Temporal Trends in Heart Disease and Diabetes Mortality in Alberta: A Negative Binomial Regression Analysis of the Impact of Fast Food Consumption*

Alexander Sun

March 16, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Wickham et al. (2019).

The remainder of this paper is structured as follows. Section 2….

In recent decades, the prevalence of heart disease and diabetes has surged globally, a trend paralleled by the increasing consumption of fast food. This paper focuses on Alberta, Canada, where these health challenges have become particularly pronounced. While numerous studies have linked fast food consumption to various health outcomes, few have directly examined its impact on mortality rates from heart disease and diabetes within this region. This gap in research motivates our study, which aims to analyze temporal trends in these mortality rates and discuss their potential association with the rise in fast food consumption, despite the absence of direct consumption data.

Using mortality data from the Alberta government, spanning two decades, we applied negative binomial regression models to analyze changes in heart disease and diabetes mortality rates. Our findings reveal significant temporal trends in these rates, with notable increases that correspond to periods of reported national and global rises in fast food consumption. While direct causation cannot be established due to the lack of specific consumption data, the temporal correlations underscore the potential health impacts of dietary habits.

*Code and data are available at: LINK.

The importance of this research lies in its contribution to the ongoing dialogue about public health strategies aimed at combating heart disease and diabetes. By highlighting the temporal association between increased mortality rates and the era of rising fast food consumption, this study emphasizes the need for targeted public health interventions and policies.

This paper is organized as follows: Following the introduction, the second section reviews existing literature on the relationship between diet and chronic diseases, establishing the theoretical foundation for the study. The third section describes the data and methodology, including the rationale behind the choice of negative binomial regression. The fourth section presents our findings, detailing the correlation between fast food consumption and mortality rates. The fifth section discusses the implications of these findings for public health policy and suggests directions for future research. The final section concludes the paper, summarizing the key contributions and urging for proactive measures in dietary education and regulation. Through this structured exploration, the paper contributes valuable insights into the diet-disease nexus, advocating for informed dietary choices as a cornerstone of public health.****

2 Data

2.1 Data source

This analysis will be carried out in **R** (**R?**) using packages **tidyverse** (**tidyverse?**), **dplyr** (**dplyr?**), **ggplot2** (**ggplot2?**), **knitr** (**knitr?**). The data set used in this paper is called Leading causes of death and was collected from the Alberta Provincial Government. The data set consists of a ranking of the 30 most common causes of death each year in Alberta. The data covers the last two decades, but for our research purpose we will focus on the last five years.

2.2 Broader Context of the Dataset

The availability of detailed public health data, such as the mortality statistics from Alberta, is crucial for the formulation of informed public health policies and strategies. Within the broader Canadian context, Alberta's commitment to data transparency enables a deeper analysis of health trends and outcomes, serving as a model for other provinces and territories. The analysis of mortality data plays a pivotal role in identifying health trends, assessing the burden of diseases, and planning public health interventions. By focusing on specific causes of death, researchers and policymakers can tailor strategies to target the underlying factors contributing to these trends, ultimately aiming to improve health outcomes and reduce preventable deaths.

Table 1: Annual Deaths by Cause (2016-2021)

Year	Acute Myocardial Infarction	All Other Forms of Ischemic Heart Disease	Atherosclerotic Cardiovascular Disease	Diabetes Mellitus	Congestive Heart Failure
2016	1102	1626	885	502	352
2017	1028	1678	817	584	374
2018	1071	1788	630	577	347
2019	1061	1886	745	569	430
2020	1067	1897	678	743	387
2021	1075	1939	463	728	403

2.3 Variables

The dataset comprises several key variables, central to this study's focus on heart disease and diabetes mortality rates:

Causes of Death: Specifically, the dataset categorizes mortality into detailed causes, including:

- **All Other Forms of Ischemic Heart Disease**: This category encompasses various conditions related to reduced blood flow to the heart muscle, excluding acute myocardial infarction.
- **Acute Myocardial Infarction (Heart Attack)**: Fatalities resulting directly from heart attack incidents.
- **Atherosclerotic Cardiovascular Disease**: Deaths caused by atherosclerosis, a condition characterized by the hardening and narrowing of the arteries due to plaque buildup, leading to cardiovascular problems.
- **Diabetes Mellitus**: Mortality attributed to complications arising from diabetes, a chronic condition affecting blood sugar regulation.
- **Congestive Heart Failure**: Deaths resulting from the heart's inability to pump blood effectively, often a consequence of other heart conditions.

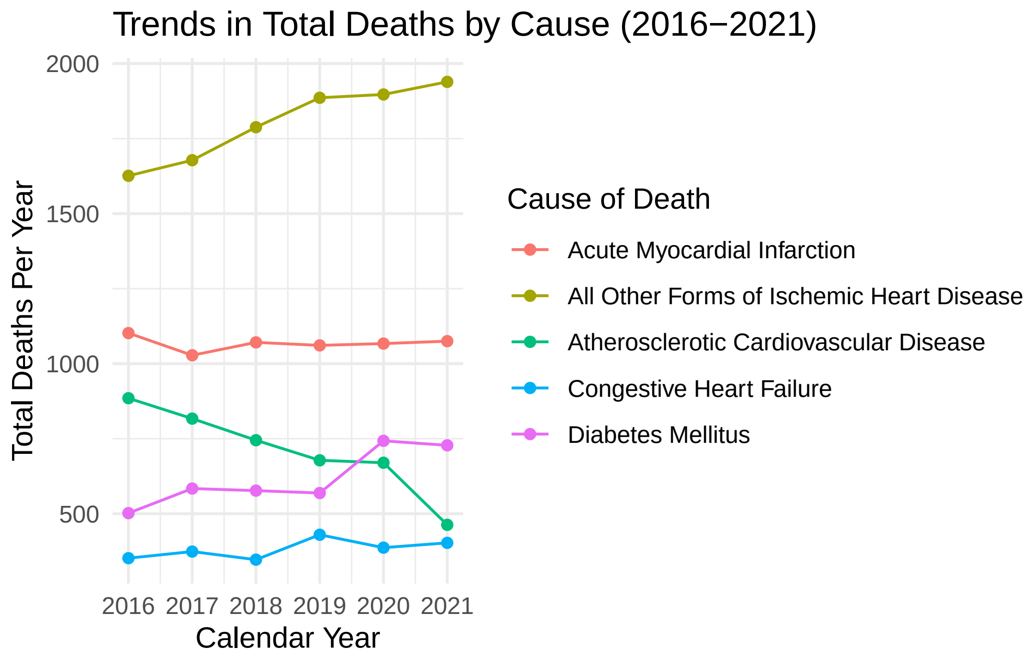
These five causes of death were chosen due to their correlation with unhealthy diets. We collected the recorded total deaths in Alberta from each of the above variables from 2016-2021 inclusive. This is shown in the table above.

2.4 Data Preparation and Cleaning

The dataset was filtered to isolate deaths attributed to our causes of interest: acute myocardial infarction, all other forms of ischemic heart disease, atherosclerotic cardiovascular disease, diabetes mellitus, and congestive heart failure. This selection was crucial to align our study with its objectives, ensuring a focused examination of these specific health outcomes. Subsequently, we removed all missing values from the dataset. The final data points were compiled for use in our analysis later in the paper.

2.5 Preliminary Observations and Exploration

Our initial analysis revealed several notable observations. First, the trend analysis suggested a correlation between the years and mortality rates for specific causes of death, hinting at the possible influence of external factors such as healthcare policies or changes in societal health behaviors. Since we are looking at a five-year span from 2016 to 2021, the onset of the COVID-19 pandemic introduced a plethora of health-related policies and regulations that may significantly impact our data. In the figure below, we plotted the total deaths from each of the causes to visualize if there were any spikes in the data.



From the graph depicted above, we see that although most causes kept their total death count per year roughly the same, Diabetes saw a noticeable spike form 2020 onwards and Atheroscleratic cardiovascular disease saw a significant decrease.

3 Model

In our analysis of the association between the total number of deaths and significant causes of death in Alberta, we employed two distinct regression models to facilitate our predictions: the Poisson model and the Negative Binomial model. Both models are well-suited for analyzing count data, where the goal is to explore the relationship between several independent variables (predictors) and a dependent variable (count outcomes). In the context of our study, the independent variables are the causes of death, including “Acute Myocardial Infarction,” "All Other Forms of Ischemic Heart Disease," "Atherosclerotic Cardiovascular Disease," "Diabetes Mellitus," and "Congestive Heart Failure." These variables serve as predictors for our dependent variable, which is the count of deaths per year in Alberta from 2016 to 2021.

3.1 Poisson Model

The Poisson regression model is characterized by its simplicity and efficiency in modeling the count data. It assumes that the event occurrence rate is constant across the observed period and that these events occur independently of each other. Despite its advantages, the Poisson model may fall short in handling overdispersion, where the variance exceeds the mean in the count data.

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

$$\log(\lambda_i) = \alpha + \beta_1 \times \text{AMI}_i + \beta_2 \times \text{OtherIHD}_i + \beta_3 \times \text{ASCVD}_i + \beta_4 \times \text{Diabetes}_i + \beta_5 \times \text{CHF}_i$$

where:

- y_i is the count of deaths due to the i-th cause.
- λ_i is the expected count of deaths for the i-th cause.
- α, β_1, β_2, β_3, β_4, and β_5 are the model parameters to be estimated.
- AMI, OtherIHD, ASCVD, Diabetes, and CHF are abbreviations of our previously selected variables.

3.2 Negative Binomial Model

To address potential overdispersion in our data, we also applied the Negative Binomial regression model. This model extends the Poisson by introducing an extra parameter to account for the overdispersion, offering a more flexible approach to fit our data. The formulation of the Negative Binomial model is:

$$y_i|\mu_i, \phi \sim \text{NegBin}(\mu_i, \phi) \tag{1}$$

$$\mu_i = \exp(\alpha + \beta x_i) \tag{2}$$

$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$

$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$

$$\phi \sim \text{Exponential}(1) \tag{5}$$

In this model, y_i again denotes the total number of deaths per year, with μ_i being the expected count adjusted for overdispersion through the dispersion parameter θ. We assume the coefficients like causes of death follow a normal distribution.

Given the observed overdispersion in our dataset, the Negative Binomial model is anticipated to offer a more accurate and reliable fit compared to the Poisson model. By incorporating the extra dispersion parameter, it allows us to better capture the variability in death counts across different causes, providing a nuanced understanding of how each cause contributes to overall mortality.

In applying these models, we aim to discern the relative impact of specified causes of death on the total number of deaths, while also accounting for the distributional characteristics of our count data. Through this comparative analysis, we seek to identify the most suitable model for our dataset, thereby enhancing the reliability of our findings and conclusions.

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

3.2.1 Model justification

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

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

4 Results

Our results are summarized in Table 2.

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Table 2: Explanatory models of flight time based on wing width and wing length

	First model
(Intercept)	1.12
	(1.70)
length	0.01
	(0.01)
width	−0.01
	(0.02)
Num.Obs.	19
R2	0.320
R2 Adj.	0.019
Log.Lik.	−18.128
ELPD	−21.6
ELPD s.e.	2.1
LOOIC	43.2
LOOIC s.e.	4.3
WAIC	42.7
RMSE	0.60

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

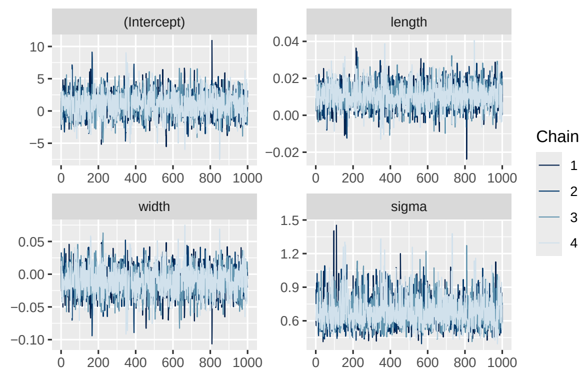
In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows...

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows...

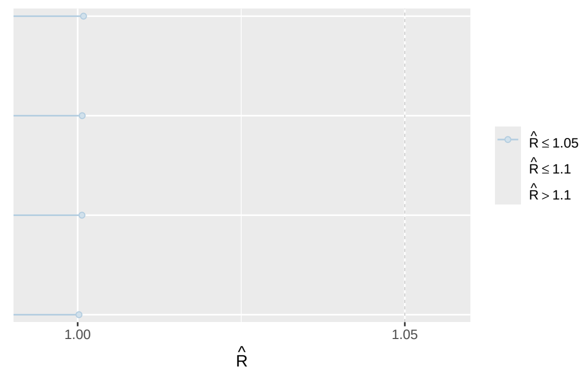
B.2 Diagnostics

Figure 1a is a trace plot. It shows... This suggests...

Figure 1b is a Rhat plot. It shows... This suggests...



(a) Trace plot



(b) Rhat plot

Figure 1: Checking the convergence of the MCMC algorithm

References

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, 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.