Noncommunicable Diseases Within Canada: Examining the Impact of Lifestyle Behaviors on Metabolic Risk Factors

Wardah Ali, Navyasri Chinthapatla, Danae McCulloch, Safeen Mridha, Ayda Takehei

DATA 604: Working with Data at Scale

FALL 2024

Table of Contents

Introduction

Noncommunicable diseases (NCDs) pose a significant challenge to public health systems worldwide as they play a major role in global mortality every year. It was stated that 86% of all deaths in Canada were accounted by NCDs in the year of 2021 (World Health Organization, n.d.)**.** Noncommunicable diseases are defined as illnesses that cannot be transmitted from one person to another, which include conditions such as cardiovascular diseases, diabetes, respiratory diseases and cancer (World Health Organization, n.d.)**.** Since NCDs are not transferrable, factors such as lifestyle behaviors play an integral role in increasing metabolic risk factors that heighten the susceptibility of these diseases. Given the ubiquitous nature of the stress that NCDs place on public health systems worldwide, understanding the impact of lifestyle behaviors is critical for the development of effective prevention and intervention strategies.

The scope of this project will be looking at these specific factors and its relation to NCDs: physical inactivity, cholesterol levels, tobacco use, hypertension, Body Mass Index (BMI), and alcohol consumption. Data regarding these factors are organized in separate datasets which have been derived from the World Health Organization (W.H.O). These listed factors were decided since they are considered mostly controllable, however this analysis does not consider uncontrollable environmental and socio-economic factors which limit the findings. Investigating these factors within the population is important as it can indicate the overall population health and give warning to public health professionals regarding potential surges.

Guiding Questions & Dataset

1. How has the number of deaths attributed to NCDs trended within Canada since 2000?

We are going to visualize the trend of deaths caused by NCDs to observe the effect of NCDs have on the general health of the Canadian population over the last decades. Given trends of rising obesity, rates of diabetes, as well as instances of heart attacks, it is fair to speculate that the number of deaths attributed to NCDs have increased since 2000; however, these trends will be obtained through analyzing and plotting data related to NCD-causing death using line graphs. Notably, the lifestyle behaviors of a person prove to play an integral role in causing and preventing noncommunicable diseases. Therefore, plotting trends can help health officials determine which factors should be focused on when it comes to increasing awareness regarding NCD prevention.

Datasets: The dataset used to visualize these trends is obtained directly from a study conducted by the W.H.O., where the number of people dying each year from four classes of NCDs was recorded every year from 2000 to 2019. Data was collected from every UN-recognized country across the world. For countries with high-quality vital registration systems, death information, as well as cause-of-death, was estimated and submitted by member states directly; for countries that lack infrastructure and do not have high-quality death registration data, the cause and number of deaths were estimated using various methods, such as household surveys and population censuses. After receiving these estimates, the weighted averages were calculated.

2. Respiratory Disease: How does smoking tobacco correlate with obesity within Canada? How do these factors correlate with the numbers of deaths attributed to respiratory diseases?

Smoking tobacco is a prevailing cause of chronic respiratory diseases, especially lung cancer. And typically, people who smoke lose their appetite causing them to lose weight and because of their lack of strength they are not physically active either. Therefore, it would be interesting to look at the relation between the two and their effect on the death rates of respiratory diseases.

Datasets: A total three datasets will be analyzed: “Age-standardized estimates of current tobacco use, tobacco smoking and cigarette smoking (Tobacco control: Monitor” , “Prevalence of obesity among adults, BMI >= 30 (age-standardized estimate) (%)” and “Number of deaths attributed to non-communicable diseases, by type of disease and sex”. Tobacco prevalence rates are given in percentage for all the countries around the world over the two decades starting from 2000 – 2018 for male and female 15 years and above who currently use any tobacco product daily or non-daily basis. The data for the smoked tobacco will be used in the analysis. For the obesity dataset the percentage of adults aged 18+ years with a body mass index (BMI) of 30 kg/m2 or higher are given for all the countries over the two decades starting from 2000 – 2022 for male and female. The data was collected from a pooled analysis of 3663

population-representative studies with 222 million children, adolescents and adults. In the third dataset, the number of deaths associated to respiratory diseases in Canda from 2000-2019 will be used.
**Limitation: The two datasets of tobacco and obesity differ in ages therefore making the population inconsistent.*

3. Diabetes: What relationship exists between physical inactivity and obesity within Canada? How are they related to the number of deaths attributed to diabetes?

The relationship between obesity and physical activity is generally inverse, meaning as levels of physical activity decrease, the likelihood of obesity increases. Physical inactivity is a key factor contributing to weight gain and the development of obesity, which, in turn, raises the risk of developing Type 2 diabetes. By examining the relationship between the two, we can gain a better understanding of the extent to which physical inactivity may contribute to obesity.

Datasets: Three datasets will be used: "Prevalence of insufficient physical activity among adults aged 18+ years (%)", "Prevalence of obesity among adults, BMI >= 30 (age-standardized estimate) (%)" and "Number of deaths attributed to non-communicable diseases, by type of disease and sex". The percentage of insufficient physical activity was determined by the percent of population attaining less than 150 minutes of moderate-intensity physical activity per week, or less than 75 minutes of vigorous-intensity physical activity per week or equivalent. The estimates were based on self-reported physical activity captured using the GPAQ (Global Physical Activity Questionnaire) and the IPAQ (International Physical Activity Questionnaire). A Bayesian hierarchical model was used to produce estimates for each country, where only the Canada data will be extracted in this report from the years 2000-2022 for male and female. As mentioned earlier, the obesity dataset for Canada from 2000-2022 will be used in the analysis. In the third dataset, the number of deaths associated to Diabetes in Canada from 2000-2019 will be used.

4. Cardiovascular Disease: How do cholesterol levels influence the prevalence of hypertension? How do these factors correlate to the numbers of deaths attributed to cardiovascular disease?

Cholesterol is easily accessible nowadays through processed and highly saturated fat foods. It is important to monitor cholesterol levels as an excess of this bulky substance can add stress to the heart causing the risk of heart issues or even heart failure (American Heart Association, 2024).

Datasets: Three datasets will be interpreted: "Mean Total Cholesterol (age-standardized estimated", "Prevalence of hypertension among adults aged 30-79" and "Number of deaths attributed to non-communicable diseases, by type of disease and sex". Cholesterol and Hypertension values will be plotted against each other on a scatter plot. Cholesterol is measured in millimoles per liter (mmol/L) which is standard to measuring the amount of glucose concentration in the blood. An unhealthy cholesterol level is defined to be more than 5.0 mmol/L (source). Hypertension is measured in millimeters of mercury (mmHg). Prevalence of hypertension is defined when systolic blood pressure is greater than or equal to 140 mmHg and the diastolic blood pressure is greater than or equal to 90 mmHg. Both datasets are population-based surveys where Cholesterol obtained 1,127 studies in ages 18 and older; whereas Hypertension obtained 1,201 studies in ages 30-79*. In the third dataset, the number of deaths associated to Cardiovascular in Canda from 2000-2019 will be used.
**Limitation: These two datasets differ in ages therefore we will be excluding data from ages 18 to 29 for Cholesterol to keep the timeframes consistent.*

5. Malignant Neoplasms (Cancer): To what extent does pure alcohol consumption impact cholesterol levels? How do these factors correlate to the numbers of deaths attributed to malignant neoplasms (cancer)?

Alcohol consumption can significantly impact cholesterol levels in the body. Excessive alcohol intake can impair metabolism, raise cholesterol levels, and increase the risk of cardiovascular diseases and certain types of cancer.

Datasets: Three datasets will be used for Analysis: "Mean Total Cholesterol (age-standardized estimate", "Alcohol, total per capita (15+) consumption (in liters of pure alcohol (SDG indicator 3.5.2) and "Number of deaths attributed to non-communicable diseases, by type of disease and sex". Alcohol Per Capita (APC) means the sum of the three-year average recorded and unrecorded pure alcohol consumed per adult in liters. Recorded alcohol consumption refers to official statistics (production, import, export, and sales or taxation data), while unrecorded alcohol consumption

refers to alcohol that is not taxed and is outside the usual system of governmental control. The Canada data for the years 2000-2020 will be extracted for use. The cholesterol data described earlier will be used to plot a scatter plot for alcohol and cholesterol. In the third dataset, the number of deaths associated to Malignant Neoplasms (Cancer) in Canada from 2000-2019 will be used.
**Limitation: These two datasets differ in ages therefore we will be excluding data from ages 15 to 17 for Alcohol to keep the timeframes consistent.*

Tasks

Data Cleaning: The datasets contain no null values, but we anticipate potential outliers that will require handling.

Data Filtering: The datasets can be filtered based on data points within Canada from the years 2000-2019.

Data Transformation: The datasets for the specified factors and the datasets for the number of deaths attributed to NCDs will be merged based on years (Guiding Questions 2 – 4). Multiple regression analysis will be performed to see the impacts of the specific factors being analyzed in each guiding question on the selected NCD. This will be performed by transformation techniques such as normalization, standardization, and possibly the minimum and maximum scaler to improve regression model accuracy.

Data Visualization: We will employ various methods to aid in the data visualization component of our project.

- **Line Chart:** To analyze trends and fluctuations in factors over time. Will be used for Guiding Question 1.
- **Heat Map:** To analyze the number of deaths attributed to the following NCDs: diabetes, cancer, respiratory and cardiovascular diseases in Canada from 2000 to 2019.
- **Bar Graph:** To help visualize Guiding Questions 2-5.
- **Scatter Plot: Will** be used in Guiding Questions 2-5.

Python Libraries

- NumPy – For handling series data and in case of any numerical computations.
- Pandas – For exploratory data analysis tasks.
- Seaborn – For heat maps
- Matplotlib – Data visualization like line charts and bar graphs.
- Plotly - For interactive visualization
- Scikit-learn – For Regression

Software/Environment

- Jupyter Notebook: For our project we will use Jupyter Notebook but will be uploading it on GitHub.
- Git Hub: We plan to make a private organization on GitHub where we all can collaborate on the code.

References

American Heart Association. (2024, October 14). What is cholesterol? www.heart.org. https://www.heart.org/en/health-topics/cholesterol/about-cholesterol.

NCD Portal. (n.d.). *Country profile: Canada*. Retrieved November 3, 2024, from https://ncdportal.org/CountryProfile/GHE110/CAN.

Nestle, M. (2020). Perspective: Challenges and controversies in food, nutrition, and health. Annual Review of Food Science and Technology, 11(1), 1- 19. https://www.annualreviews.org/content/journals/10.1146/annurev-food-032519-051827.

U.S. Department of Health and Human Services. (2024, June 17). Control your cholesterol. National Institutes of Health. https://newsinhealth.nih.gov/2019/02/control-your-cholesterol.

World Health Organization. (n.d.). *Canada: Noncommunicable diseases country profiles*. https://ncdportal.org/CountryProfile/GHE110/CAN

World Health Organization. Chronic respiratory diseases. https://www.who.int/health-topics/chronic-respiratory-diseases#tab=tab_3.

World Health Organization. (n.d.). Noncommunicable diseases: Risk factors. WHO. https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/ncd-risk-factors.