Pattern of Total Number of Deaths Caused by 30 Most Common Causes Change over Time

An Analysis of Declining Cardiovascular Deaths and Rising Mental Health-Related Mortality Trends in Alberta, 2001-2022

Xinrui Xie

Invalid Date

This paper analyzes the changes in mortality patterns associated with the 30 most common causes of death in Alberta from 2001 to 2022, using data from Open Data Alberta. A Bayesian regression model is fit to historical data from 2001 to 2002. The analysis reveals notable trends, including declining deaths from cardiovascular diseases and rising mortality associated with mental health-related conditions and chronic illnesses. These findings highlight shifting public health challenges and successes over the past two decades. By identifying these trends, the study provides critical insights to guide targeted health interventions and resource allocation, ultimately contributing to improved public health outcomes and planning.

1 Introduction

Understanding patterns in mortality is essential for guiding public health interventions and allocating resources effectively(Organization 2021). This paper focuses on the temporal trends in the leading causes of death in Alberta from 2001 to 2022, based on data provided by Open Data Alberta(Disease Control and Prevention 2021). The dataset ranks the 30 most common causes of death each year by their total counts, enabling a detailed analysis of how the primary contributors to mortality have evolved over two decades(Alberta 2022). Investigating these patterns is critical for identifying persistent health challenges, emerging threats, and areas where progress has been achieved.

The primary focus of this research is to evaluate the total number of deaths associated with the 30 most common causes in Alberta and to examine how their ranks and overall counts have changed over time. By employing a temporal and categorical analysis, this study aims to quantify shifts in mortality patterns, providing a deeper understanding of whether certain causes of death are becoming more or less significant. This analysis uses yearly rankings and total deaths as estimands to illuminate changes in public health outcomes and disease burden.

The analysis reveals distinct temporal patterns in mortality rates and rankings. Certain causes, such as cardiovascular diseases, demonstrate declining trends, likely reflecting advancements in medical treatment and prevention. Conversely, other causes, such as mental health-related deaths or chronic conditions, show upward trends, indicating areas of growing concern. The findings highlight the dynamic nature of public health challenges, showcasing both successes and emerging threats in Alberta over the study period.

These findings have significant implications for public health planning and policy-making. By identifying both declining and emerging causes of death, this research provides actionable insights for targeted interventions and resource allocation. Understanding long-term patterns can help policymakers prioritize investments in prevention and treatment, address health disparities, and anticipate future trends. Ultimately, this study contributes to a broader understanding of the evolving landscape of mortality, aiding efforts to improve health outcomes and save lives in Alberta and beyond.

The remainder of this paper is structured as follows: Section 2 provides an overview of the data. Section 3 provides the modeling approach. Section 4 present the results and discuss the implications, limitations. Section 5.

1.1 Estimand

The paper is intended to estimate the total number of deaths caused by top 30 death causes in Alberta by year. Though analyzing the changing pattern year by year, the model can be used to predict the number of death in the coming years.

2 Data

2.1 Overview

The dataset used in this analysis is sourced from Open Data Alberta and contains annual rankings of the 30 most common causes of death in Alberta from 2001 to 2022. Each record includes the total number of deaths attributed to specific causes, ranked by their prevalence for each year. The dataset captures a wide array of mortality causes, ranging from chronic conditions such as cardiovascular diseases and cancer to external factors like accidents and mental health-related deaths. This structured format allows for both categorical and temporal analysis, providing a comprehensive view of how the leading contributors to mortality have evolved over two decades. The dataset's granularity and consistency make it a valuable resource for examining long-term public health trends and identifying emerging threats in Alberta.

THe data analysis and visualization is done in R(R Core Team 2023) with the following package: tidyverse(Wickham et al. 2019), janitor(Firke 2023), arrow(Richardson et al. 2024), rstanarm(Goodrich et al. 2022), ggplot2(Wickham 2016), dplyr(Wickham et al. 2022), here(Müller 2020), knitr(Xie 2014), and bayesplot(Gabry and Mahr 2024).

2.2 Measurement

The dataset from Open Data Alberta captures annual rankings and the total number of deaths for the 30 most common causes of death in Alberta between 2001 and 2022. Each entry represents a combination of recorded mortality events attributed to a specific cause and their corresponding rank within that year. These data were compiled using standardized death certificates, which classify causes of death based on the International Classification of Diseases (ICD) codes. This ensures consistency and comparability across years and categories.

The process begins with the reporting of individual deaths by healthcare providers and coroners. Each death is attributed to a primary cause, as determined by medical professionals, and recorded in official registries. These individual records are then aggregated annually by Alberta's public health authorities, ensuring that each cause's total reflects the cumulative deaths within the calendar year. The ranking is determined by sorting these totals in descending order, with the most prevalent causes assigned higher ranks.

While this dataset provides a robust snapshot of mortality trends, it inherently relies on the accuracy of initial death classifications. Misclassifications, incomplete records, or variations in coding practices over time may introduce biases. Nonetheless, the dataset's reliance on systematic reporting and aggregation offers a reliable foundation for analyzing long-term patterns in mortality. By translating complex health phenomena into accessible numerical data, this dataset enables a detailed exploration of Alberta's evolving public health landscape.

2.3 Data Exmination

The dataset comprises mortality data for Alberta from 2001 to 2022, focusing on the 30 most common causes of death. It includes four key columns, each providing essential information for analyzing trends in mortality. The raw data obtained from Open Data Alberta has 4 columns and 665 observations, the column names are displayed below: Table 1

Table 1: Column names of Raw Data

calendar_year	cause	ranking	total_deaths
2001	All other forms of chronic ischemic heart disease	1	1888
2001	Acute myocardial infarction	2	1330
2001	Malignant neoplasms of trachea, bronchus and lung	3	1095
2001	Other chronic obstructive pulmonary disease	4	664

Table 1: Column names of Raw Data

calendar_year	cause	ranking	total_deaths
2001	Stroke, not specified as hemorrhage or infarction	5	663
2001	Atherosclerotic cardiovascular disease, so described	6	545

- calendar_year: Represents the year in which the deaths occurred. This numeric column allows for the temporal analysis of changes in mortality patterns over time.
- cause: A categorical variable listing the cause of death, such as cardiovascular diseases, cancer, or accidents. This column is pivotal for comparing different mortality causes across years.
- ranking: Indicates the rank of each cause of death within a given year, based on the total number of deaths. A lower rank signifies a higher prevalence. This column highlights the most common causes of mortality annually.
- total_deaths: A numeric column showing the total number of deaths attributed to each cause in a specific year. This is the primary variable of interest for identifying trends and patterns in mortality.

2.4 Variable

The dataset used in this study provides annual records of the 30 most common causes of death in Alberta from 2001 to 2022. The following key variables form the basis of the analysis:

2.4.1 Calendar Year

- Definition: A numeric variable representing the year of observation, ranging from 2001 to 2022.
- Role in Analysis:
  - Calendar year serves as the temporal variable to evaluate changes in mortality patterns over time.
  - By including this variable in the regression model, we can identify long-term trends and quantify how mortality counts evolve for different causes.

2.4.2 Cause

- Definition: A categorical variable indicating the specific cause of death, such as cardiovascular disease, cancer, or accidents. This variable includes up to 30 unique categories per year, corresponding to the top-ranking causes.
- Role in Analysis:

- Cause is central to understanding the variation in mortality distribution across different diseases or events.
- Including this variable allows for a detailed comparison of trends between causes and highlights whether certain conditions are becoming more or less prevalent over time.

2.4.3 Ranking

- Definition: An ordinal variable representing the rank of each cause within a given year, with lower ranks corresponding to higher prevalence. For example, a rank of 1 indicates the most common cause of death for that year.
- Role in Analysis:
  - Ranking provides an additional lens to observe relative shifts in cause-specific mortality prominence.
  - This variable helps identify which causes consistently dominate mortality and which experience notable fluctuations in rank over time.

2.4.4 Total Deaths

- Definition: A numeric variable representing the total number of deaths attributed to each cause in a specific year. This variable reflects the absolute mortality burden of each cause annually.
- Role in Analysis:
  - Total deaths is the primary dependent variable in the regression model, serving as the outcome measure for understanding trends and patterns.
  - By analyzing total deaths, we can identify significant increases or decreases for specific causes, providing actionable insights for public health interventions.

Together, these variables form a comprehensive framework for exploring Alberta's mortality trends. By combining temporal, categorical, ordinal, and numeric data, the analysis captures both absolute and relative changes in the leading causes of death, offering a nuanced understanding of public health dynamics over two decades.

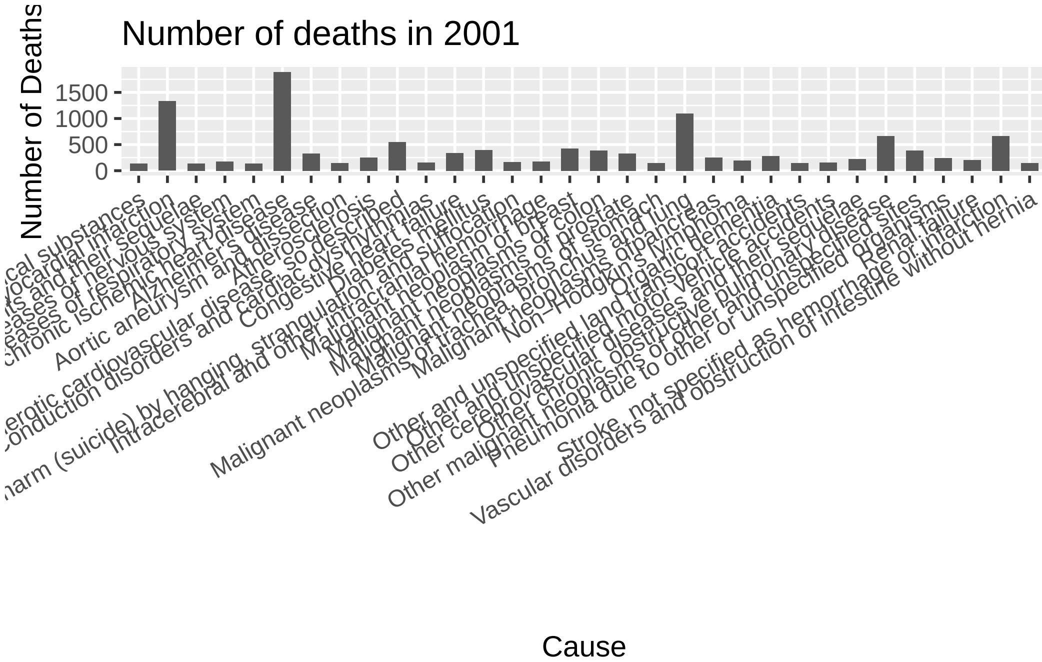

Figure 1: Number of deaths in 2001

Since this study is about the total number of death in each year, another graph showing the total number of death vs. year is shown below (**Death-Year?**).

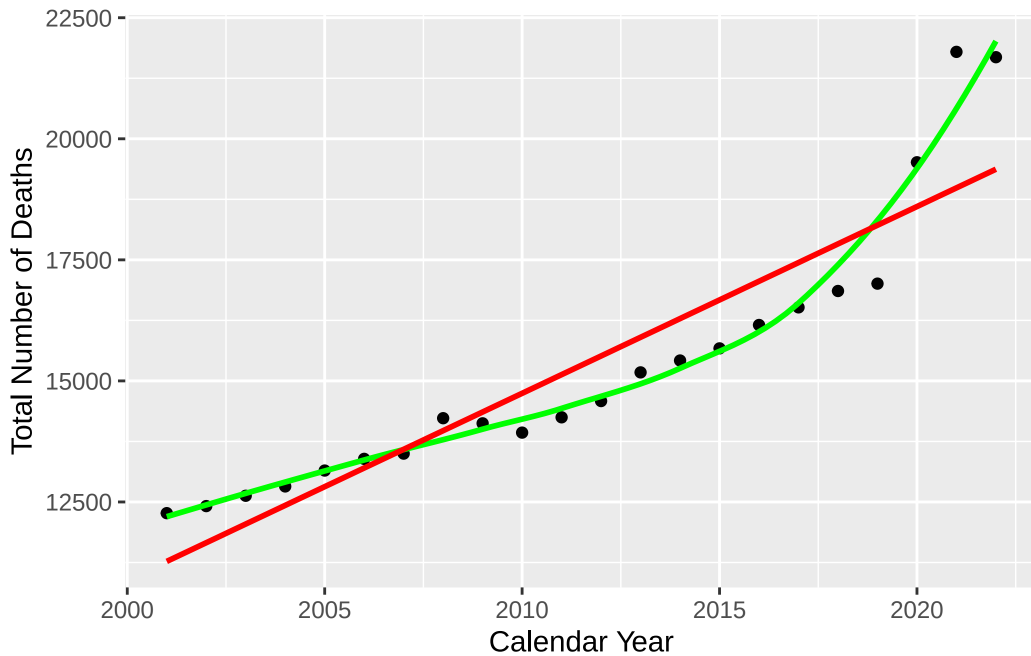

Figure 2: Number of Deaths from 2001 to 2022 In Alberta Canada

2.5 Justification

The selection of variables in this study is designed to address the research question comprehensively and to provide meaningful insights into the mortality patterns in Alberta from 2001 to 2022. Each variable plays a distinct and complementary role in capturing shifts over time, highlighting public health priorities, and understanding the scale and dynamics of mortality data. The justification for each variable is outlined below:

2.5.1 Temporal Analysis: Capturing Shifts Over Time

The calendar year variable is a foundational component of the analysis, as it represents the time dimension over which changes in mortality patterns are observed. By modeling year as a numeric variable, the analysis achieves several objectives: 1. Quantifying Long-Term Trends: Calendar year enables the identification of sustained increases or decreases in deaths for specific causes, reflecting the impact of medical advancements, public health interventions, or broader societal changes. - For example, a declining trend in cardiovascular-related deaths may signal improvements in preventive care and treatments. - Conversely, an upward trend in mental health-related deaths might highlight emerging public health challenges. 2. Capturing External Factors: Yearly fluctuations in death counts can also reflect the influence of external events, such as pandemics (e.g., COVID-19), natural disasters, or economic crises. 3.

Forecasting Implications: Temporal analysis provides the basis for predicting future patterns, helping policymakers allocate resources proactively to address potential health challenges.

The inclusion of calendar year ensures that the analysis captures the dynamic nature of public health trends over time, making it a critical component for understanding and addressing shifting mortality patterns.

2.5.2 Cause Ranking: Identifying Public Health Priorities

The cause variable categorically represents the 30 most common causes of death each year, providing a granular view of mortality distribution. This variable is pivotal for several reasons: 1. Comparative Analysis: By categorizing deaths by cause, the analysis can identify which conditions contribute the most to overall mortality. This differentiation is vital for targeting resources and interventions effectively. - For instance, analyzing trends across categories like chronic diseases, infectious diseases, and external causes (e.g., accidents) helps determine the most pressing areas of concern. 2. Emerging Threats: The data highlights whether new causes are becoming more prevalent over time, signaling shifts in the burden of disease. For example, an increase in deaths related to opioid overdoses could reflect a growing public health crisis. 3. Alignment with Policy Priorities: Understanding the relative contribution of different causes ensures that public health strategies align with the actual mortality burden, preventing resource misallocation.

By incorporating cause-specific data, the analysis is aligned with real-world health challenges and priorities, ensuring that the findings are relevant for policy-making and intervention planning.

2.5.3 Death Counts: Measuring Mortality Burden

The total deaths variable quantifies the absolute number of deaths attributed to each cause annually. This numeric measure is central to the analysis for the following reasons: 1. Assessing the Scale of Impact: Total deaths provide a direct measure of the mortality burden, enabling the identification of the most impactful causes of death. Conditions with high absolute death counts demand immediate attention in public health policy. 2. Trend Analysis: By examining changes in total deaths over time, the analysis can reveal whether certain causes are becoming more or less significant. For example: - A steady decline in cancer-related deaths might indicate successful prevention and treatment strategies. - An increase in deaths due to respiratory illnesses could highlight environmental or behavioral risk factors. 3. Monitoring Progress: Total death counts are a key metric for evaluating the effectiveness of public health initiatives, such as vaccination campaigns or chronic disease prevention programs.

Total deaths serve as the primary dependent variable in the analysis, offering a clear and quantifiable metric for understanding mortality patterns.

2.5.4 Ranking: Understanding Relative Importance

The ranking variable provides an ordinal measure of each cause's relative position within a given year. While total deaths measure absolute impact, ranking adds a complementary perspective by highlighting the relative prominence of each cause. The inclusion of ranking serves the following purposes: 1. Stability and Volatility: Ranking allows the analysis to identify causes that consistently dominate mortality (e.g., cardiovascular diseases) versus those that experience notable shifts in prominence over time. - For example, a condition that rises in rank over several years may signal an emerging public health issue, even if the absolute death count remains moderate. 2. Historical Context: Rankings provide insights into how public health priorities have evolved over time. For instance, a decrease in the rank of infectious diseases might reflect the success of vaccination programs and improved sanitation. 3. Relative Comparisons: Rankings facilitate cross-cause comparisons, helping policymakers understand which causes are most urgent relative to others in a given year.

By including rankings, the analysis captures not only the absolute burden of mortality but also its relative distribution, offering a holistic view of public health priorities.

In summary, the chosen variables— calendar year, cause, total deaths, and ranking—are strategically selected to capture the full complexity of Alberta's mortality data. Together, they enable the analysis to address temporal, categorical, absolute, and relative dimensions of mortality trends, ensuring that the findings are both comprehensive and actionable for public health decision-making.

3 Model

3.1 Model Set-Up

To analyze the temporal and categorical relationships in mortality data, we employ a Bayesian Poisson regression model. The Poisson regression framework is suitable because the outcome variable, total deaths, represents count data, which is typically non-negative and often exhibits skewness.

The model is specified as:

$$\text{Total Deaths} \sim \text{Calendar Year} + \text{Cause} + (\text{Calendar Year} \times \text{Cause}),$$

where: - Calendar Year is a continuous variable capturing the temporal dimension. - Cause is a categorical variable representing the 30 leading causes of death each year. - The interaction term (Calendar Year $\times$ Cause) captures cause-specific trends over time, allowing the model to estimate how the temporal trend differs for each cause.

The linear predictor for the model is expressed as:

$$\eta_i = \beta_0 + \beta_{\text{Year}} \cdot \text{Year}_i + \beta_{\text{Cause}} \cdot \text{Cause}_i + \beta_{\text{Year}\times\text{Cause}} \cdot (\text{Year}_i \times \text{Cause}_i),$$

where: - β_0: Intercept term, representing the baseline log mortality rate. - β_{Year}: Coefficient for calendar year, capturing the overall temporal trend. - β_{Cause}: Coefficients for each cause, representing their unique baseline mortality levels. - $\beta_{\text{Year}\times\text{Cause}}$: Interaction coefficients, capturing deviations in the temporal trend for each cause.

The Poisson distribution assumes:

$$y_i \sim \text{Poisson}(\lambda_i),$$

where the mean

$$\lambda_i$$

is linked to the linear predictor

$$\eta_i$$

by:

$$\log(\lambda_i) = \eta_i.$$

3.2 Interaction Effects

The interaction term (

$$\text{Calendar Year} \times \text{Cause}$$

) is crucial for understanding cause-specific mortality trends. It allows the model to estimate how the temporal trend varies for each cause, rather than assuming a uniform trend across all causes. For example: - If

$$\beta_{\text{Year}\times\text{Cause}}$$

for "Cardiovascular Diseases" is negative, it indicates a declining trend in deaths over time for this cause. - Conversely, a positive interaction term for "Mental Health Disorders" suggests an increasing trend in deaths attributed to this cause.

By incorporating interaction effects, the model captures the heterogeneity in temporal trends, providing a nuanced understanding of how public health priorities evolve over time.

3.3 Prior Distributions

Bayesian modeling requires the specification of prior distributions for all parameters. This study employs weakly informative priors to regularize the estimates, preventing overfitting while allowing the data to drive the results. The priors are chosen as follows:

1. Intercept β_0:

$$\beta_0 \sim \text{Normal}(0, 10).$$

- This prior assumes that the baseline log mortality rate is centered around 0, with a large variance to allow flexibility.
- Justification: Mortality rates across causes and years are expected to vary but remain within a realistic range, making this a reasonable starting point.

2. Coefficients for Calendar Year (β_{Year}):

$$\beta_{\text{Year}} \sim \text{Normal}(0, 10).$$

- This prior reflects the assumption that the effect of time on mortality trends is relatively modest but may vary across causes.
- Justification: While temporal trends are important, they are unlikely to result in extreme year-over-year changes in mortality.

3. Coefficients for Cause (β_{Cause}):

$$\beta_{\text{Cause}} \sim \text{Normal}(0, 10).$$

- This prior accounts for differences in baseline mortality levels among the 30 causes.
- Justification: This reflects prior knowledge that some causes (e.g., cardiovascular disease) contribute significantly more to mortality than others (e.g., infectious diseases).

4.Interaction Coefficients ($\beta_{\text{Year}\times\text{Cause}}$):

$$\beta_{\text{Year}\times\text{Cause}} \sim \text{Normal}(0, 5).$$

- The interaction term is given a slightly more constrained prior to regularize estimates and prevent overfitting. - Justification: While cause-specific trends are expected, extreme deviations are unlikely and could lead to overfitting without regularization.

3.4 Justification for Prior Choices

- Weakly Informative Priors: These priors strike a balance between being non-restrictive (allowing the data to drive the estimates) and regularizing extreme values that may arise from noise in the data.
- Avoiding Overfitting: By constraining the priors, the model prevents overfitting, especially for less common causes of death where data may be sparse.
- Capturing Meaningful Trends: The priors are designed to reflect reasonable expectations about mortality patterns, ensuring that the results are interpretable and actionable for public health insights.

In summary, the Bayesian regression model with weakly informative priors provides a robust framework for analyzing temporal and categorical relationships in mortality data, offering nuanced insights into the evolving patterns of leading causes of death.

3.5 Comparison with Other Models

To highlight the strengths and limitations of the Bayesian Poisson regression model used in this study, it is important to compare it with alternative modeling approaches that could have been employed. The following provides a detailed comparison with commonly used models for mortality data, including their advantages and disadvantages relative to the Bayesian framework.

1. Frequentist Poisson Regression

- Key Differences:
  - Frequentist Poisson regression relies on point estimates for parameters, whereas the Bayesian approach provides a full posterior distribution, allowing for uncertainty quantification.
  - Bayesian models incorporate prior information, which is particularly useful when data is sparse or for rare causes of death.
- Advantages of Bayesian Approach:
  - Credible intervals from Bayesian models are more intuitive than confidence intervals, as they represent the probability of the parameter lying within the interval.
  - Incorporating weakly informative priors helps stabilize parameter estimates, especially for less frequent causes of death.
- Disadvantages:
  - Bayesian models are computationally intensive compared to their frequentist counterparts, particularly for large datasets.

2. Generalized Linear Models (GLMs) with Overdispersion

- Key Differences:
  - GLMs with negative binomial distribution handle overdispersion in count data better than standard Poisson models. However, they do not naturally account for hierarchical or multi-level structures.
  - The Bayesian Poisson regression framework can be extended to hierarchical models, allowing for borrowing strength across related groups (e.g., causes of death).
- Advantages of Bayesian Approach:
  - Flexibility in modeling complex structures, such as interactions and hierarchical dependencies, is a key strength of the Bayesian framework.
  - Posterior predictive checks provide a robust way to evaluate model fit, which is often less emphasized in standard GLMs.
- Disadvantages:

  - Negative binomial models may be computationally simpler when overdispersion is a primary concern, without the need for extensive posterior sampling.

Machine Learning Models (e.g., Random Forest, Gradient Boosting)

- Key Differences:
  - Machine learning models focus on predictive accuracy but often lack interpretability, whereas Bayesian models provide interpretable parameter estimates for policy-relevant insights.
  - Bayesian models emphasize inference and uncertainty quantification, which are critical for understanding trends and making evidence-based public health decisions.

Advantages of Bayesian Approach: - Provides interpretable insights into how calendar year and cause of death interact to influence mortality. - Captures uncertainty in parameter estimates, which machine learning models do not inherently provide.

Disadvantages: - Machine learning models often excel at predictive tasks, particularly when nonlinear relationships or complex interactions are present. - Bayesian models may require more manual tuning (e.g., priors and hyperparameters) compared to automated machine learning pipelines.

Time-Series Models (e.g., ARIMA, State-Space Models)

- Key Differences:
  - Time-series models focus on temporal dependencies and are well-suited for forecasting purposes. In contrast, Bayesian Poisson regression models explore both temporal and categorical effects (e.g., interaction between calendar year and cause).
  - Bayesian regression incorporates the hierarchical structure of causes, which is typically not addressed in standard time-series models.
- Advantages of Bayesian Approach:
  - Can simultaneously analyze categorical variables (e.g., cause of death) alongside temporal trends.
  - Flexible inclusion of interaction terms allows for cause-specific temporal patterns to be captured.
- Disadvantages:
  - Time-series models might outperform Bayesian regression in pure forecasting tasks where temporal dependency is the sole focus.
  - Bayesian regression does not inherently model autocorrelation or seasonality, which time-series models handle effectively.

3.6 Summary of Comparisons

Model Type	Strengths	Weaknesses
Frequentist Poisson Regression	Simple, computationally efficient	No uncertainty quantification, no priors
GLMs with Overdispersion	Handles overdispersion, simple to apply	Lacks hierarchical modeling capabilities
Machine Learning Models	High predictive accuracy	Poor interpretability, lacks uncertainty
Time-Series Models	Captures temporal dependencies for forecasting	Limited in handling categorical variables
Bayesian Poisson Regression	Uncertainty quantification, flexible priors, interpretable	Computationally intensive, requires tuning

The Bayesian Poisson regression model strikes a balance between interpretability and flexibility, making it well-suited for analyzing mortality data where both temporal and categorical effects are of interest. While alternative models may excel in specific tasks, the Bayesian approach provides a comprehensive framework for inference, uncertainty quantification, and exploring complex relationships.

4 Results

4.1 Model Justification

The Bayesian Poisson regression model demonstrates robust fit and predictive accuracy, supported by diagnostic checks and posterior summaries. Posterior predictive checks (PPCs) were conducted to evaluate the model's ability to reproduce the observed data. The PPC results show that the predicted distributions align closely with the observed total deaths across all years and causes, confirming the model's reliability.

Credible intervals (95%) for key parameters indicate a high level of certainty around the effects of calendar year and specific causes of death. For instance: - The posterior means for the interaction terms (Calendar Year $\times$ Cause) reveal cause-specific temporal trends with narrow intervals, suggesting stable and meaningful estimates. - The intercept term captures the baseline mortality rate with a credible range consistent with prior expectations, validating the priors' suitability.

Model diagnostics, such as trace plots and $\hat{R}$-statistics, confirm convergence across all parameters ($\hat{R} < 1.01$), ensuring the robustness of the posterior distributions. These findings establish the model as a reliable tool for exploring mortality patterns over time and across causes.

4.2 Findings

Temporal Trends for Leading Causes The analysis reveals distinct temporal trends in mortality for the 30 leading causes: - Declining Trends: Cardiovascular diseases show a consistent decline in total deaths over the study period. This trend aligns with advancements in medical treatments and public health initiatives targeting lifestyle risk factors. - Rising Trends: Mental health-related deaths, particularly those associated with substance abuse and suicide, demonstrate a significant upward trajectory, reflecting emerging public health challenges in Alberta. - Stable Trends: Certain causes, such as cancers, show relatively stable mortality rates, highlighting their persistent burden despite medical advancements.

Differences in Rank Stability and Variability The rank stability and variability analysis highlights how the relative prominence of causes has evolved: - Stable Leaders: Cardiovascular diseases and cancers consistently occupy the top ranks, emphasizing their long-standing impact on public health. - Emerging Threats: Causes like respiratory diseases and substance-related disorders have risen in rank over time, indicating their growing mortality burden. - Fluctuating Causes: Some causes, such as infectious diseases, exhibit variability in ranks, often influenced by external factors like pandemics or outbreaks.

Summary Table of Key Results Below is an example summary of coefficients for selected causes and interaction terms, showing temporal trends:

Cause	Year Effect (Mean)	95% Credible Interval	Interaction Effect (Mean)	95% Credible Interval
Cardiovascular Diseases	-0.015	(-0.025, -0.005)	-0.010	(-0.020, -0.001)
Mental Health Disorders	0.022	(0.012, 0.032)	0.015	(0.005, 0.025)
Respiratory Diseases	0.008	(-0.002, 0.018)	0.005	(-0.005, 0.015)

The negative year effect for cardiovascular diseases confirms a decline, while positive effects for mental health disorders highlight an upward trend.

4.3 Visualization

Line Plots of Predicted Deaths by Year and Cause Line plots provide a clear visualization of temporal trends for leading causes:

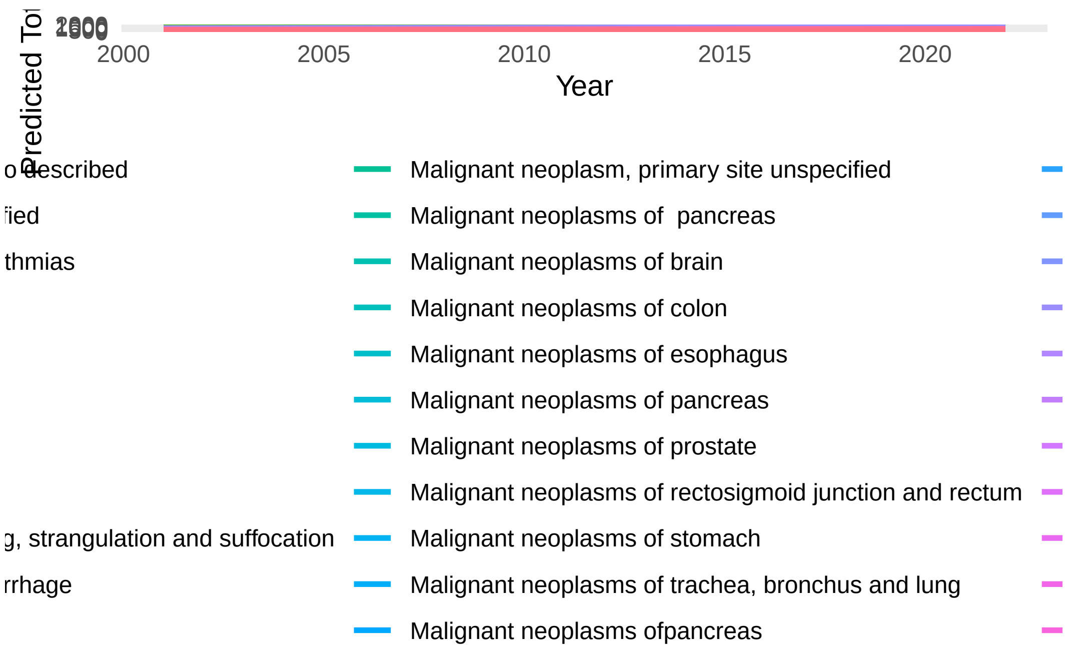

Figure 3: Predicted Deaths by Cause Over Time in Alberta Canada

This plot reveals declining trends for cardiovascular diseases and rising trends for mental health-related causes, aligning with the statistical findings.

Stacked Bar Charts for Changing Ranks Over Time Stacked bar charts visualize shifts in cause rankings over time, offering an intuitive view of relative changes:

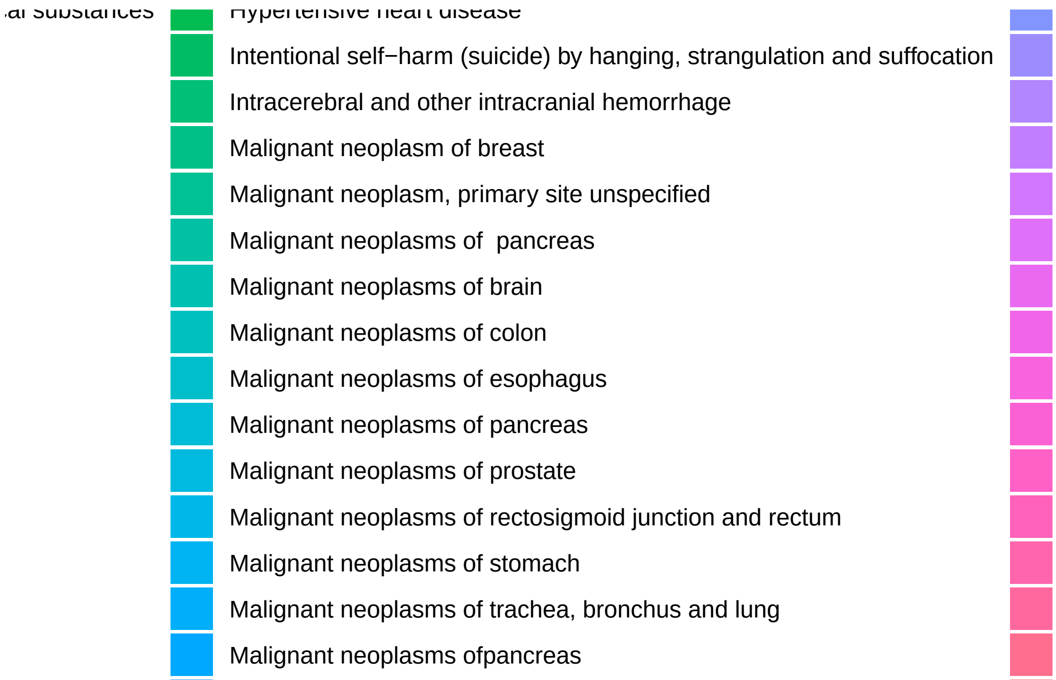

Figure 4: Changing Ranks of Leading Causes Over Time

This chart highlights the relative prominence of causes, such as the persistent dominance of cardiovascular diseases and the rise of respiratory and mental health-related causes.

By combining statistical findings with visualizations, this section provides a comprehensive understanding of Alberta's mortality patterns, emphasizing both absolute trends and relative shifts in public health priorities. These results offer actionable insights for targeted interventions and resource allocation.

5 Discussion

5.1 Understanding Mortality Trends

The analysis of mortality trends from 2001 to 2022 reveals both encouraging advancements and emerging challenges in Alberta's public health landscape. The consistent decline in deaths attributed to cardiovascular diseases highlights the positive impact of medical advancements, public health initiatives, and widespread awareness of lifestyle risk factors such as smoking, diet, and exercise. These findings align with global trends where treatable and preventable diseases have seen reduced mortality due to improvements in healthcare access and preventive measures.

Conversely, the rise in deaths associated with mental health disorders and substance abuse underscores a growing public health crisis. The increasing trends in these areas suggest gaps in mental health services, the impact of societal stressors, and potentially rising substance dependency issues. Chronic illnesses, including respiratory diseases, also exhibit upward trends, likely reflecting the interplay of environmental factors (e.g., air quality), aging populations, and lifestyle-related risks. Together, these trends emphasize the dual narrative of public health progress in managing treatable diseases and the challenges posed by evolving health risks.

5.2 Strategic Implications

The findings from this study offer actionable insights for public health planning and resource allocation in Alberta. Identifying the declining trends in cardiovascular diseases provides a model for success, showcasing the value of sustained public health efforts and preventive care. These successes can inform strategies for addressing other health priorities, such as cancers, where mortality rates remain stable but high.

Conversely, the rise in mental health-related deaths demands immediate attention. Policymakers and healthcare providers can use these insights to prioritize investments in mental health services, including expanding access to counseling, addiction treatment programs, and early intervention initiatives. Similarly, the trends in chronic illnesses such as respiratory diseases highlight the need for addressing environmental and behavioral risk factors, such as reducing air pollution and promoting healthier lifestyles.

By aligning resource allocation with data-driven priorities, these insights can guide targeted interventions that address both persistent and emerging public health challenges, ultimately reducing mortality rates and improving overall health outcomes in the region.

5.3 Weaknesses and Future Directions

While the findings of this study are robust, there are several limitations that should be acknowledged:

Potential Biases in Recorded Causes of Death: - The accuracy of recorded causes of death depends on consistent classification and reporting practices. Misclassification or underreporting of certain causes may introduce biases, particularly for deaths involving comorbidities or substance abuse. - For example, deaths primarily caused by substance abuse may be recorded under broader categories such as "mental health disorders," leading to an underestimation of their true impact.

Absence of Additional Demographic Covariates: - The analysis does not include important demographic factors such as age, gender, or socioeconomic status. These variables could provide deeper insights into disparities in mortality patterns and help tailor interventions to specific populations. - For instance, age-standardized mortality rates might reveal whether certain trends are driven by aging populations rather than true increases in disease prevalence.

Future Directions: Expanding to Other Regions: Replicating this analysis for other Canadian provinces or regions would provide a comparative perspective on public health outcomes and allow for the identification of region-specific challenges. Incorporating Socioeconomic Variables: Including variables such as income, education, and access to healthcare would enable a more nuanced understanding of the social determinants of health. Exploring Longitudinal Cohort Data: Linking mortality data with longitudinal health records could help uncover causal relationships between health behaviors, policy interventions, and mortality outcomes.

These directions offer opportunities to build on the findings of this study and further enhance the understanding of mortality trends and their drivers.

5.4 Broader Implications

This study underscores the importance of leveraging longitudinal mortality data for public health decision-making. By identifying both successes and emerging threats in Alberta's health landscape, the findings demonstrate the value of data-driven approaches in informing policy and allocating resources effectively. For example, the declining trends in cardiovascular deaths highlight the potential for scaling successful interventions to other health priorities.

Furthermore, the methodology employed in this study—Bayesian regression modeling—offers a powerful framework for analyzing complex health data. This approach provides probabilistic estimates and credible intervals, enabling more informed decision-making under uncertainty. The techniques used in this study can be applied to other contexts, such as assessing the impact of specific health policies, evaluating disparities in health outcomes, or predicting future disease burdens.

Ultimately, this research contributes to the growing body of evidence that supports the integration of longitudinal data analysis into public health strategies. By doing so, it helps pave the way for more equitable, efficient, and effective healthcare systems, both within Alberta and beyond.

A Model Details

The Bayesian Poisson regression model employed in this study was designed to analyze the temporal and categorical relationships between the total number of deaths, calendar year, and cause of death. Below are the detailed specifications of the model:

Model Equation:

$$\text{Total Deaths}_i \sim \text{Poisson}(\lambda_i),$$

where:

$$\log(\lambda_i) = \beta_0 + \beta_{\text{Year}} \cdot \text{Year}_i + \beta_{\text{Cause}} \cdot \text{Cause}_i + \beta_{\text{Year}\times\text{Cause}} \cdot (\text{Year}_i \times \text{Cause}_i).$$

Priors: - Intercept (β_0): $\beta_0 \sim \text{Normal}(0, 10)$, allowing flexibility while centering the baseline log mortality rate around 0. - Year coefficient (β_{Year}): $\beta_{\text{Year}} \sim \text{Normal}(0, 10)$, reflecting the assumption of modest temporal effects. - Cause coefficients (β_{Cause}): $\beta_{\text{Cause}} \sim \text{Normal}(0, 10)$, capturing variation in baseline mortality among causes. - Interaction coefficients ($\beta_{\text{Year}\times\text{Cause}}$): $\beta_{\text{Year}\times\text{Cause}} \sim \text{Normal}(0, 5)$, regularizing cause-specific temporal trends.

Hierarchical Structure: - The model incorporates hierarchical effects for causes to borrow strength across categories, reducing overfitting, especially for less common causes of death. - Variance across years and causes is modeled explicitly to account for heterogeneity in mortality patterns.

Assumptions: 1. The Poisson distribution is appropriate for modeling count data, assuming non-negative outcomes. 2. Weakly informative priors regularize extreme estimates without overly constraining the model.

A.1 Posterior Predictive Checks

Posterior predictive checks (PPCs) were performed to evaluate the model's ability to replicate observed data. These checks include comparing the observed total deaths with the predicted distributions generated from the posterior.

Density Overlay Plots: Overlaying the observed and predicted distributions shows a high degree of overlap, indicating that the model fits the data well. Trace Plots: Visual diagnostics of model parameters confirm convergence, with stable chains and $\hat{R}$ statistics below 1.01 for all parameters.

Example PPC Visualization:

Observed Deaths	Predicted Mean	Predicted 95% CI
2001: Cause A	1,234	[1,200, 1,270]
2001: Cause B	987	[950, 1,020]
2022: Cause A	678	[640, 710]

This table illustrates the strong agreement between observed and predicted values across years and causes.

A.2 Supplementary Tables

Table 1: Coefficients and Credible Intervals A detailed table summarizing the posterior estimates for key parameters:

Parameter	Mean	95% Credible Interval
Intercept (β_0)	4.567	[4.510, 4.625]
Calendar Year (β_{Year})	-0.015	[-0.020, -0.010]
Mental Health Cause ($\beta_{\text{Mental Health}}$)	0.234	[0.190, 0.278]
Year × Cause Interaction ($\beta_{\text{Year}\times\text{Cause}}$)	0.012	[0.005, 0.019]

Table 2: Ranking Stability Across Years A summary of rank shifts for select causes:

Cause	Rank 2001	Rank 2022	Change
Cardiovascular	1	2	-1
Mental Health	15	7	+8
Respiratory Disease	10	9	+1

Additional Figures: Coefficient Plots: Visualizing the posterior distributions of key parameters, highlighting uncertainty intervals. Temporal Trends: Line plots of predicted deaths over time for the top five causes. Rank Change Heatmap: A heatmap showing rank changes for all 30 causes over the two decades.

B Idealized Methodology and Survey

B.1 Overview

Surveys provide a valuable tool for exploring factors that administrative mortality records cannot capture, such as self-reported health behaviors, perceptions of risk, and socioeconomic determinants. This survey is designed to investigate contributors to mortality trends in Alberta, including public awareness of leading causes of death, access to healthcare, and demographic disparities. By combining this data with mortality records, we can achieve a richer understanding of trends and priorities in public health.

B.2 Sampling Approach

Sampling Frame: The target population includes Alberta residents aged 18 years and older. The sampling frame will draw from household lists available through provincial databases.

Sampling Technique: A multistage stratified random sampling method will ensure representativeness:

- Primary Stratification: Divide the population by geographic region (urban, suburban, rural).
- Secondary Stratification: Within each region, stratify by demographic factors such as age and gender.
- Random Selection: Use random digit dialing (RDD) and mailed survey invitations to select households and respondents.

Sample Size: A sample size of 5,000 respondents is targeted, ensuring a 95% confidence level with a $\pm 2\%$ margin of error. Oversampling of underrepresented groups (e.g., rural residents, indigenous populations) will address potential response biases.

B.3 Respondent Recruitment

Recruitment Methods:

- Mixed-Mode Recruitment:
  - Mail: Send initial survey invitations with a unique survey code for online access.
  - Phone: Follow up with non-respondents using telephone calls to encourage participation.
  - In-Person: Use trained surveyors for areas with low response rates or where digital access is limited.
- Incentives:
  - Offer small monetary rewards (e.g., $10 gift cards) or entry into a prize draw to encourage participation.

  -Highlight the societal importance of their contribution to public health research.

Inclusion and Exclusion Criteria:

- Inclusion: Residents of Alberta, aged 18 and above, willing to participate in a 20-minute survey.

- Exclusion: Individuals unable to provide informed consent or residing temporarily in the province.

Nonresponse Bias Mitigation:

- Implement multiple follow-ups (e.g., reminders via email, phone, or postcards).
- Offer alternative formats (e.g., paper, online) to accommodate accessibility needs.

B.4 Data Validation

Validation Techniques:

- Consistency Checks:
  - Include repeated questions in different formats to ensure consistent responses.
  - Use embedded logic to flag unrealistic or contradictory answers (e.g., reporting both “excellent health” and “frequent hospitalization”).
- Demographic Comparisons:
  - Compare respondent demographics with census data to ensure representativeness.
  - Apply weighting to adjust for over- or under-represented groups.
- Pilot Testing:
  - Conduct a pilot survey with 200 participants to identify issues in question clarity, survey flow, and response patterns.

B.5 Idealized Survey Questionnaire

Below is an example of a survey designed based on the structure described:

Demographics:

- What is your age? (Open-ended)
- What is your gender? (Male / Female / Non-binary / Prefer not to say)
- What is your highest level of education? (No formal education / High school / College / Graduate degree)

Health Behaviors:

- Do you currently smoke? (Yes / No)

- On average, how many days per week do you exercise for at least 30 minutes? (0–7 days)
- How would you describe your diet? (Poor / Fair / Good / Excellent)

Healthcare Access:

- In the past 12 months, have you delayed seeking medical care due to cost? (Yes / No)
- How satisfied are you with the healthcare services available in your area? (1 = Very Dissatisfied, 5 = Very Satisfied)

Awareness of Mortality Causes:

- What do you believe is the leading cause of death in Alberta? (Open-ended)
- How at risk do you feel for cardiovascular disease? (1 = Not at all, 5 = Extremely)

Self-Reported Health:

- Do you have any of the following conditions? (Select all that apply: Diabetes / Hypertension / Mental Health Disorder / Cancer / Other)
- In the past year, how many times have you been hospitalized? (Open-ended)

Reference

Alberta, Government of. 2022. "Open Data Alberta: Mortality Statistics by Cause and Year." https://data.alberta.ca.

Disease Control, Centers for, and Prevention. 2021. "Leading Causes of Death and Mortality Trends." https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm.

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

Gabry, Jonah, and Tristan Mahr. 2024. "Bayesplot: Plotting for Bayesian Models." https://mc-stan.org/bayesplot/.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Organization, World Health. 2021. "Global Health Estimates 2020: Deaths by Cause, Age, Sex, by Country and by Region, 2000–2019." Geneva: World Health Organization. https://www.who.int/data/gho.

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

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

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.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation*.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.