gender

3 Model

3.1 Model set-up

This section elucidates the development of a logistic regression model, tailored to predict the likelihood of suicide among homeless individuals using demographic factors. Before delving into the model, we visually explore the relationship between suicide rates and demographic variables—age and gender—using ggplot2.

We employ (**graph-suicide-gender-group?**) to create a comprehensive visual representation, showcasing the relationship between suicide rates and the key demographic variables, age and gender. This visualization aids in understanding the data distribution and any apparent trends that could influence the model.

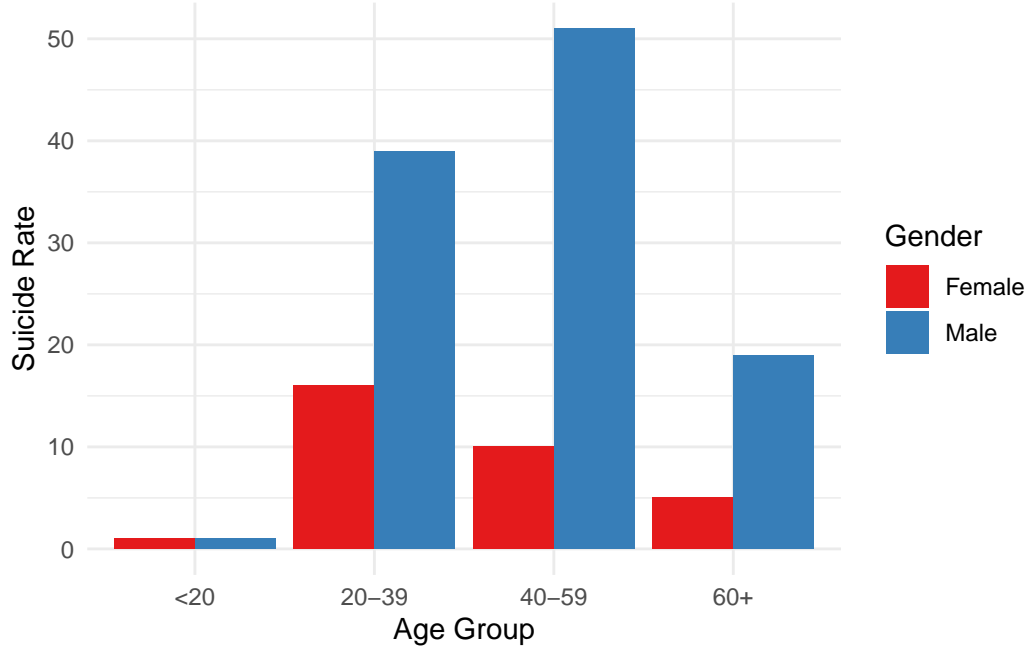


Figure 1 Summary of Suicide Rates by Age and Gender

The logistic regression model is designed to predict the probability of suicide cases among homeless individuals, utilizing sociodemographic factors such as age and gender. In a Bayesian logistic regression context, the model's setup can be represented with the following hierarchical structure:

$$y_i | p_i \sim \text{Bernoulli}(p_i) \quad (1)$$

$$\log \left(\frac{p_i}{1 - p_i} \right) = \alpha + \beta_1 \times \text{AgeGroup}_{i1} + \beta_2 \times \text{AgeGroup}_{i2} + \beta_3 \times \text{Gender}_i \quad (2)$$

$$\alpha \sim \text{Normal}(0, 2.5) \quad (3)$$

$$\beta_j \sim \text{Normal}(0, 2.5) \text{ for } j = 1, 2, 3 \quad (4)$$

$$(5)$$

In this adapted model:

- y_i represents the binary outcome for each individual (i.e., the presence or absence of a suicide case).
- p_i is the probability of observing a suicide case for the i th individual, linked to the predictors through the logistic function.

- The coefficients α, β_1, β_2 , and β_3 are assigned Normal prior distributions, reflecting our prior beliefs about these parameters' distributions before observing the data. The Normal priors are centered at 0 with a standard deviation of 2.5, indicating moderate certainty in the prior information.
- The logit link function (log-odds) is the natural logarithm of the odds $p_i / 1 - p_i$ and linearly relates the predictors to the probability of the outcome.

3.1.1 Model justification

This Bayesian logistic regression framework not only estimates the parameters but also quantifies uncertainty in the estimates, providing a more comprehensive understanding of the model's predictions and the effects of the predictors. The use of Normal priors for the coefficients is a standard choice in Bayesian modeling, offering flexibility and conjugacy that facilitate both analytical and computational solutions. The hierarchical model structure allows for the incorporation of additional layers of complexity and variability, accommodating more nuanced relationships and potential hierarchical data structures.

4 Results

5 Discussion

References

Sinyor, Mark, Nicole Kozloff, Catherine Reis, and Ayal Schaffer. 2017. "An Observational Study of Suicide Death in Homeless and Precariously Housed People in Toronto." *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie* 62 (May): 706743717705354. <https://doi.org/10.1177/0706743717705354>.