2023 Ontario Monthly Communicable Disease Specified Cases Study*

Discussions and Graphs

Ruiyang Wang

September 24, 2024

This paper examines the monthly trends of three communicable diseases—HIV, Salmonellosis, and Influenza(sporadic) - in 2023 of Ontario. We use data from the City of Toronto's Open Data Portal, and conduct an analysis to identify seasonal patterns and peaks for each disease. For HIV, the data shows a notable peak, suggesting potential influences on inappropriate social activities in the certain season. Salmonellosis shows a significant rise of case number during the summer, highlighting the importance of the role of environmental and food safety problems. Influenza cases, increased dramatically in December, marking the onset of the flu season. These trends underscore the need for targeted public health interventions, such as enhanced awareness campaigns during identified peak periods and improved vaccination coverage for influenza. Understanding these patterns is crucial for optimizing resource allocation,and improving healthcare systems decisions in Ontario.

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.

---

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

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.

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.

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.

```
# A tibble: 3 x 20
  Disease     `2023 YTD Cases` `2023 YTD Rate1` January February March April   May
  <chr>                  <dbl> <chr>               <dbl>    <dbl> <dbl> <dbl> <dbl>
1 HIV                      732 24.70                  56       54    55    47    75
2 Salmonel~                456 15.40                  40       34    34    34    41
3 Influenz~               2892 97.50                 143       78   203   178   218
# i 12 more variables: June <dbl>, July <dbl>, August <dbl>, September <dbl>,
#   October <dbl>, November <dbl>, December <dbl>,
#   `December 3 yr. Average2` <chr>, `2022 YTD Cases` <dbl>,
#   `2022 YTD Rate1` <chr>, `2022 Total Cases` <dbl>, `2022 Total Rate1` <chr>
```

- **Variable Explanation**

- "Disease": This variable represents the specific communicable diseases being analyzed in the dataset.

- "January - December" : These variables indicate the number of reported cases of each disease for each month in the year 2023. They provide a month-by-month breakdown of disease prevalence.

- I have cleaned the data by eliminating all unused diseases, and cancelled out all NA values or labels. I also changed the data type of numbers from str to double, in order to make sure the graphing process is accurate.
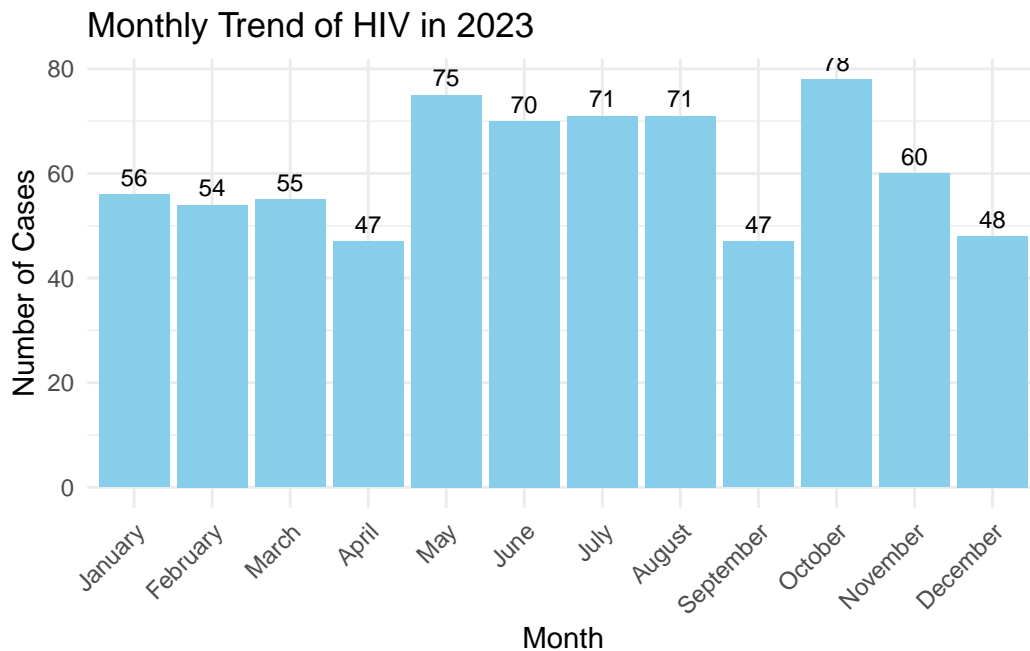
3 Disease Analysis

3.1 HIV

Figure 1: Monthly cases of HIV

3.1.1 Characteristics of HIV in 2023

- According to Figure.1, 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.
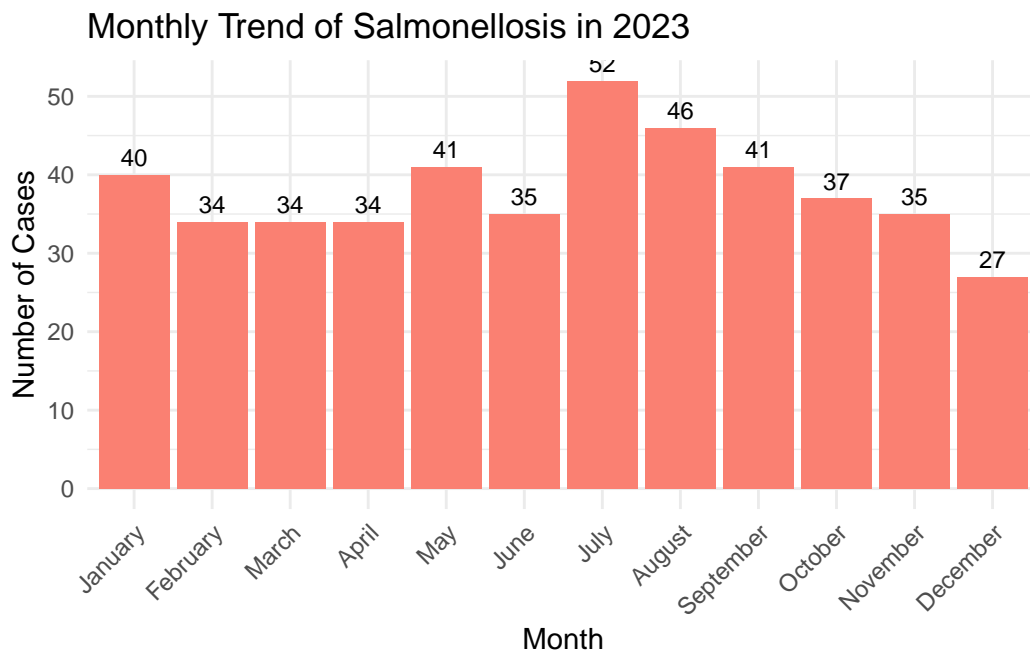
3.2 Salmonellosis

Figure 2: Monthly cases of Salmonellosis

3.2.1 Characteristics of Salmonellosis

- According to Figure.2, the number of cases ranges 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.
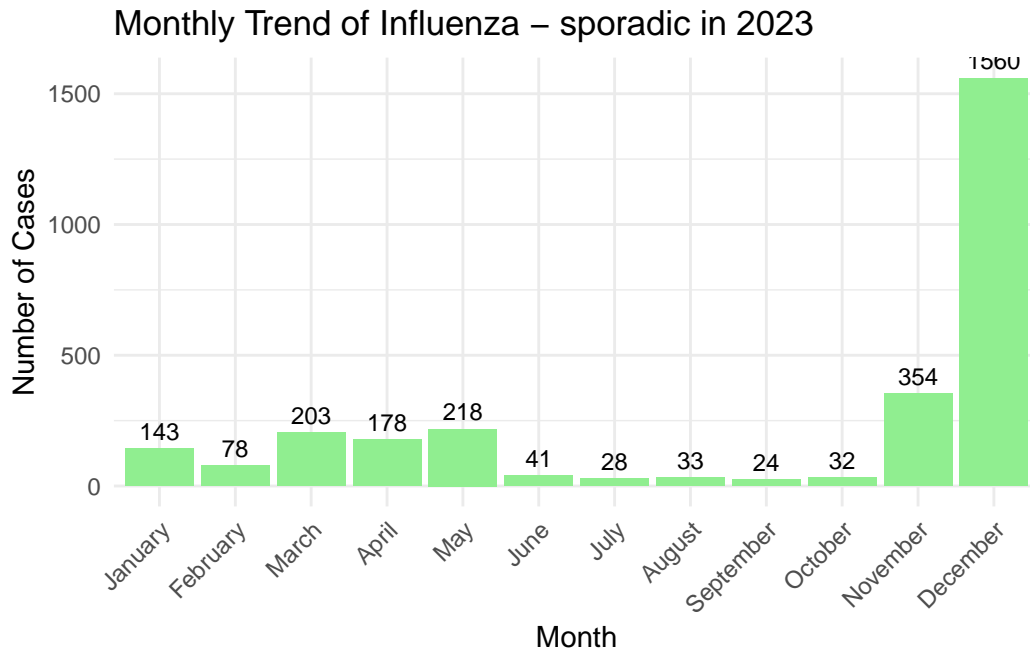
3.3 Sporadic Influenza

Figure 3: Monthly cases of Sporadic Influenza

3.3.1 Characteristics

- According to Figure.3, there is a sharp increase in cases in December, reaching a peak of 1560 cases. 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. **Temporal and Spatial Limitations:**

   - The analysis is limited to the year 2023, providing a short view without the ability to observe longer-term trends or compare with previous years.
   - This data might be affected by the COVID-19 symptoms due to pharmacological reasons.

2. **Lack of Contextual Factors:**

   - The data does not include contextual information such as socioeconomic factors, healthcare access, or specific public health interventions, which might be crucial for understanding the causes of disease trends.

3. **Statistical Perspective:**

   - 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 cases, such as changes in social behaviors, festivals, or events, which might influence the 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 an investigation of foodborne outbreaks, and link the reported cases with specific food sources and outbreak reports from health agencies.
   - **Environmental and Behavioral Correlations:** Use environmental data and behavioral data (e.g., outdoor gatherings, food festivals) to build a model to explain seasonal fluctuations for cases of Salmonellosis.

3. **Sporadic Influenza**

- **Seasonal Modeling:** Use additional data of vaccination and climates, develop a predictive model for influenza cases, in order to forecast future outbreaks and improve Ontario health systems.

- **Healthcare System Impact:** Use additional data, analyze the impact of the winter peak on healthcare resources, like hospitalizations and ICU admissions, 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.