My title*

My subtitle if needed

Yanyu Wu Ziyi Liu Hechen Zhang

March 16, 2024

We examined the distribution of the top 30 causes of death for each year between 2001 and 2022 in Alberta, Canada. The conifer model and negative binomial regression were used to analyse the long-term leading causes of death and the sudden emergence of specific causes of death. According to our findings, negative binomial regression improves our ability to predict outcomes when the data is too spread out by fitting the data more accurately. The results could not only help policymakers design more effective preventive measures to reduce mortality from these conditions, but also help researchers and policymakers make more precise decisions.

Table of contents

*Code and data are available at: https://github.com/HechenZ123/Cause-of-Deaths-in-Alberta.git

1.0 Introduction

The mortality rate, often referred to as the death rate, represents an approximation of the fraction of a population that dies within a given time frame (Porta 2014). Mortality rates can serve as a crucial indicator of a population's health status, and it also reveals the impact of diseases and other health-related issues over a period of time. This paper explores the leading causes of death in Alberta for crafting effective public health strategies and policies and understanding the most significant health threats affecting a population for researchers(Alberta 2015). By identifying the main causes of mortality, health authorities can prioritise research funding towards diseases and conditions that have the highest impact on community health and lifespan (Vargas et al. 2019).

As discussed in the data section, we used data from Service Alberta (Alberta 2015) on the leading causes of deaths, in which the five most significant causes in 2022 were analysed. These five causes are Organic dementia, Other causes not clearly defined, COVID-19, and Cancers of the trachea, bronchus, and lungs. It was noted that, among the examples mentioned, the negative binomial regression is more accurate compared to the Poisson model, while Poisson regression is prone to errors.

1.1 Importing Important Packages.

In this analysis, we employ a range of R (R Core Team 2023) packages tailored for data cleaning, transformation, analysis, and reporting. `Tidyverse` by Wickham et al. (2019) is used for data wrangling, `janitor` package by Firke (2021) is used for data cleaning operations, and `knitr` by Xie (2021) for data presentation in data tables.The following code section aims at importing the important packages that are essential for examining the missing values in the data set.We run the model in R R Core Team (2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`. For comprehensive mixed effects model analysis, we leverage the `broom.mixed` package (Bolker and Robinson 2022), which extends the `broom` package functionalities to mixed models, facilitating the extraction, tidying, and representation of model outputs. Furthermore, the `modelsummary` package (Arel-Bundock 2022) provides tools for creating customizable summary tables of model results, enhancing the interpretability and dissemination of statistical findings. By calculating the LOO-CV scores for different models with `loo` (Yao et al. 2017), we could compare them based on their out-of-sample predictive accuracy. Lower values of LOOIC indicate better model performance. The following code sections aim to import these crucial packages, essential for conducting a thorough analysis and addressing the research questions at hand, while ensuring data integrity and transparent reporting of results.

2.0 Data

Our data is of leading causes of death (Figure 1), from Alberta (2015).

2.1 Data Sources

The Alberta Government created the dataset.The official website of the Government of Alberta provides a single platform for services and information pertaining to Alberta, Canada. It provides information on government news, jobs, social services, health, and education, making it an essential tool for both locals and tourists to be updated about provincial issues.

2.2 Variables

Order by total number of deaths and a ranking of the top 30 causes of death in Alberta each year.Our table lists the top eight causes of death in Alberta in 2022. Examine each variable in detail: *Year*: This denotes the data gathering year, which for all entries is 2022. *Cause*: This represents the medical condition or event that led to death. The causes listed are shown in Figure 1: `Organic dementia All other forms of chronic...`, `Other ill-defined and unknown...`, `COVID-19, virus identified`, `Malignant neoplasms of the trachea, bronchus, and lung`, `Acute myocardial infarction`, `Accidental poisoning by and...`, `Other chronic obstructive pulmonary diseases`, *Ranking*: This is a ranking by the number of deaths caused by each disease, with 1 being the highest. *Deaths*: The number of deaths attributed to each cause. *Year*: Indicates the number of years in which data was collected for that reason.

Year	Cause	Ranking	Deaths	Years
2022	Organic dementia	1	2,377	22
2022	All other forms of chronic …	2	2,098	22
2022	Other ill-defined and unkno…	3	1,714	4
2022	COVID-19, virus identified	4	1,547	3
2022	Malignant neoplasms of trac…	5	1,523	22
2022	Acute myocardial infarction	6	1,240	22
2022	Accidental poisoning by and…	7	1,200	10
2022	Other chronic obstructive p…	8	1,183	22

Figure 1: Top-teight causes of death in Alberta in 2022

2.3 Data Description

It accurately reflects health issues specific to the region, but this information is somewhat limited by location. Health policies vary from country to country, and in some regions and countries, health coverage may be less developed than in others, and low-income groups may not be able to afford basic medical care, which increases their risk of chronic diseases and other health problems. The main types of employment in a city also have an impact on health; In an industrial city, some jobs may require exposure to hazardous substances or strenuous physical labor, which increases the likelihood of developing occupational diseases. Risks associated with the workplace may lead to an increase in diseases such as cardiovascular and respiratory diseases. In addition, communities with higher levels of education and economic prosperity generally have stronger social support systems, which are critical for disease prevention.

The demographics of a society also have a significant impact on cause-of-death data. Age distribution is important because certain age groups are more susceptible to certain diseases. For example, young children may be more susceptible to infectious diseases, while older people may be more susceptible to degenerative and chronic diseases. Given that certain medical conditions are more common in one sex than the other - men may have higher rates of heart disease, while women may have higher rates of certain cancers - the sex ratio may also have an impact on causes of death. In addition, differences in the ethnic makeup of the population may be related to lifestyle choices, genetic makeup, and disease risk. For example, genetic factors may contribute to the higher incidence of certain chronic diseases in particular ethnic groups. Elucidating how resources are allocated to public health problems therefore requires a thorough understanding and study of population characteristics.

We can also see that in the data table Figure 1 for 2022, COVID-19 ranks fourth, with 1,547 deaths. Alberta's cause-of-death data proves that the coronavirus pandemic is one of the most serious public health crises of the early 21st century, with global implications. This figure not only speaks to the outbreak's death rate, but also suggests that the region's health care system may be under strain. The COVID-19 pandemic has disrupted the distribution of deaths from previously common causes, which may include some chronic diseases that have long held the top spot, such as cancer and heart disease, even though it was not the leading cause of death in the years shown in the table.

As large datasets are difficult to observe, this report will analyze only specific aspects. The original dataset contains eight different causes of death. Only the first five causes will be analyzed in this paper.

For simplicity we restrict ourselves to the five most common causes of death in 2022 of those that have been present every year.

Table 1 # 3.0 Model

The goal of our modelling strategy is twofold. Firstly,…

Table 1: Top Five Leading causes of death in 2022

```
[1] "Organic dementia"                "All other forms of chronic ..."
[3] "Other ill-defined and unkno..." "COVID-19, virus identified"
[5] "Malignant neoplasms of trac..."
```

Table 2

Here we briefly describe the Bayesian analysis model used to investigate… Background details and diagnostics are included in Appendix

3.1 Model set-up

Define y_i as the number of deaths recorded in Alberta. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

Poisson model:

$$y_i|\lambda_i \sim \text{Poisson}(\lambda_i) \tag{1}$$
$$log(\lambda_i) = \beta_0 + \beta_1 \cdot x_i \tag{2}$$
$$\beta_0 \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta_1 \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\tag{5}$$

Negative binomial model:

$$y_i|\lambda_i, \theta \sim \text{NegativeBinomial}(\mu_i, \theta) \tag{6}$$
$$log(\mu_i) = \beta_0 + \beta_1 \cdot x_i \tag{7}$$
$$\beta_0 \sim \text{Normal}(0, 2.5) \tag{8}$$
$$\beta_1 \sim \text{Normal}(0, 2.5) \tag{9}$$
$$\tag{10}$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ.

Results

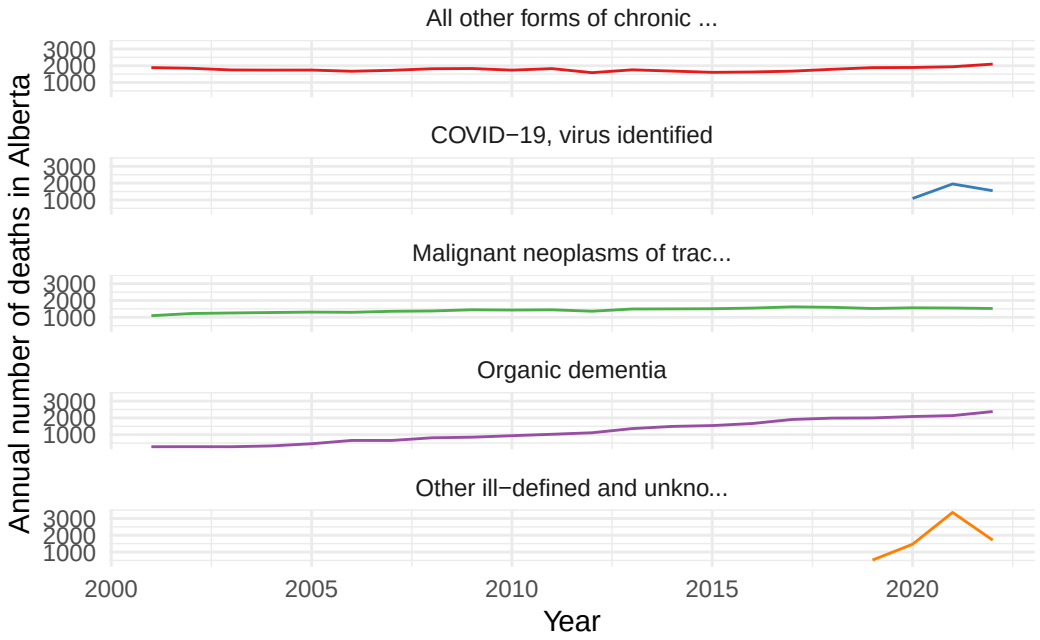


Figure 2: Annual number of deaths for the top-five causes in 2022, since 2001, for Alberta, Canada

Table 3: Summary by cause of the number of yearly deaths in Alberta, Canada

Min	Mean	Max	SD	Var	N
280	1483.411	3362	520.4136	270830.3	73

Our results are summarized in Table 4.

Table 4: Modeling the most prevalent cause of deaths in Alberta, 2001-2022

	Poisson	Negative binomial
(Intercept)	7.484	7.482
		(0.093)
causeCOVID-19, virus identified	−0.152	−0.129
		(0.262)
causeMalignant neoplasms of trac...	−0.223	−0.220
		(0.131)
causeOrganic dementia	−0.400	−0.396
		(0.131)
causeOther ill-defined and unkno...	−0.007	0.017
		(0.241)
Num.Obs.	73	73
Log.Lik.	−6421.556	−565.317
ELPD	−6731.0	−570.5
ELPD s.e.	1418.0	6.3
LOOIC	13 462.1	1140.9
LOOIC s.e.	2836.0	12.6
WAIC	14 288.6	1140.4
RMSE	457.92	458.07

Discussion

Addressing Public Health Challenges in Alberta: Strategies for Health Policy and Social Regulation

Based on the top five causes of death in Alberta, it is imperative that we consider initiatives to improve health policy and social regulation to address health challenges, prioritizing public health issues. Looking at the data, given the significant impact of Covid-19 on mortality rates, it is crucial to continue efforts to implement public health interventions. This includes increasing mass vaccination activities, continuing to promote mask-wearing and social distancing measures, and enhancing testing and contact tracing capabilities. Furthermore, it is essential to ensure that the healthcare system has sufficient capacity to handle an increase in cases. It is worth noting that in the coming years, due to the passage of several years since the virus initially emerged, Covid-19 may not continue to be such a significant cause of death, as the virus gradually becomes less virulent or severe (Talic et al. 2021). Implementing policies for cancer prevention and control can help reduce mortality from malignant neoplasms of the trachea, bronchus, and lung. This may include implementing tobacco control measures, such as increasing tobacco product taxes, comprehensive smoking cessation programs, and restricting tobacco advertising and promotion. Additionally, promoting healthy lifestyles, early cancer screening programs, and providing high-quality cancer treatment services are also crucial (Eastman 2023).

Second discussion point

Third discussion point

Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

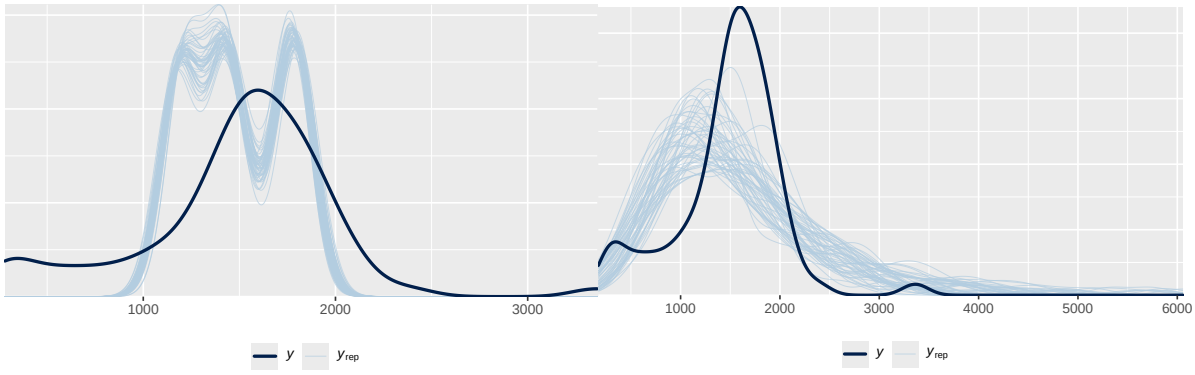
Additional data details

Model details

Posterior predictive check

In Figure 3a we implement a posterior predictive check. This shows...

In Figure 3b we compare the posterior with the prior. This shows...



(a) Poisson model

(b) Negative binomial model

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

```
Warning: Found 20 observations with a pareto_k > 0.7. With this many problematic observations
```

```
                                   elpd_diff se_diff
cause_of_death_alberta_neg_binomial     0.0       0.0
cause_of_death_alberta_poisson      -6160.6    1412.1
```

References

Alberta, Service. 2015. "Leading Causes of Death." *Leading Causes of Death - Open Government*. https://open.alberta.ca/dataset/leading-causes-of-death/resource/3e241965-fee3-400e-9652-07cfbf0c0bda.

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Bolker, Ben, and David Robinson. 2022. *Broom.mixed: Tidying Methods for Mixed Models*.

Eastman, Peggy. 2023. "NCI Releases New National Cancer Plan to Realize Vision of Cancer Moonshot." LWW.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Porta, Miquel. 2014. *A Dictionary of Epidemiology*. Oxford university press.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Talic, Stella, Shivangi Shah, Holly Wild, Danijela Gasevic, Ashika Maharaj, Zanfina Ademi, Xue Li, et al. 2021. "Effectiveness of Public Health Measures in Reducing the Incidence of Covid-19, SARS-CoV-2 Transmission, and Covid-19 Mortality: Systematic Review and Meta-Analysis." *Bmj* 375.

Vargas, Ashley J, Sheri D Schully, Jennifer Villani, Luis Ganoza Caballero, and David M Murray. 2019. "Assessment of Prevention Research Measuring Leading Risk Factors and Causes of Mortality and Disability Supported by the US National Institutes of Health." *JAMA Network Open* 2 (11): e1914718–18.

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.

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

Yao, Yuling, Aki Vehtari, Daniel Simpson, and Andrew Gelman. 2017. "Using Stacking to Average Bayesian Predictive Distributions." *Bayesian Analysis*. https://doi.org/10.1214/17-BA1091.