Demystify Demise: A Comprehensive Analysis of Top 6 Causes of Death in 2020, Led by Organic Dementia*

Zhijun Zhong & Xincheng Zhang

March 12, 2024

Mortality rates and annual death numbers are critical indicators of population dynamics, reflecting both growth and contraction of population. In this report, we conducted a complete investigation into the myriad causes of death in the year 2020, developing multiple models to elucidate the six leading causes of mortality detailed in this comprehensive report. Our findings indicate that Organic Dementia, characterized as a form of cognitive impairment, emerged as the predominant cause, accompanied by five other significant conditions that collectively delineate the mortality of 2020. Unveiling the escalation in mortality attributable to diverse factors, we propose strategic interventions aimed at mitigating these lethal influences. Our objective is to enhance overall well-being and elevate human happiness by preventing the detrimental impact of these fatal factors.

Table of contents

*Files, Codes and Manuscript are Availiable at . The original study which we replicated is availiable at https://tellingstorieswithdata.com/13-ijaglm.html#negative-binomial-regression

Introduction

Birth and death rates are generally used as two determining factors to judge population growth or decline. The definition of death is mainly judged by the extreme decline and cessation of heart or cardiopulmonary function (Sarbey 2016). The common causes of death are tumors, epidemic diseases, etc. (Mokdad et al. 2004) Our report provides a comprehensive exploration of the mortality pattern in Alberta by taking the top six different causes of death over the last decade. These diverse causes were led by Organic Dementia, which is a cognitive disorder, and were followed by five other causes, including COVID-19.

This report discussed the meaning of the variables in the Data section and, through narrative methods, introduced all the data sources used. A detailed explanation of each variable shown in the dataset is included with a table. The Model part attempts to build Poisson and Negative binomial models for the top 6 different causes of individual death in 2020 and fitted a model to predict the cause of death by these 6 factors. The changing trends in deaths from six different causes from 2001 to 2022 are clearly shown in different colors on the line graph in result section, which will give us a lot of insights in the Discussion section and make accurate predictions and effective suggestions for the future. We are conducting general analyses and closing existing gaps in our understanding of mortality using literature from a variety of sources. Rather than analyzing the details of each cause of death on a case-by-case basis, our paper prefers to combine these various factors into a cohesive observation and provide the reader with an overview through the narrative of the many perspective of mortality in the particular year we selected. There are also several reasonable suggestions are raised to pursue the improvement of human happiness.

After data collection and sketching drawing, we found that Organic Dementia plays a dominant role in increasing mortality among different causes. COVID-19, which has not been significantly reflected in other years, has also been included in the top six causes of death in 2020. We can use mortality in Alberta to grapple with unexpected causes of death in the 2020 mortality pattern and face pressing questions about social healthcare policy, public awareness of disease, which is the importance of this report. The findings in this report can also provided as a compass for the future, guiding Canadian residents to predict and investigate frequent and dynamic mortality in the future and focus on solving or weakening the six leading causes of death.

Data

Data Source And Methodology

All data involved in this report are obtained from the original data (Service Alberta 2023) by downloading the data package. The raw data URL is a collection of data that is stored and uploaded by certified specialists. The basic information of the data is related to social issues in Alberta, Canada, and is published and uploaded by Service Alberta. We selected a dataset called "Leading causes of death", which describes a ranking of the 30 most common causes of death each year in Alberta. This database has a high level of activity since it was last updated on September 22, 2023.

During the process of sorting the data, we found that the original data package contained one row of missing value and three rows of column "cause" marked as "(blank)". We use the Deletion method to deal with the row with missing value since the data missing type belongs to MCAR and the deletion of one row will not affect the overall results for 663 rows. For the three rows of data marked as "(blank)", which are 2014, 2015, and 2018 respectively. We use the Single Imputation method to try to replace blank value with a similar data. We looked forward to the causes of death corresponding to the rankings of the following year, that is, 2015, 2016, and 2019 respectively, and used them as substitutes for the missing data. This method can ensure that the gap with the actual results is narrowed while ensuring the sample size since the difference between the year and its following year is relatively small. We also created a new variable called "cause_grouped". This variable is to combine the ones who died due to all neoplasms types for each year as a one single row. This new data will be used in the Result section to analyze trends in various causes of death from 2001 to 2022.

In the original database website (Service Alberta 2023), the data set which has a purpose of revealing the cause of death can also be found on the website such as "Deaths, cause by gender and age" and "Infant deaths, cause by gender and age", which may be regarded as replaceable or similar dataset. However, we are focusing on a more general populations while the subjects used to investigate and form the data set are only focused on infants in Alberta, which can lead to biased results. Therefore, compared with the "Leading causes of death" dataset, there is a certain degree of incompatibility with our report with these two similar data sets.

Attributes

Table 1: Variable Description for raw data

Column	Type	Description
calendar year	num	Specific year in the range from 2001 to 2022
causes	str	Specific cause of death, such as Stroke
ranking	num	Organizes the causes of death by their severity or frequency within a given calendar year counted by total_death
total deaths	num	Represents the total number of deaths attributed to each cause in a specific calendar year

For the data set we use, there are 5 columns representing different variables, and the meaning of each variable is organized in Table 1.

The relationships between different variables in the "Leading causes of death" dataset are mainly presented in "ranking" and "total_deaths". Specifically, for each year from 2001 to 2022, the column ranking arranged cause of death from "1" being the top deadly to "30" being the least deadly. For "total_deaths", represents the number of deaths caused by each cause, which will be negatively related to the "ranking" variable. For example, in 2020, Organic dementia ranked on top in the "ranking" and has a whopping 2,081 deaths number, which is higher than any other cause of death in the same year.

All data manipulated and presented in this paper were sourced from the Alberta open government website (Service Alberta 2023). The original paper we trying to replicate is written by Professor Alexander Rohan on the course textbook website (Rohan 2023). The data processing, and analyzing for this replication is using R (R Core Team 2023) along with other support while very useful packges knitr(Xie 2023), janitor(Firke 2023), tidyverse(Wickham, Averick, et al. 2023), lubridate(Grolemund and Wickham 2023), ggplot2(Wickham 2023a), dplyr(Wickham, François, et al. 2023), readr(Wickham, Hester, and Bryan 2024), gridExtra(Auguie and Antonov 2017), modelsummary(Arel-Bundock 2022), broom.mixed(Bolker and Robinson 2022), stringr(Wickham 2023b), rstanarm(Brilleman SL and R 2024) .

Model

Table 2: Top 6 Cause of Death in 2020

cause	total_deaths
Organic dementia	2081
All other forms of chronic ...	1897
Malignant neoplasms of trac...	1563
Other ill-defined and unkno...	1464

Table 2: Top 6 Cause of Death in 2020

cause	total_deaths
Other chronic obstructive p...	1178
COVID-19, virus identified	1084

By studying the database used, We focus on the top 6 cause of death of specific year 2020 on Table 2. We plan to use two different model methods to depict the graph and analyze it separately, namely Poisson regression and Negative binomial regression when we conduct data analysis in this section. For Poisson regression, it emphasized modeling through the fluctuation of counting data and distributing probability among non-negative integers. The models of Poisson regression include many different types such as e-log-linear, quasilinear, etc. Predict the variable values and changes that need to be inferred through the relationship between the independent variables and the dependent variables shown in the model (Frome 1983). In this case, the entire graph will be spread over the top with variance and mean showing a positive correlation. For Negative binomial regression, the relationship between the variance and the mean mentioned above is changed and allows the graph to be relatively dispersed through variables.

Poisson Regression

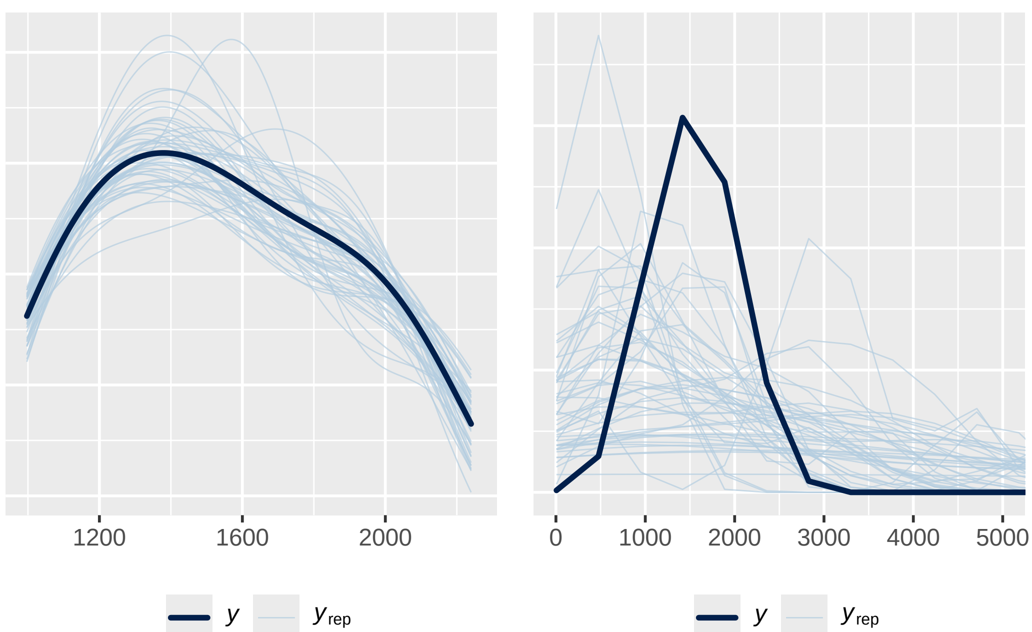

Figure 1: Comparison of Poisson Regression and Negative Binomial Regression

Table 3

	Poisson	Negative binomial
(Intercept)	7.547	7.581
		(0.780)
causeCOVID-19, virus identified	−0.560	−0.484
		(1.149)
causeMalignant neoplasms of trac...	−0.194	−0.125
		(1.125)
causeOrganic dementia	0.093	0.181
		(1.103)
causeOther chronic obstructive p...	−0.477	−0.406
		(1.095)
causeOther ill-defined and unkno...	−0.260	−0.187
		(1.098)
Num.Obs.	6	6
Log.Lik.	−28.782	−50.767
ELPD	−34.0	−54.5
ELPD s.e.	0.5	0.5
LOOIC	67.9	109.0
LOOIC s.e.	1.0	1.0
WAIC	64.8	107.3
RMSE	1.58	164.98

We organize the downloaded data package and establish a Poisson model on the left side of Figure 1. In the image, we first observe that the black line represents the observation data and is marked as "y", and the light blue relatively thin line surrounding the black line represents the observation data generated after establishing the Poisson model, which is replicated data about the distribution and is labeled "y_rep". Upon closer inspection of the graph, the trend of the main black line used to describe the image of the database is well described by the thin blue line acting as a support. Specifically, the lines of the replicated data and the main data have very similar curves and share a high overlap rate in the range of 1200-1600 deaths number on the x-axis. In contrast, there is some variability for the second half of the graph in the range 1600-2000 but the overall trend is still the same.

Table 4: Variable Description for equation

Variable	Description
a1 & b1	COVID19 virus
a2 & b2	Malignant neoplasms of trachea, bronchus and lung
a3 & b3	Organic dementia

Table 4: Variable Description for equation

Variable	Description
a4 & b4	Other chronic obstructive pulmonary disease
a5 & b5	other ill-defined and unknown causes of mortality

$$log_E(y_1) = 7.547 - 0.560a_1 - 0.194a_2 + 0.093a_3 - 0.477a_4 - 0.260a_5 \tag{1}$$

Through the above analysis of the Poisson model, the probability equation for the overall proportion of the top six causes of death in 2020 can be expressed Equation 1. The specific steps are to establish equations by observing various data observed in Table 3. For Poisson, we first have the Intercept value of 7.547 and by observing the subsequent data, we use a1/a2/a3/a4/a5 to represent each cause of death in order, which can be known through Table 4 and establish the initial formula related to log.

$$E(y_1) = e^{7.547-0.560a_1-0.194a_2+0.093a_3-0.477a_4-0.260a_5} \tag{2}$$

Then, the expected count of an outcome variable E(y1) is obtained by adding the index e. The specific steps are the replacement of the log on both sides of the Equation 1. The final formula for the mean value of death caused by different reasons in the Poisson model was obtained Equation 2.

In this formula, we can quickly predict conclusions based on the research we need by changing different values of a. For example, if we want to predict the number of deaths due to COVID-19 in 2020 by the Poisson model. We can set values other than a1 to 0 and set a1 to 1, and the entire formula becomes Equation 3.

$$E(y_1) = e^{7.547-0.560*1-0.194*0+0.093*0-0.477*0-0.260*0} \tag{3}$$

Negative Binomial Regression

For the Negative Binomial model building, we use the contents of the previous data packet and create the Negative Binomial Regression on the right side of Figure 1. In the graph, like the Poisson model, we first observed that the black line represents the observation data and is marked as “y”. The light blue relatively thin line around the black line represents the observation data generated after establishing the Negative Binomial Regression, which emphasizes distribution according to the replicated data and marked as “y rep”. In this very different case, the trend of the main black line used to describe the database graph is not fully described by the surrounding thin blue line. To be more specific, for the peak of the data

Table 5: LOO-CV Comparison of Model Predictive Performances

	elpd_diff	se_diff
cause_of_death_alberta_poisson	0.0	0.0
cause_of_death_alberta_neg_binomial	-20.5	0.3

trend, the lines for the replicated data and the previous data describe two completely different situations. When the x-axis is 500 and 2700, which means that the number of deaths is 500 and 2700, the replication data shows two obvious peaks. The previous data with a peak at 1400 does not show any sharp changes at 500 and 2700. Besides, the copied data constructed by Negative Binomial Regression is more unstable and non-centralized than the previous data graph, which means that it is incompatible with the previous data based on the disordered distribution of lines and points.

Through the above analysis of the Negative Binomial Regression, the probability equation for the overall proportion of the top six causes of death in 2020 can be also expressed Equation 4. The specific steps are to establish equations by observing various data observed in Table 3. For the Negative Binomial, we first have the Intercept value of 7.581 and by observing the subsequent data, we use b1/b2/b3/b4/b5 to represent each cause of death in order, which can be known through Table 4 and establish the initial formula related to log.

$$log_E(y_2) = 7.581 - 0.484b_1 - 0.125b_2 + 0.181b_3 - 0.406b_4 - 0.187b_5 \tag{4}$$

Similarly, the expected count of an outcome variable E(y2) is obtained by adding the index e. The specific steps are the replacement of the log on both sides of the Equation 4. The final formula for the mean value of death caused by distinctive factors in the Negative Binomial Regression model was obtained Equation 5.

In this formula, we can quickly predict conclusions based on the research we need by changing different values of a. For example, if we want to predict the number of deaths due to COVID-19 in 2020 by the Negative Binomial Regression model. We can set values other than b1 to 0 and set b1 to 1, and the entire formula becomes Equation 6.

$$E(y_2) = e^{7.581 - 0.484b_1 - 0.125b_2 + 0.181b_3 - 0.406b_4 - 0.187b_5} \tag{5}$$

$$E(y_2) = e^{7.581 - 0.484*1 - 0.125*0 + 0.181*0 - 0.406*0 - 0.187*0} \tag{6}$$

Comparison

Based on the above description of the two models and the detailed description of the formulas derived from the two models respectively. The Poisson model is relatively more suitable for our analysis in this report. This is because when we observe the two graphs produced by Poisson Regression and Negative Binomial Regression in Table 3, the thin blue line on the left (Poisson) has a certain inconsistency when the x-axis is equal to 1800. But the overall trend fits the black line representing the original data better than the right side (Negative Binomial). This phenomenon occurs since the Poisson model better explains the overdispersion in this case.

In addition, based on the measurements comparing the two model approaches value, which can be found in Table 5. The Expected log pointwise predictive density difference, shown as "elpd_diff" represents the log pointwise predicted density difference between the two models. When this value is larger, the model has better sample predictability. In this case, we can see that poisson had a greater value than the negative binomial, indicating that Poisson is more accurate than Negative Binomial in predicting the data. As for "se_diff", which refers to the standard error of the difference in elpd values in the models. When this value is higher, it means that the error between prediction and reality is larger. By observing Table 5, we can also find that nagative binomial is a little bigger than poisson, which means that Poisson has a relatively small error.

These two reasons will convince us to actively choose the Poisson regression model instead of the Negative Binomial regression model in analyzing the case of 2020 about 6 different causes of death.

Results

For the Result part, we built a line graph Figure 2 using the original dataset named "Leading causes of death" to analyze the annual mortality numbers due to various causes in a certain range of years where the x-axis represents calendar year showing from 2001 to 2022 with the units of one year. The y-axis showed total deaths measured in population. By examining the key on the right side of the graph, we can identify the specific causes of death, with each color representing a distinct cause. In terms of the variable "Neoplasms", it is the newly created illnesses where "cause_grouped" mentioned in the Data section that we created by summarizing all neoplasm-related diseases in each year.

By observing the sketched graph, it is evident that throughout the specified time frame, diseases caused by "Neoplasms" consistently dominate as the top cause of death, significantly surpassing other factors. It peaked in 2016, with approximately 3,700 deaths due to neoplasm. Moreover, for organic dementia, the relatively high slope indicates that it has the fastest growing trend in the range from 2001 to 2022. We notice that in 2001, less than 500 people died because this cause. However, by year 2022, organic dementia has killed nearly 2,400 people.

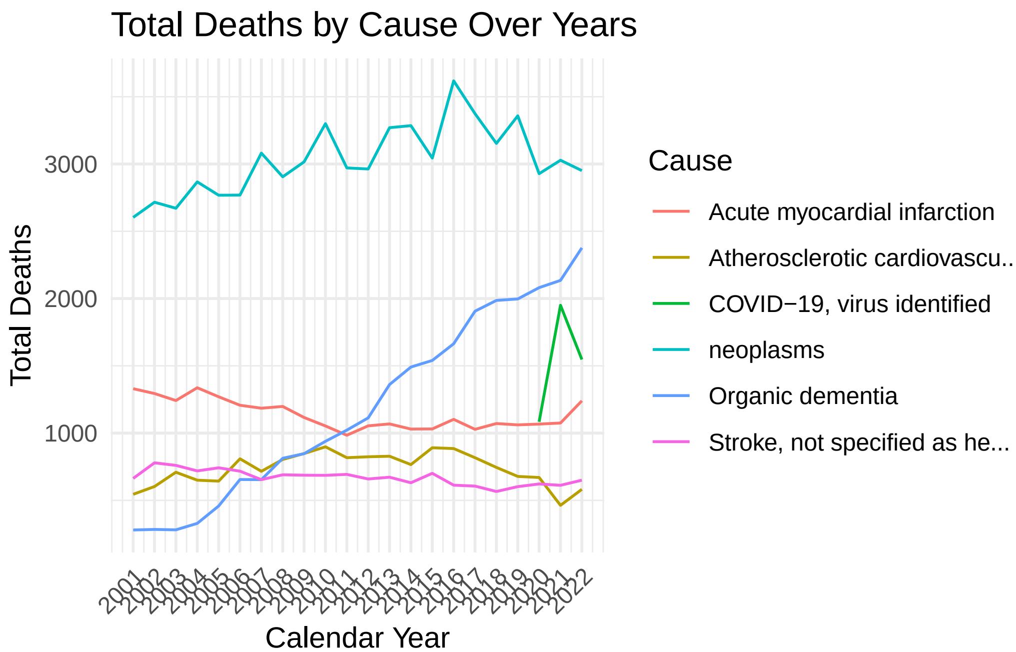

Figure 2

This means that an average of 100 more people will die because of this disease every year, with an average annual growth rate of 33.3%.

For disease like Acute myocardial infarction, Atherosclerotic cardiovascular disease and Stroke, there is no significant change within a given year range and they remain at a certain level of approximately 1200, 800 and 700 deaths per year respectively. But for COVID-19 which outbreak in 2020 meaning that all data points presented are only from 2020 onwards. In 2021, the number of deaths due to COVID-19 has increased by nearly 900 in one year, with an extremely high slope. The growth rate is 81%, which is the period with the most prominent slope in the entire graph.

Discussion

In this paper, we have explored 6 top causes in year 2020 and we try to fit a model so that we are able to predict the number of death for each illnesses. We found out that poisson regression is a better fit in our dataset than negative binomial regression. Thus, we can use Equation 2 to predict the total death of each of the top five cause of deaths. We also used a line graphFigure 2 to show a comparison of the number of deaths caused by six different causes from 2001 to 2022. Among them, we learned that the number of deaths caused by various neoplasms is far greater than that of other causes. For the remaining ones, organic dementia showed a significant growth rate within a restricted year range indicating an increase death rate where people need to be careful about. For Acute myocardial infarction, Atherosclerotic cardiovascular disease and Stroke, there is no significant change within a given year scope meaning that there is small discovery in treating these disease effectively. Due to COVID-19 rise in 2020, we expect the cause of death rate for it to raise. But we didn't expect it to grow in such fast way from 2020 to 2021.

Things Learned

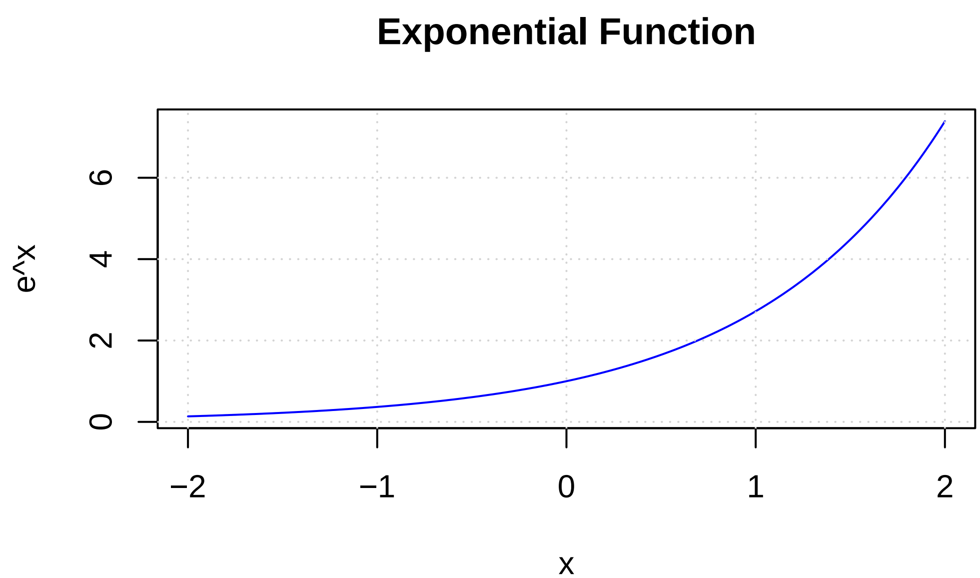

Figure 3

Through the analysis of the "Leading causes of death" dataset. We can use the mortality data in Alberta, Canada to derive a formula for predicting future causes of death by building a model and selecting the optimal model. According to Equation 2, we can infer the contribution of each of the six factors to the cause of death in a given 2020 by changing variables. The specific method is to set the "a_i" value (i is a positive integer from 1-5) represented by the cause of death that you want to explore to 1 and set the rest to 0. This method combined with the derivation formula can easily help us predict the annual number of deaths.

Also, combined with the increasing function graph Figure 3 of the exponential equation in mathematics, for E(y1), the larger the index, the higher the number of deaths. For Equation 2, COVID-19 represented by a1 has the lowest contribution to the number of deaths in 2020, which is equivalent to the smallest proportion of overall deaths. For this situation, when we know the total number of deaths and want to deduce the distribution of each cause of death, we can judge it by the coefficient before "a_i" (i is a positive integer 1-5). The smaller the coefficient, the greater the proportion of the corresponding cause of death, and vice versa.

In addition, by observing the Figure 2 in the result section, we can clearly observe the fluctuations in the factors responsible for most deaths in the past 20 years or so. This result can motivate us to prevent certain diseases. For example, for organic dementia, which has a rapid growth rate, we need to make some changes to reduce the growth rate of deaths caused by it as much as possible. This will provide a lot of inspiration and warning to relevant departments responsible for medical treatment and prevention.

Future Plan

For the graphs in the Result section, future planning will be divided into 4 parts.

First of all, Figure 2 in the Result section tells us that organic dementia, which is urgently needed to be curbed, is rising steadily at an annual growth rate of up to 33.3%. Organic dementia is a deterioration of mental functioning in its cognitive and emotional aspects specifically manifested as observable decline in mental abilities (Gustafson 1996). And because it is a disease that cannot be cured by itself or even treatments. So a more active prevention need to be setup. The main recommended methods for reducing this alarming growth rate are provided. Research shown that staying physically active (regular aerobic or strength training to strengthen your heart and lungs), limiting alcohol consumption (no alcohol of any kind), and eating a healthy, balanced diet (a balance of carbohydrates, protein, and fat) are the top priorities. These are three most effective ways to prevent organic dementia (Public Health Agency of Canada 2024). Therefore, we can enhance public awareness of the prevention of such diseases through various means, such as medical publicity or health lectures and other types of social activities. We believe that active prevention by raising citizens' awareness in this way can effectively and gradually reduce the growth rate of fatalities in the next few years.

The second point is that we need to weaken the neoplasmas that maintain a high lethality rate all year round. For all types of neoplasm, we can mainly divide it into two categories, Benign Neoplasm and Salivary Gland Neoplasm of Uncertain Malignant Potential (Baloch et al. 2023). For the first type of Benign Neoplasm, it can be performed through surgery because it does not have any transfer ability. It can be completely removed and allowed it to heal on its own. This can be solved by strengthening the success of Neoplasm surgery. Specifically, the government can strengthen the national medical equipment and academic improvement for Benign Neoplasm resection surgery through financial assistance. On the contrary, malignant neoplasm is proliferative and has great harm and malignant potential to life characteristics. Specific solutions require early detection, diagnosis and treatment, which requires the public to have sufficient knowledge for judging their own health condition. This can also be done by conducting academic lectures in various cities to popularize those knowledge about Neoplasm and make sure everyone with potential Neoplasm to be cured as soon as possible to reduce the number of deaths.

Thirdly, as for COVID-19, which emerged in 2020 in the figure, the death rate has increased by nearly 81% in just one year. By comparing all growth curves side by side, this part has a very steep slope. But what comforts us is that there is also a significant decline in 2022 compared with 2021. This phenomenon means that for the residents of Alberta, Canada, medical advancement and increased awareness of protection have profoundly reduced the number of deaths caused by COVID-19. Regarding this point in future planning, we are inclined to maintain the status quo, that is, to continue to gradually exclude this fatal factor through a steady downward trend. This goal also still requires every citizen to pay attention to this epidemic and raise their awareness of regulations.

Lastly, We noticed that among the three conditions mentioned above, environmental factors can be the main cause of aggravating people's health condition. When people face environmental problems, chemical pollutants is a primary cause of neoplasm (Trichopoulos, Li, and Hunter 1996). The accumulation of dust in the air will intensify the condition of COVID-19 (Marazziti et al. 2021). Food processing can also cause organic dementia due to pollutants. The governance environment is one of the issues that we as analysts recommend the government to deal with or pay attention to first. Specific methods can be to limit the total amount of combustion materials in chemical plants every day and to control it within the limits of air purification, also can be to avoid direct pollution by building special sewage pipelines when discharging pollutants. When ecological issues are taken seriously, we believe that the growth rate of many causes of disease will decrease and public happiness will improve.

Strength and Weaknesses

In the original data set of the "Leading causes of death" dataset", the strength is mainly reflected in the integrity of the data. Although we have effectively processed a total of 4 rows of missing data through Deletion and Single Imputation. But overall, it does not affect the performance of the research results in a relatively large data set. In addition, there is no inconsistency in the data types of the corresponding columns, which is beneficial to our efficiency in data sorting.

For flaws, the data set has regional biases in studying the causes of death each year. Specifically, the entire data set only contains ranking statistics for each cause of death in Alberta, Canada. This will lead to delayed treatment of the disease and an increase in deaths due to specific reasons such as insufficient regional medical facilities or citizens lacking awareness of prevention. For research values such as mortality or death volume that require a wide range of regions, a single region is not enough to explain the situation. Besides, if the time variable "calendar_year" has a more detailed division in the original dataset, for example, the ranking of different fatal factors into different months. We can then correlate mortality factors with months and even seasons to explore broader themes in this report.

References

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.

Auguie, Baptiste, and Anton Antonov. 2017. *gridExtra: Miscellaneous Functions for "Grid" Graphics*. https://CRAN.R-project.org/package=gridExtra.

Baloch, Zubair, Guido Fadda, Christopher C Griffith, Jerzy Klijanienko, Jeffrey F Krane, Lester Layfield, Zahra Maleki, Ritu Nayar, and Marc P Pusztaszeri. 2023. "Neoplasm." In *The Milan System for Reporting Salivary Gland Cytopathology*, 71–103. Springer.

Bolker, Ben, and David Robinson. 2022. *Broom.mixed: Tidying Methods for Mixed Models*. https://github.com/bbolker/broom.mixed.

Brilleman SL, Moreno-Betancur M, Crowther MJ, and Wolfe R. 2024. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. https://mc-stan.org/rstanarm.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Frome, Edward L. 1983. "The Analysis of Rates Using Poisson Regression Models." *Biometrics*, 665–74.

Grolemund, Garrett, and Hadley Wickham. 2023. *Lubridate: Make Dealing with Dates a Little Easier*. https://CRAN.R-project.org/package=lubridate.

Gustafson, L. 1996. "What Is Dementia?" *Acta Neurologica Scandinavica* 94: 22–24.

Marazziti, Donatella, Paolo Cianconi, Federico Mucci, Lara Foresi, Ilaria Chiarantini, and Alessandra Della Vecchia. 2021. "Climate Change, Environment Pollution, COVID-19 Pandemic and Mental Health." *Science of the Total Environment* 773: 145182.

Mokdad, Ali H, James S Marks, Donna F Stroup, and Julie L Gerberding. 2004. "Actual Causes of Death in the United States, 2000." *Jama* 291 (10): 1238–45.

Public Health Agency of Canada. 2024. "Dementia: Risk Factors and Prevention." https://www.canada.ca/en/public-health/services/diseases/dementia/risk-factors-prevention.html.

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

Rohan, Alexander. 2023. "Negative Binomial Regression." https://tellingstorieswithdata.com/13-ijaglm.html#negative-binomial-regression.

Sarbey, Ben. 2016. "Definitions of Death: Brain Death and What Matters in a Person." *Journal of Law and the Biosciences* 3 (3): 743–52.

Service Alberta. 2023. "Leading Causes of Death." Government of Alberta; https://open.alberta.ca/opendata/leading-causes-of-death; Service Alberta.

Trichopoulos, Dimitrios, Frederick P Li, and David J Hunter. 1996. "What Causes Cancer?" *Scientific American* 275 (3): 80–87.

Wickham, Hadley. 2023a. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

———. 2023b. *Stringr: Simple, Consistent Wrappers for Common String Operations*. https://stringr.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2023. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. https://readr.tidyverse.org.

Xie, Yihui. 2023. *Knitr: A Comprehensive Tool for Reproducible Research in R*. Edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.