2023 Ontario Monthly Communicable Disease Cases Study*

Discussions and Graphs

Ruiyang Wang

Invalid Date

This study analyzes the Monthly Communicable Disease cases for Ontario in 2023, focusing on 3 mainly cosiderred disease: HIV, Salmonellosis, and Sporadic Influenza. This paper presenst the general graphs for each disease, dicusses the genral trends, and further analysis.

1 Introduction

In 2023, Ontario, Canada, experienced distinct trends in the prevalence of communicable diseases such as HIV, Salmonellosis, and Sporadic Influenza. Monitoring and analyzing the seasonal patterns of these diseases is crucial for public health.

HIV, a chronic and life-threatening condition caused by the human immunodeficiency virus, has shown relatively stable case numbers throughout the year.

Salmonellosis is an infection caused by the Salmonella bacteria, typically contracted through the consumption of contaminated food or water. It commonly results in gastrointestinal symptoms such as diarrhea, fever, and abdominal cramps.

Sporadic Influenza, commonly known as the flu, is a highly contagious respiratory illness caused by influenza viruses.

This paper aims to analyze the statistical trends of these three diseases in Ontario throughout 2023, providing a general overview of their seasonal patterns. We also identify potential drivers of these trends and discuss their implications for public health policy. The findings will contribute to a better understanding of disease dynamics in the region and then support the development of prevention and strategies.

*Code and data are available at: https://github.com/Ruiyang-Wang/STA304-Paper-1.git

We use data and codes from R Core Team (2023), Gelfand (2022), and Wickham et al. (2019a). The exact data is from Toronto (2023).

The remainder of this paper is structured as follows. -Section 2 -Section 3 -Section 3.1 -Section 3.2 -Section 3.3 -(**trd-first-point?**)

2 Data Summary

Our data is from Gelfand (2022), named "Monthly Communicable Disease Surveillance". The time range of the whole dataset is from 2016 to 2024. We only extract 3 classic diseases in this dataset for simplification, and set the year of analysis to 2023, which is the most recent full-year data and most unaffected data by COVID-19 after the 2019 Pandemic.

3 Disease Analysis

3.1 HIV

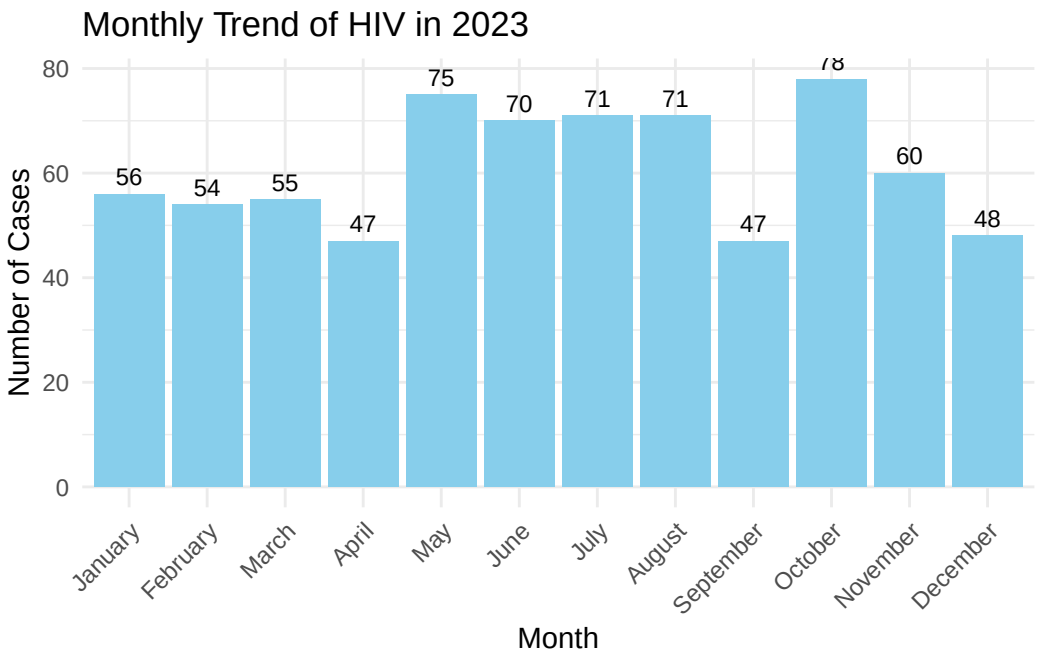

Figure 1: Monthly cases of HIV

3.1.1 Characteristics of HIV in 2023

- The number of HIV cases ranges from around 47 to 78 per month.
- There are noticeable fluctuations throughout the year, with peaks in May and October and a significant drop in April and September.
- The case count is relatively stable in other months, hovering between 50 and 75 cases.

3.1.2 Summary

HIV cases in Ontario displayed moderate monthly fluctuations, with notable peaks in May and October. These variations could be influenced by changes in testing rates or social behaviors.

3.2 Salmonellosis

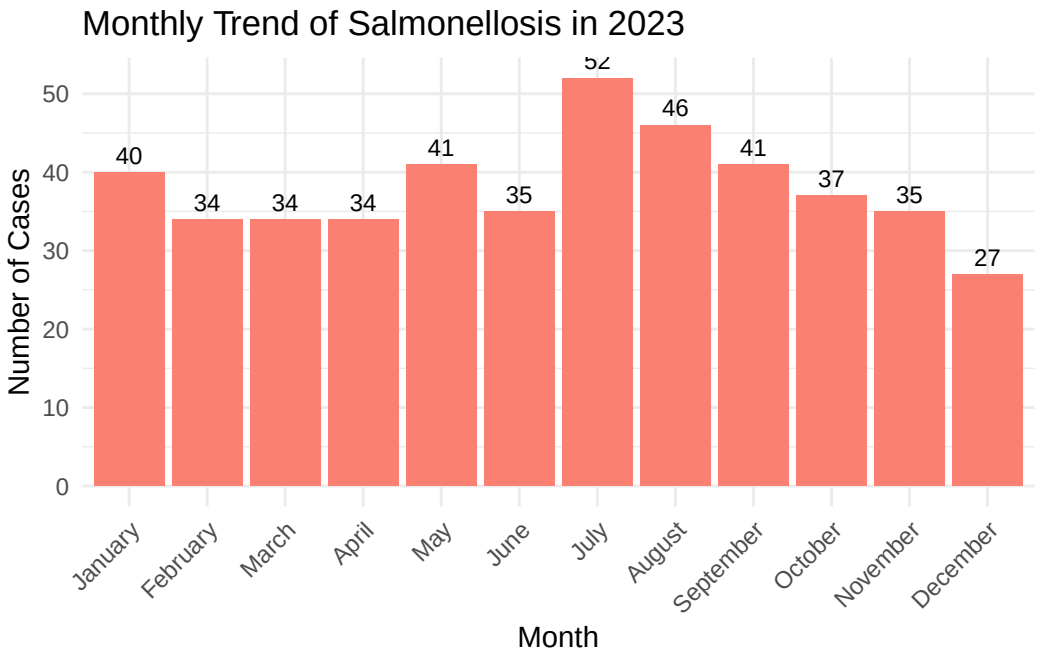

Figure 2: Monthly cases of Salmonellosis

3.2.1 Characteristics of Salmonellosis

- The number of cases is relatively low, ranging from 27 to 52 per month.

- There is a significant peak in July, with a steady increase from April, followed by a decline from August to December.
- Salmonellosis cases are highest in summer months, which is consistent with the bacteria's transmission patterns.

3.2.2 Summary

Salmonellosis cases peaked in the summer months, particularly in July, likely reflecting increased risk factors such as outdoor food consumption and higher temperatures.

3.3 Sporadic Influenza

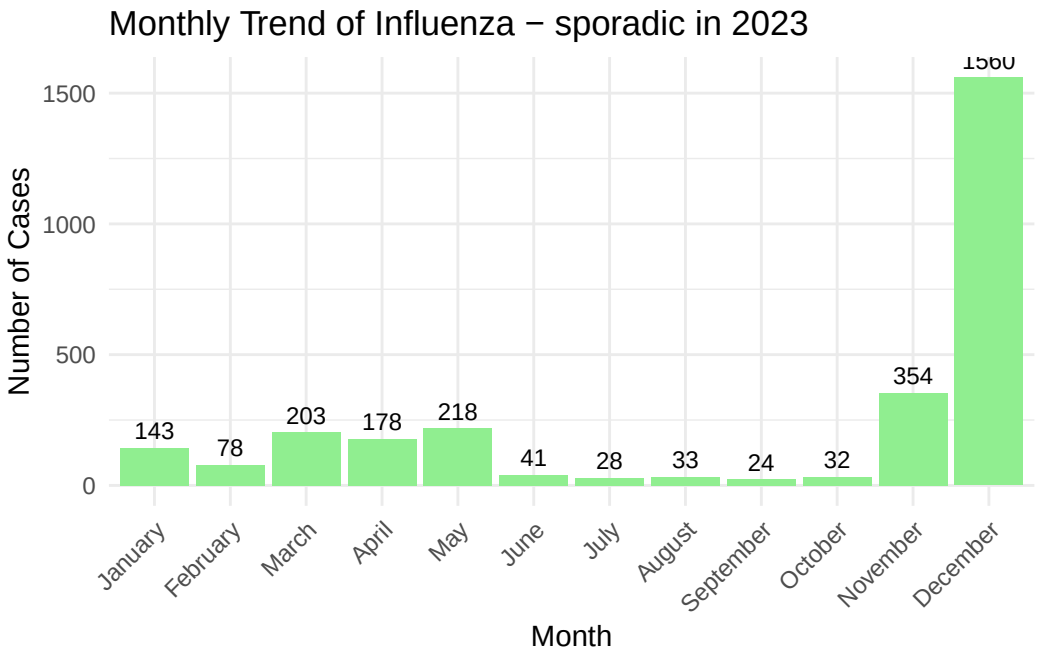

Figure 3: Monthly cases of Sporadic Influenza

3.3.1 Characteristics

- There is a sharp increase in cases in December, reaching a peak of 1560 cases, while the rest of the year shows relatively low numbers.
- A smaller peak in March and a gradual increase from October to November indicate sporadic activity.

3.3.2 Summary

Sporadic Influenza cases spiked sharply in December, marking the onset of the flu season. A smaller peak in March suggests occasional outbreaks earlier in the year.

3.4 Weaknesses and Next Steps

3.4.1 Limitations and Weaknesses of the Data Processing Process

1. **Data Completeness and Quality:**
    - The dataset used in this analysis may contain missing or incomplete data for certain months or diseases, which can skew the results and affect the reliability of the observed trends.
    - Inconsistencies in data recording and reporting, such as varying definitions or case counting methods, could lead to potential biases in the analysis.
2. **Temporal and Spatial Limitations:**
    - The analysis is limited to the year 2023, providing a snapshot view without the ability to observe longer-term trends or compare with previous years.
    - Data is aggregated at the provincial level, which may obscure regional variations within Ontario that could offer more nuanced insights.
3. **Lack of Contextual Factors:**
    - The data does not include contextual information such as socioeconomic factors, healthcare access, or specific public health interventions, which are crucial for understanding the underlying causes of disease trends.
    - External factors like changes in public health policies, seasonal events, or unrecorded outbreaks were not accounted for in this analysis.
4. **Statistical Assumptions:**
    - The analysis assumes that monthly case counts follow a consistent pattern, but this may not hold true for diseases influenced by sporadic outbreaks or variable reporting practices.
    - The use of basic descriptive statistics and visualizations limits the depth of insights that can be derived, lacking more advanced statistical or predictive modeling techniques.

3.4.2 Further Analysis for Each Disease

1. **HIV**
    - **Investigate Seasonal Factors:** Conduct a more detailed analysis of potential seasonal factors affecting HIV case counts, such as changes in social behaviors, festivals, or events that might influence transmission rates.
    - **Public Health Interventions:** Evaluate the impact of specific public health initiatives, such as awareness campaigns, increased testing, or outreach programs, to determine their effectiveness in reducing HIV transmission during peak months.
    - **Longitudinal Analysis:** Extend the analysis to include data from previous years to identify long-term trends and changes in HIV incidence over time, adjusting for any policy or healthcare changes.
2. **Salmonellosis**
    - **Foodborne Outbreak Analysis:** Perform a detailed investigation of foodborne outbreaks, linking reported cases with specific food sources, handling practices, and outbreak reports from health agencies.
    - **Environmental and Behavioral Correlations:** Use environmental data (temperature, humidity) and behavioral data (e.g., outdoor gatherings, food festivals) to build a model explaining seasonal fluctuations in Salmonellosis cases.
    - **Geographical Sub-analysis:** Break down the data by regions within Ontario to identify hotspots for Salmonellosis and evaluate the effectiveness of local food safety regulations and practices.
3. **Sporadic Influenza**
    - **Seasonal Modeling:** Develop a predictive model for influenza cases using historical data, vaccination rates, and climate data to forecast future outbreaks and inform public health preparedness.
    - **Vaccination Coverage Impact:** Analyze the relationship between vaccination coverage rates and influenza case trends to understand the effectiveness of the flu vaccination program in reducing the disease burden.
    - **Healthcare System Impact:** Assess the impact of the influenza peak on healthcare resources, including hospitalizations, ICU admissions, and healthcare worker absenteeism, to evaluate the preparedness and resilience of the healthcare system during the flu season.

A Appendix

In this analysis, we utilized several data sources and tools. The dataset for communicable disease trends was obtained from the City of Toronto's Open Data Portal (Toronto 2023). Data processing and visualization were performed using the `opendatatoronto` package (Gelfand 2022) in the R environment (R Core Team 2023).

The `tidyverse` package was a crucial tool for data wrangling and visualization (Wickham et al. 2019b). Additionally, the `Lahman` package was used for data manipulation (Friendly et al. 2020), and Bayesian regression models were fitted using `rstanarm` (Goodrich et al. 2022).

The analysis of the communicable disease trends showed significant seasonal variations in the data (Gebru et al. 2021).

Future work should focus on integrating more comprehensive datasets to provide a better understanding of the disease dynamics (Wickham et al. 2019a).

References

Friendly, Michael, Chris Dalzell, Martin Monkman, and Dennis Murphy. 2020. *Lahman: Sean "Lahman" Baseball Database*. https://CRAN.R-project.org/package=Lahman.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. "Datasheets for Datasets." *Communications of the ACM* 64 (12): 86–92.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

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/.

Toronto, City of. 2023. "Monthly Communicable Disease Surveillance Data." https://open.toronto.ca/dataset/monthly-communicable-disease-surveillance-data/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019b. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

———, et al. 2019a. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.