Canada_leading_causes_of_death*

Scarlet Ruoxian Wu

March 12, 2024

Table of contents

1 Introduction

Mortality involves the complex interplay between lifestyle, environment, healthcare access, and genetic predisposition. Every year, mortality statistics shed light on a nation's health and the effectiveness of its healthcare system. The Canadian government has kept records of causes of death from 2000 to 2022. This study aims to examine the leading causes of death during this period, identify patterns, highlight healthcare challenges, and suggest potential improvements in public health policy. The primary estimate of this study is the annual number of deaths attributed to various causes.

This analysis aims to understand the complexity of mortality data and the variability in death counts. The study employs the simulation of datasets using the negative binomial distribution. This particular distribution is well-suited to model count data, especially in cases where the data exhibit over-dispersion. By simulating datasets, this study aims to gain insights into the potential distributions of deaths by cause and year, allowing for a deeper understanding of the trends and variability in the actual data.

This study employs both Poisson and negative binomial regression models to analyze the Candian death data. The reason to use both models is to accommodate the variable nature of

*Code and data are available at: https://github.com/ScarletWu/Canada_leading_cause_of_death.git mortality counts. While Poisson regression is a conventional choice for count data, it assumes equality between the mean and variance of the data, which is often not the case in mortality statistics due to over-dispersion. In contrast, the negative binomial model provides greater flexibility by accommodating over-dispersion, making it a more realistic tool for analyzing the complex nature of mortality data.

The critical comparison between Poisson and negative binomial models in this analysis is not merely a statistical preference but a methodological necessity. It underscores the importance of selecting a model that accurately reflects the data's underlying distribution, ensuring the reliability and validity of the findings. This comparison is instrumental in identifying the model that best captures the nuances of mortality data, thereby providing a solid foundation for concluding the leading causes of death in Canada and the potential implications for public health policies and initiatives.

In summary, this paper comprehensively analyzes mortality data from Canada by using Poisson and negative binomial models to capture the nuances of death-related statistics. The primary estimand —the annual number of deaths by various causes—sets the stage for a detailed exploration of mortality patterns within Canada. Besides quantifying the burden of mortality, this study reveals underlying trends that can inform future health policies and interventions to reduce preventable deaths and improve Canadians' health and well-being.

2 Data

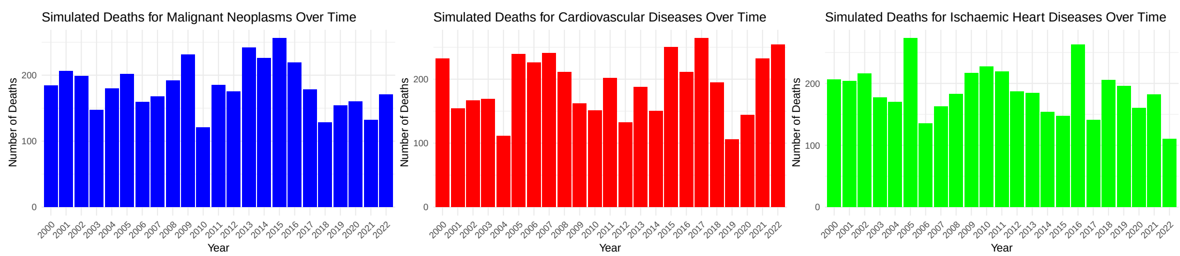
This analysis examines mortality data in Canada from 2000 to 2022 obtained from the comprehensive database maintained by Statistics Canada. The data includes the annual number of deaths categorized by different causes, offering a detailed view of mortality trends in the country. The primary variables of interest are the year of death, the cause of death according to the ICD-10 classification, and the total number of deaths attributed to each cause. These variables provide an overview of how mortality patterns have changed over the past two decades, reflecting the impact of healthcare advancements, public health initiatives, and emerging challenges.

To prepare the dataset for analysis, we made significant efforts to ensure that the data was clean and appropriately structured. This included truncating the cause of death descriptions for readability and consistency, which helped ensure that our analyses were accurate and easy to interpret. We chose the Canadian dataset specifically due to its comprehensive coverage and the high quality of data reporting standards maintained by Statistics Canada. This allowed us to gain a nuanced understanding of mortality within the Canadian context, which may differ from individual provincial trends due to various socio-economic, environmental, and healthcare factors.

To explore and visualize the data, we used a suite of R packages, each chosen for its specific capabilities. The dplyr package facilitated efficient data manipulation, while ggplot2 helped us visualize trends and patterns in mortality. We used Poisson and negative binomial regression techniques to model the data, and rstanarm was chosen for its advanced Bayesian modeling capabilities. The modelsummary package efficiently summarized the results, providing clear and concise insights into the findings. We also utilized the broom and broom.mixed packages to tidy the outputs of our statistical models, making the results more accessible and interpretable.

By systematically examining the variables within the dataset and using rigorous statistical modeling, this analysis provides a deeper understanding of mortality trends in Canada, underscoring the critical role of data-driven approaches in public health planning and evaluation.

3 Result



(a) Malignant Neoplasms (b) Cardiovascular Diseases (c) Ischaemic Heart Diseases

Figure 1: negative binomial Death simulation for each cause of death

Table 1: Top-ten causes of death in Canada in 2022

Year	Cause	Deaths	Rank	Years
2022	Malignant neoplasms [C00-C97]	82412	1	23
2022	Major cardiovascular diseas...	76639	2	23
2022	Diseases of heart [I00-I09,...	57357	3	23
2022	Ischaemic heart diseases [I...	34830	4	23
2022	Dementia [F010-F019, F03]	25994	5	6
2022	Unspecified dementia [F03]	23896	6	6
2022	Other forms of chronic isch...	20126	7	23
2022	COVID-19 [U07.1, U07.2, U10.9]	19716	8	3
2022	Malignant neoplasms of trac...	19151	9	23
2022	Other heart diseases [I26-I51]	18913	10	23

```
poisson <- loo(cause_of_death_poisson, cores = 2)
```

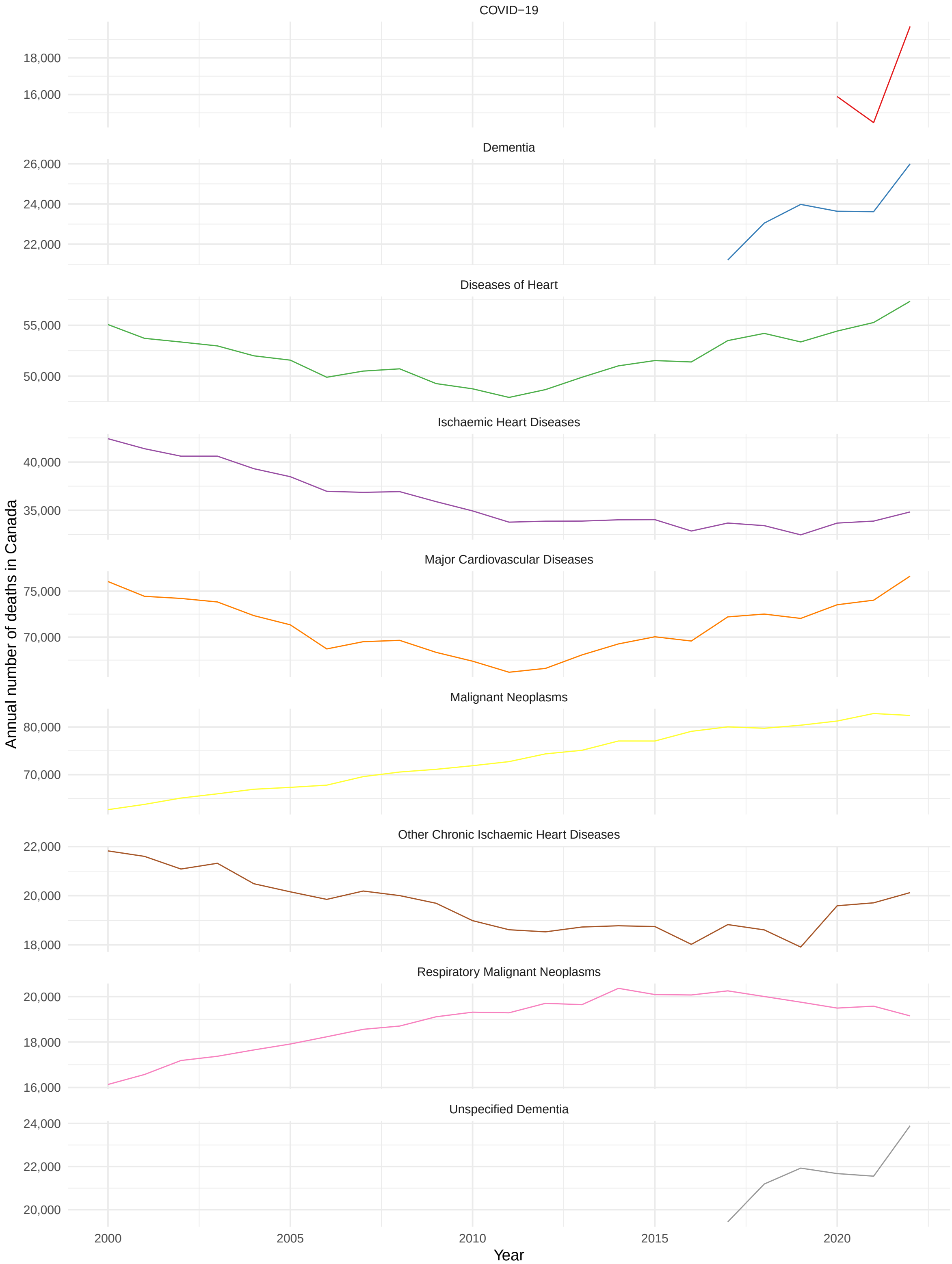

COVID−19
18,000
16,000
Dementia
26,000
24,000
22,000
Diseases of Heart
55,000
50,000
Ischaemic Heart Diseases
40,000
35,000
Major Cardiovascular Diseases
75,000
70,000
Malignant Neoplasms
80,000
70,000
Other Chronic Ischaemic Heart Diseases
22,000
20,000
18,000
Respiratory Malignant Neoplasms
20,000
18,000
16,000
Unspecified Dementia
24,000
22,000
20,000
2000
2005
2010
2015
2020
Year
Annual number of deaths in Canada

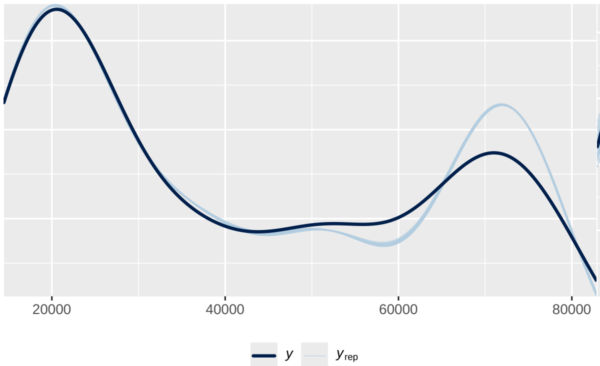
Figure 2

Table 2: Summary statistics of the number of yearly deaths, by cause, in Canada

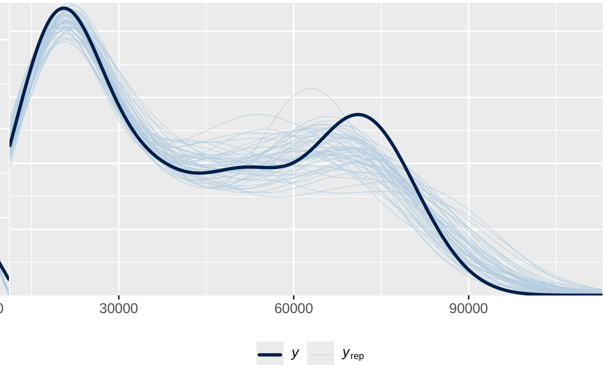
	Min	Mean	Max	SD	Var	N
value	14 466	42 831	82 822	22 480	505 344 387	153

Table 3: Modeling the most prevalent cause of deaths in Canada, 2001-2020

	Poisson	Negative Binomial
Malignant Neoplasms	1.479	1.472
		(0.082)
Diseases of Heart	1.137	1.129
		(0.083)
Respiratory Malignant Neoplasms	0.123	0.116
		(0.082)
Major Cardiovascular Diseases	1.450	1.443
		(0.085)
Ischaemic Heart Diseases	0.770	0.761
		(0.083)
Unspecified Dementia	0.259	0.253
		(0.096)
Num.Obs.	153	153
Log.Lik.	−14 947.833	−1463.624
ELPD	−15 593.5	−1468.3
ELPD s.e.	1752.0	7.1
LOOIC	31 186.9	2936.6
LOOIC s.e.	3504.0	14.3
WAIC	31 838.1	2936.3
RMSE	3153.30	3153.84



(a) Poisson model



(b) Negative binomial model

Figure 3: Comparing posterior prediction checks for Poisson and negative binomial models

```
neg_binomial <- loo(cause_of_death_neg_binomial, cores = 2)

loo_compare(poisson, neg_binomial)
```

```
                            elpd_diff se_diff
cause_of_death_neg_binomial      0.0       0.0
cause_of_death_poisson      -14125.2    1748.0
```

#Discussion One important lesson we can learn from this analysis is how different diseases affect mortality rates. The comparison of different models shows that certain causes, such as heart diseases and cancer, consistently rank as the leading causes of death. This information emphasizes the chronic nature of these conditions and may also reflect the impact of an aging population on healthcare systems. The inclusion of COVID-19 as a separate category in recent years highlights how emerging health threats can quickly alter mortality rates.

Another valuable insight we gain from this investigation is how mortality causes change over time. The graph that shows the annual number of deaths per cause reveals the dynamics of how these causes evolve. Some causes, such as cancer and heart diseases, show a steady increase that aligns with population growth and aging. On the other hand, causes such as COVID-19 show a sudden spike due to the pandemic. This trend information can inform healthcare planning, such as resource allocation, to address these evolving challenges.

However, this study has limitations. While the models used are adept at handling the data, they may not fully account for all the complexities of real-world scenarios, such as socioeconomic factors, healthcare access, and lifestyle changes over time. The Poisson model, in particular, with its assumption of equal mean and variance, may oversimplify the complexity of mortality data. Furthermore, our understanding is limited to the data provided by Statistics Canada, which may have its own reporting biases and gaps.

Moving forward, it is important to continue this research by incorporating more nuanced data that considers the broader determinants of health. Future studies could benefit from a more granular approach, such as analyzing sub-populations, to understand disparities in mortality rates. Additionally, as new health threats emerge and societal factors evolve, continuous updating and refining of models will be crucial. Longitudinal studies could also shed light on the lifetime risks of various demographics, providing a more comprehensive picture of mortality in Canada.

In conclusion, this study is a starting point for a deeper understanding of mortality and its causes. It highlights the persistent and emerging health threats that dominate mortality statistics and emphasizes the importance of tailored health interventions. Further research can provide more detailed insights, ultimately guiding better health policies and outcomes.

#Reference