Trends in Top 10 Causes of Death in Ontario*

Jacob Gilbert Liam Wall

March 16, 2024

Ontario, Canada, mortality data has been available on the internet for years and we have utilised data from 2000 to 2022 to predict the leading causes of death in 2023 to 2028. We examine the top ten leading causes of death and observe the trend in them over 23 years and train a model to use cause of death and the year as predictors for the number of yearly deaths for that cause. We find that year and cause prove to be very good predictors of total death for the years 2000 to 2022 and extends the trends observed into the future until 2028. However, this model does not take into account many of the external factors in medicine and society that are the true causes of these trends, and so we expect that this model explains the future of Ontario mortality if no intervention is made.

1 Introduction

In this paper, we attempt to analyze data on Ontario mortality from 2000 to 2022. In the analysis we create a model predicting the number of deaths per year for a certain cause. From this model we were able to create a prediction for the next five years of the top ten causes of death in Ontario and the corresponding predicted number of deaths.

The creation of this report was made possible by R Core Team (2023), Alexander (2023), Wickham et al. (2019) for data cleaning, Goodrich et al. (2022) for data modeling and posterior prediction checks, Wickham (2011) to test many aspects of the code used for the analysis, Zhu (2024) for adjusting and displaying code, Chang (2015) to download the data, and Canada (2023) for making the data available.

*Code and data are available at: https://github.com/JfpGilbert0/Ontario-CA-Mortality.

2 Data

The data is this paper was retrieved from Canada (2023). The data set contains information on Ontario residents' mortality from 2000 to 2022. The data set lists information on the top 43 leading causes of death each year. It has values for rank of leading cause of death in Ontario, total number of deaths per cause, percentage of deaths pertaining to each cause compared to the yearly total, and age-specific mortality rate. It lists all this information for males, females, and both sexes combined individually as well as for each age group of around 5 years and a combined statistics for all ages. The data set also lists annual information like total deaths per cause per year and total deaths per year. For the purposes of our analysis we will only look at the top ten causes of death in Ontario per year **?@fig-all_values**.

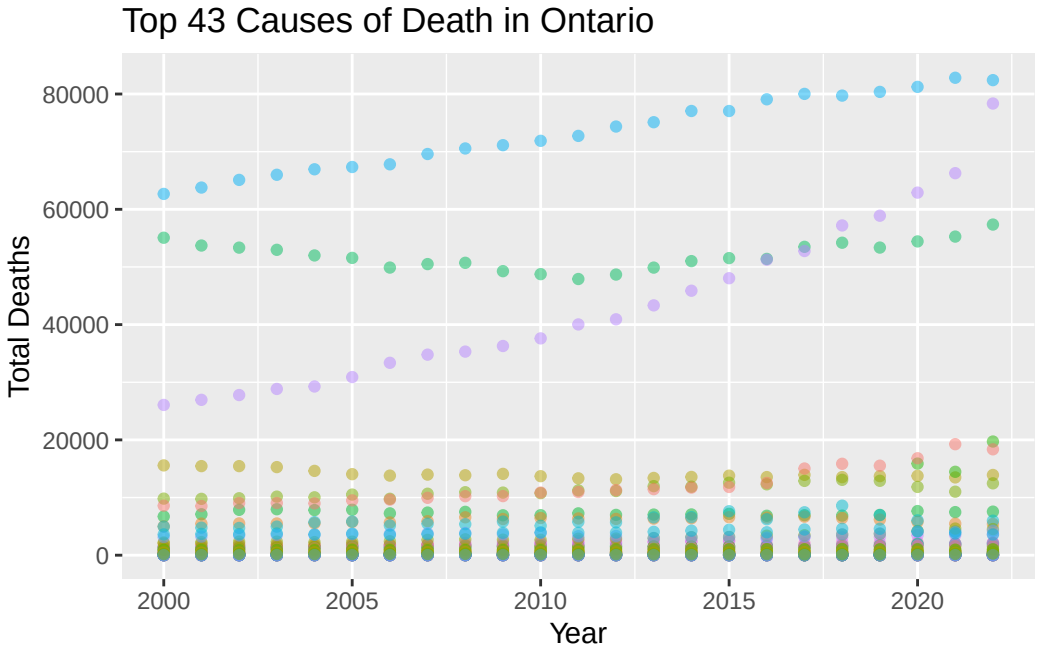

Figure 1: Visual of each value in the dataset, the number of deaths per year per cause. Each color corresponds to a cause

For this research we look at the data pertaining to both sexes and all age groups, which are the cumulative results from the individual information from male and female and from all the age groups combined. The purpose of this paper is not to define age specific or gender specific trends in the mortality of Ontario residents, but rather the overall trend of the leading causes of death and predictions for the future. In this way way we can use the combined data from these groups and still achieve meaningful results in our analysis **?@tbl-header-1**.

The original data set from STAT.CANADA.CA contains 407,334 observations of 18 different variables. In our analysis, using the combined data of gender and age, we look at 230 observa-

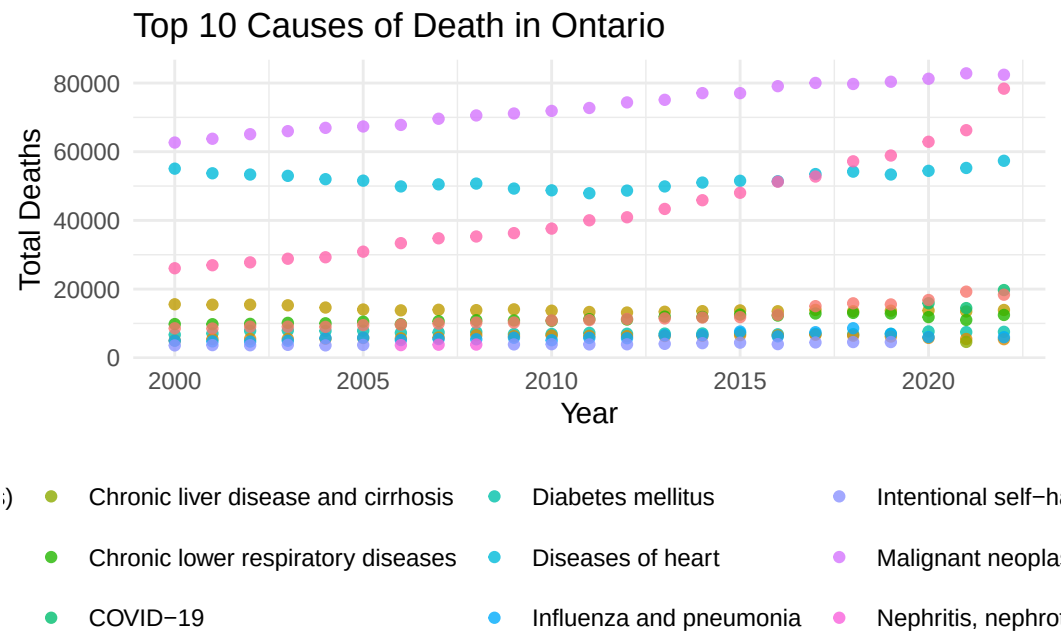

Figure 2: Visual of each value in the dataset, the number of deaths per year per cause. Each color corresponds to a cause

tions of 5 variables: year, cause, total deaths per cause, yearly rank of leading cause of death, and total deaths that year. There are 230 observations for the top ten causes of death over 23 years, 2000 to 2022. In the 23 years, there has only been 13 different causes ranked as the top ten leading causes of death **?@tbl-header-2**. The yearly deaths for each of the top ten causes range from 3,606 to 82,822. The annual deaths in Ontario from 2000 to 2022 range from 218,062 to 334081.

3 Model

The model used to simulate, analyse, and predict the next five years of mortality data for the top ten causes of death in Ontario was produced using the rstanarm package and the stan_glm function. Taking advice from [INSERT ROHAN TEXTBOOK REFERENCE OF CHAPTER 13 ALBERTA DATA] we set out to fit our data to a negative binomial distribution. [INSERT FORMULA FOR NEGATIVE BINOMIAL PMF]. We regressed total deaths for each cause against cause of death dependent on the year. We used the negative binomial distribution because our data has a variance far exceeding the mean. Table 4. We also specify a log link function because as we are dealing with number of deaths, any negative value as a result of the model will not be telling of the real world. In this way using the log link function the predictions are always positive.

Table 4: Summary statistics of the number of yearly deaths, by cause, in Ontario, Canada.

Min	Mean	Max	3	SD	Var	N
3606	23051.95	82822		23534.65	553879519	230

The model we built is very well fitting to a negative binomial distribution as shown in **?@tbl-model-summary**. We can also look at the residuals of the observed data compared to that which was predicted by our model. We can see this distribution is very normal looking with a clear mean of 0. The outcome of this model is discussed further in the discussion and results

Table 1: Visual of the original dataset and the top 13 leading causes of death from 2000 to 2022.

Year	Leading Cause of Death	Deaths	Rank	Total Annual Deaths
2000	Malignant neoplasms [C00-C97]	62672	1	218062
2000	Diabetes mellitus [E10-E14]	6714	7	218062
2000	Alzheimer's disease [G30]	5007	8	218062
2000	Diseases of heart [I00-I09, I11, I13, I20-I51]	55070	2	218062
2000	Cerebrovascular diseases [I60-I69]	15576	4	218062
2000	Influenza and pneumonia [J09-J18]	4966	9	218062

Leading Causes of Death	Fequency
Chronic liver disease and cirrhosis	1
COVID-19	3
Nephritis, nephrotic syndrome and nephrosis	3
Intentional self-harm (suicide)	17
Influenza and pneumonia	22
Accidents (unintentional injuries)	23
Alzheimer's disease	23
Cerebrovascular diseases	23
Chronic lower respiratory diseases	23
Diabetes mellitus	23
Diseases of heart	23
Malignant neoplasms	23
Other causes of death	23

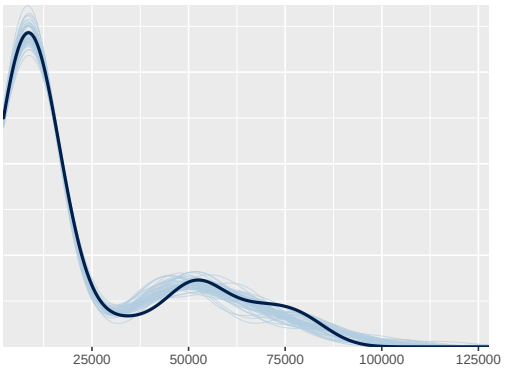

Figure 3: Modeling the most prevalent cause of deaths in Ontario, 2000 - 2022.

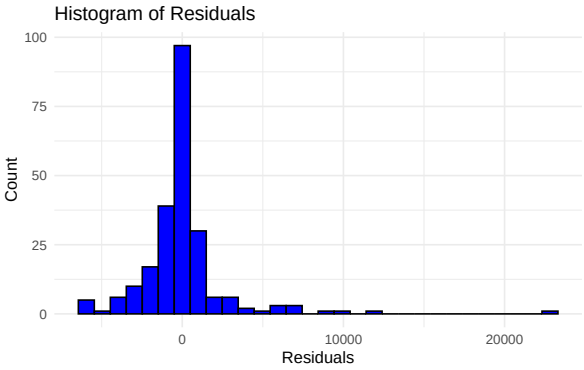

Figure 4: Modeling the most prevalent cause of deaths in Ontario, 2000 - 2022.

remained at rank number one Figure 5. As for the fourth through tenth leading causes, there has been many changes in rank however the amount of deaths for each of these causes does not 20,000, and rarely surpasses 15,000. In this way a small change in the amount of deaths per cause could lead to a change of rank much easier than in the case of the top three leading causes.

In the case of the fourth through tenth leading causes, there is one outlier that has been trending upwardly in recent years: accidents and unintentional injuries. We can see this as the orange points in Figure 6. In 2000 this was ranked sixth with only 8,589 deaths to rank four in 2021 with 19,257 deaths.

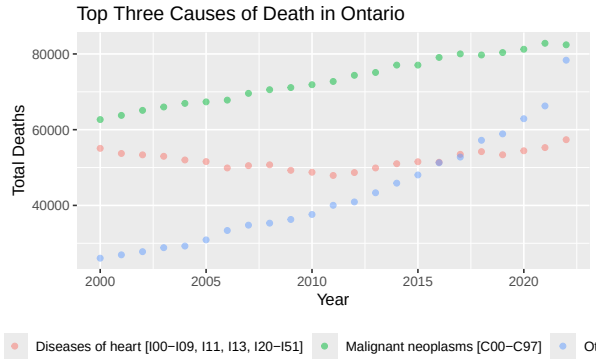

Figure 5: Visual of the trends in total deaths from the top three causes versus the next seven.

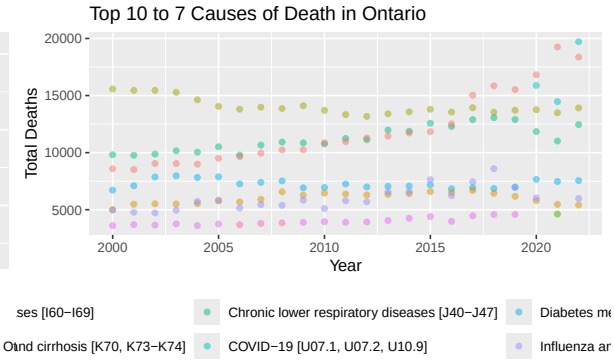

Figure 6: Visual of the trends in total deaths from the top three causes versus the next seven.

Using the model we built, having regressed cause and year together to predict the total number of deaths per cause each year, we can use the posterior predictors to predict value for the future. We did this for the next 5 years after 2022. We can see the visual extension of the predicted data with the observed data from 200 to 2022. In Figure 7 we can see the predicted values on the right.

5 Discussion

We can see in Figure 8 the total deaths in Ontario have increased greatly, and in recent years has had a somewhat exponential increase. Our model continues this trend and we can see the deaths continue to rise in Figure 9, however, it seems that our model is mistaken in that in 2023 it predicts less than 290,000 deaths whereas in 2022 we observed over 330,000 deaths. Further, when we look at Figure 7, we see again see a similar extension by the model in a linear fashion. It predicts the death for each cause to increase in the next years, which is most likely accurate, however in a completely linear way. Naturally this is reflected in the total deaths and we begin to learn of some of the flaws of this model.

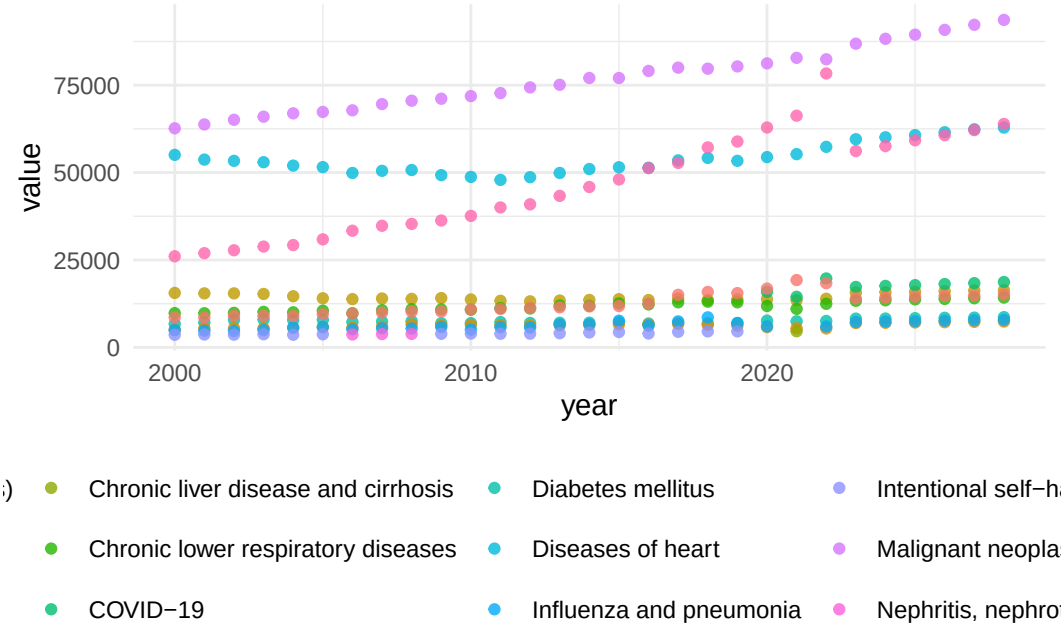

Figure 7: A visual extension of the observed data with predictions from the neagtive binomial model from 2023 to 2028.

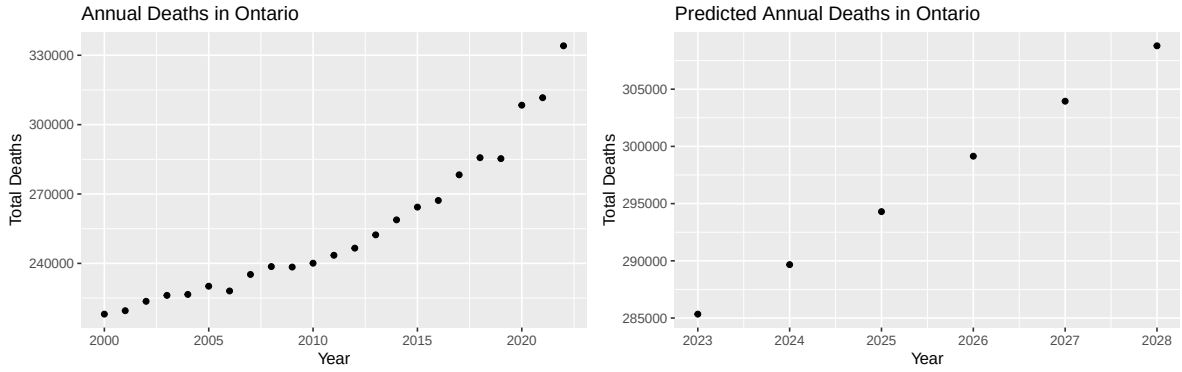

Figure 8: Visual of the total deaths in Ontairo from 2000 - 2022.

Figure 9: Visual of the total deaths in Ontairo from 2000 - 2022.

Having seen that model fits very well for the observed data and that in the future it predicts very linear changes, one may conclude that the real predictors are not year and cause of death but rather much more specific and less quantifiable factors. Year and cause fo death are very good predictors when we have the data to match, however we cannot get mortality data for the future and so the model can only replicate trends it has seen in the data from 2000 to 2022. It is replicating the almost linear gradual change in the values of our data we can see in Figure 2. What the model does nto pick up on well enough is that there is exponential growth of the total deaths per cause, and total deaths in Ontario over the recent past, more specifically since 2017.

6 Appendix

6.1 Additional data details

6.2 Model details

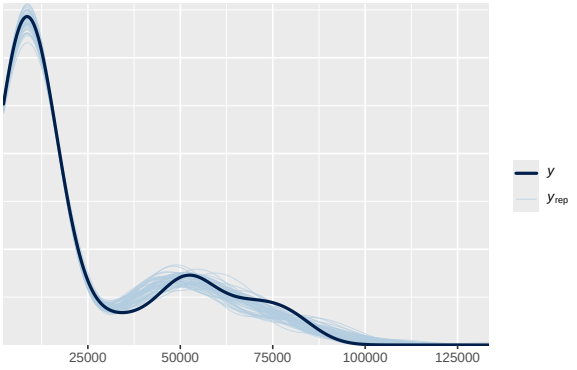

Figure 10: Residuals plotted displaying a normal-like distribution and graph of the posterior predection check.

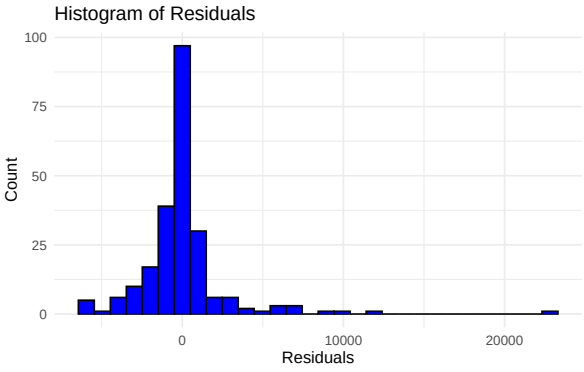

Figure 11: Residuals plotted displaying a normal-like distribution and graph of the posterior predection check.

6.3 Posterior predictive check

References

Alexander, Rohan. 2023. *Telling Stroies with Data*. https://tellingstorieswithdata.com/13-ijaglm.html#negative-binomial-regression.

Canada, atistics. 2023. *Mortality Rates, by Age Group*. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310071001.

Chang, Winston. 2015. *Downloader: Download Files over HTTP and HTTPS*. https://CRAN.R-project.org/package=downloader.

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

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

Wickham, Hadley. 2011. "Testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

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.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.