My title*

My subtitle if needed

Yanyu Wu   Ziyi Liu   Hechen Zhang

March 12, 2024

When the continuity assumption of linear regression does not apply, we can work with counting data using Poisson regression, which deals with discontinuous results but still incorporates predictor variables into the model in a linear manner. We demonstrate the validity of negative binomial regression through an extensive analysis of mortality in Alberta, Canada, where the data is too scattered for the classical method (Poisson regression). Based on our findings, negative binomial regression improves our ability to predict outcomes by fitting the data more precisely, especially in cases where the variance is substantially higher than the mean. This study highlights how important it is to use the right statistical models to get more accurate findings, which will ultimately enhance our understanding of patterns and trends in many academic fields.

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.

1.2. Data Overview.

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

Examining the top ten causes in 2021 reveals several notable findings (Figure 1). For example, ...

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

2.0 Data

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

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

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    280    1297    1547    1483    1757    3362
```

3.0 Model

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

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

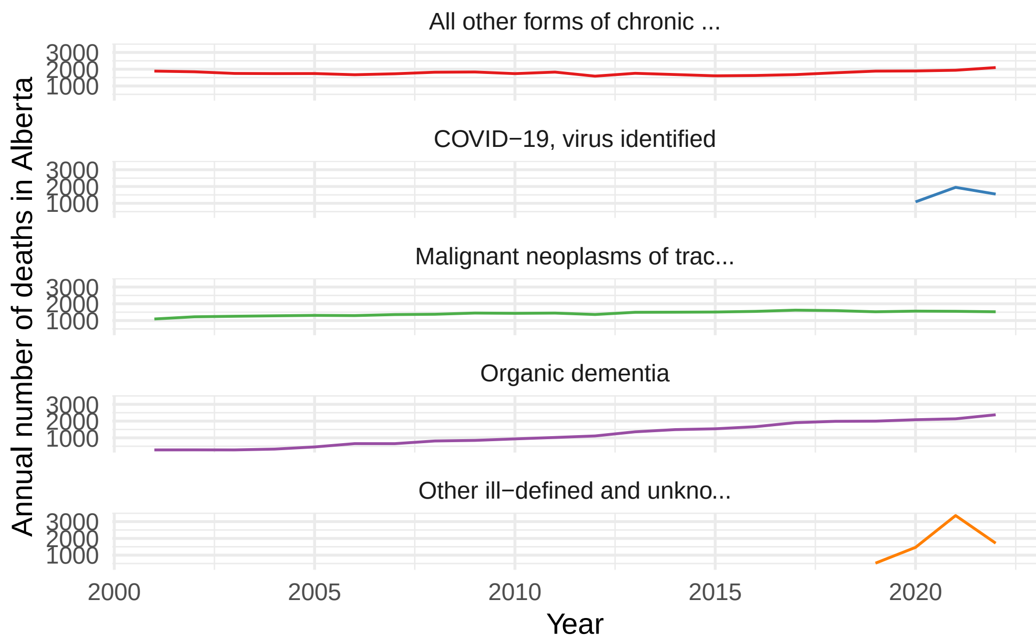


Figure 2: Annual number of deaths for the top-five causes in 2022, since 2001, for Alberta, Canada

Table 1

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$
\begin{aligned}
y_i|\mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) && (1)\\
\mu_i &= \alpha + \beta_i + \gamma_i && (2)\\
\alpha &\sim \text{Normal}(0, 2.5) && (3)\\
\beta &\sim \text{Normal}(0, 2.5) && (4)\\
\gamma &\sim \text{Normal}(0, 2.5) && (5)\\
\sigma &\sim \text{Exponential}(1) && (6)
\end{aligned}
$$

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`.

Table 2: 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

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

Our results are summarized in Table 2.

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

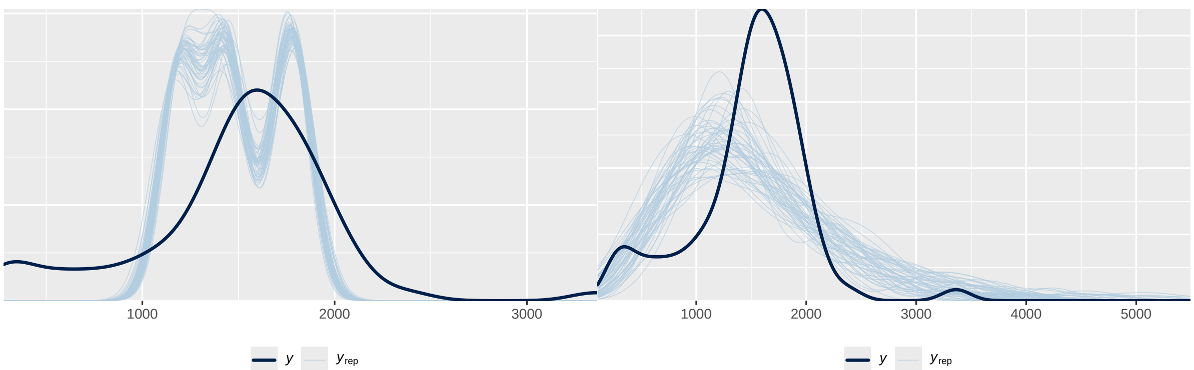
Additional data details

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

?@fig-stanareyouokay

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

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

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

loo_compare(poisson, neg_binomial)
```

```
                                    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.