Analyzing and Enhancing Outbreak Response in Ontario Healthcare Institutions: A Comprehensive Study Under the Health Protection and Promotion Act*

Monica Sainani

January 25, 2024

This project investigates outbreaks of gastroenteric and respiratory illnesses in Ontario healthcare institutions, focusing on response patterns mandated by the Health Protection and Promotion Act. Through detailed analysis, we aim to uncover insights into outbreak occurrence, durations, and causative agents. Findings will inform strategies to optimize outbreak management and enhance the efficiency of Toronto Public Health's response, contributing to improved public health outcomes.

Table of contents

*Code and data are available at: https://github.com/sainanim/Disease-Outbreaks-in-Toronto.git

0.1

```
#message: false
#echo: false

#### Preamble ####
# Purpose: Read daata from dataset of Disease Outbreaks in toronto
# Author: Monica Sainani
# Date: 23 January 2024
# Contact: monica.sainani@mail.utoronto.ca
# License: MIT
# Pre-requisites: run scripts/01-download_data.R followed by scripts/02-data_cleaning.R

#### Workspace setup ####
# install.packages("tidyverse")
# install.packages("knitr")
# install.packages("here")
# install.packages("ggplot2")

library(ggplot2)
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.2     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v lubridate 1.9.2     v tibble    3.2.1
v purrr     1.0.2     v tidyr     1.3.0
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
library(knitr)
library(here)
```

```
here() starts at /cloud/project/outputs/paper
```

```r
library(readr)

# read in the data
Disease_data <- read_csv(file = here("cleaned_dataset.csv"),
  show_col_types = FALSE)

head(Disease_data)
```

```
# A tibble: 6 x 12
   X_id Institution.Name   Institution.Address Outbreak.Setting Type.of.Outbreak
  <dbl> <chr>              <chr>               <chr>            <chr>
1     1 Amica On the Aven~ 1066 Ave Rd         Retirement Home  Respiratory
2     2 UHN Toronto Rehab~ 130 Dunn Ave        Hospital-Chroni~ Respiratory
3     3 Fairview Nursing ~ 14 Cross St         LTCH             Respiratory
4     4 Hellenic Home - U~ 2411 Lawrence Ave E LTCH             Respiratory
5     5 Hillcrest Reactiv~ 47 Austin Terr      Hospital-Chroni~ Enteric
6     6 SHSC - Sunnybrook~ 2075 Bayview Ave    LTCH             Enteric
# i 7 more variables: Causative.Agent.1 <chr>, Causative.Agent.2 <chr>,
#   Date.Outbreak.Began <date>, Date.Declared.Over <date>, Active <chr>,
#   Date_Outbreak_Began <dbl>, Date_Declared_Over <dbl>
```

1 Introduction

Under the Ontario Health Protection and Promotion Act (HPPA), healthcare institutions are obligated to monitor and report outbreaks of gastroenteric and respiratory illnesses. Toronto Public Health responds to these outbreaks, working collaboratively with healthcare facilities to implement infection prevention and control measures. This project aims to analyze and enhance the response to outbreaks in Ontario healthcare institutions, utilizing a dataset that captures key information under the HPPA.

2 Data

2.1 About the dataset

The dataset comprises records of outbreaks in healthcare institutions, including hospitals, long-term care homes, and retirement homes. It includes essential details such as the outbreak setting, types, causative agents, and key dates like the start and end of outbreaks. The data is a valuable resource for understanding the dynamics of infectious outbreaks within healthcare facilities and evaluating the effectiveness of response protocols.

2.2 Preprocessing and data cleaning

The data underwent preprocessing to ensure its suitability for analysis. Steps included handling missing values, converting date columns to the appropriate format, and categorizing variables. Imputation was performed judiciously, and categorical variables were transformed into factors for effective analysis. This rigorous preprocessing lays the foundation for meaningful insights from the dataset.

2.3 Variables of interest

The analysis will focus on key variables, including outbreak settings, types, causative agents, and temporal aspects such as the duration of outbreaks. By investigating these variables, we aim to identify patterns, trends, and factors influencing the occurrence and resolution of outbreaks. The project will contribute to a deeper understanding of outbreak dynamics and guide improvements in response strategies for better public health outcomes.

3 Results

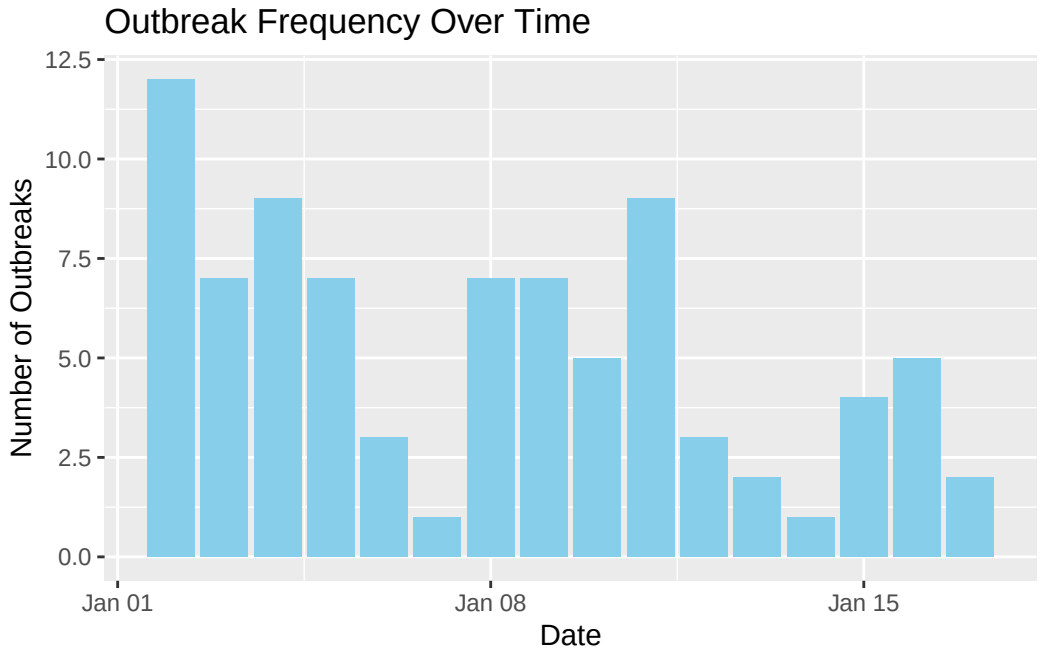
```
# Summary statistics
summary(Disease_data)
```

```
     X_id       Institution.Name   Institution.Address Outbreak.Setting
Min.   : 1.00   Length:84          Length:84           Length:84
1st Qu.:21.75   Class :character   Class :character    Class :character
Median :42.50   Mode  :character   Mode  :character    Mode  :character
Mean   :42.50
3rd Qu.:63.25
Max.   :84.00

Type.of.Outbreak   Causative.Agent.1  Causative.Agent.2  Date.Outbreak.Began
Length:84          Length:84          Length:84          Min.   :2024-01-02
Class :character   Class :character   Class :character   1st Qu.:2024-01-04
Mode  :character   Mode  :character   Mode  :character   Median :2024-01-08
                                                         Mean   :2024-01-07
                                                         3rd Qu.:2024-01-11
                                                         Max.   :2024-01-17
```

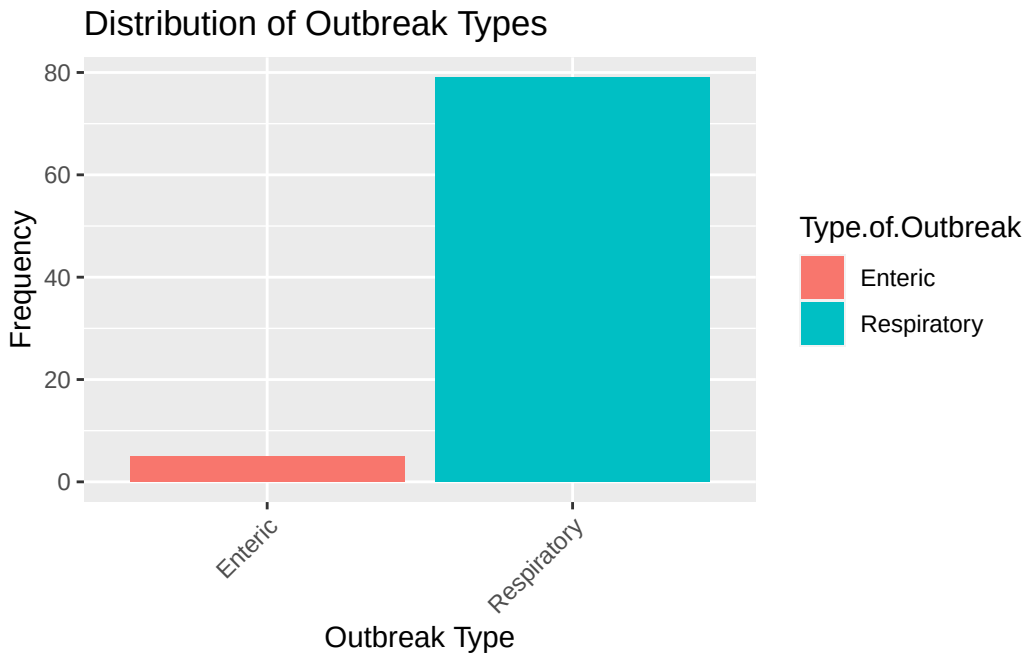
```
Date.Declared.Over      Active           Date_Outbreak_Began Date_Declared_Over
Min.   :2024-01-04   Length:84          Min.   :19724       Min.   :19726
1st Qu.:2024-01-10   Class :character   1st Qu.:19726       1st Qu.:19735
Median :2024-01-13   Mode  :character   Median :19730       Median :19735
Mean   :2024-01-12                      Mean   :19730       Mean   :19735
3rd Qu.:2024-01-16                      3rd Qu.:19733       3rd Qu.:19735
Max.   :2024-01-17                      Max.   :19739       Max.   :19739
NA's   :56
```

```r
# Exploratory Data Analysis (EDA)
# 1. Outbreak Frequency Over Time
ggplot(Disease_data, aes(x = Date.Outbreak.Began)) +
  geom_bar(stat = "count", fill = "skyblue") +
  labs(title = "Outbreak Frequency Over Time", x = "Date", y = "Number of Outbreaks")
```



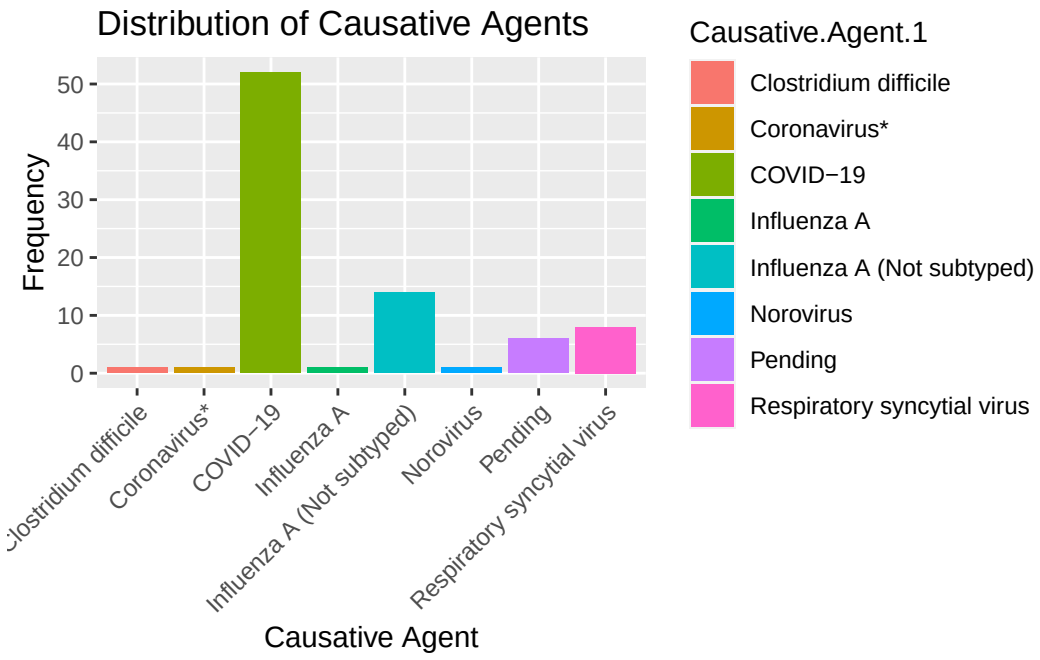

```r
##The time series plot reveals patterns in the frequency of outbreaks over time. Peaks or

# 2. Outbreak Types Distribution
ggplot(Disease_data, aes(x = Type.of.Outbreak, fill = Type.of.Outbreak)) +
  geom_bar() +
  labs(title = "Distribution of Outbreak Types", x = "Outbreak Type", y = "Frequency") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
##The bar plot illustrates the distribution of outbreak types, offering insights into the

# 3. Causative Agents Analysis
ggplot(Disease_data, aes(x = Causative.Agent.1, fill = Causative.Agent.1)) +
  geom_bar() +
  labs(title = "Distribution of Causative Agents", x = "Causative Agent", y = "Frequency") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
##The bar plot provides an overview of the distribution of causative agents for outbreaks.
```