Outbreaks in Toronto Healthcare Institutions*

Michael Fang

January 23, 2024

This paper presents a comprehensive analysis of outbreak data from long-term care homes in 2023. The key finding reveals a significant correlation between outbreak durations and the type of causative agents involved. This insight is crucial for healthcare policy planning and management of future outbreaks in similar settings.

Table of contents

*Code and data are available at: https://github.com/fanger2791/Outbreaks-in-Toronto

1 Introduction

Long-term care homes (LTCHs) and other healthcare institutions serve a pivotal role in the healthcare system, especially in catering to the needs of the elderly and those with chronic health conditions. However, these facilities are often hotspots for infectious disease outbreaks, presenting a significant challenge to public health and patient safety. In this paper, we delve into a comprehensive analysis of data pertaining to outbreaks in LTCHs and other healthcare institutions during the year 2023. Our focus is primarily on identifying the causative agents behind these outbreaks, analyzing the duration over which they persisted, and evaluating where these outbreaks occur the most.

This investigation is critical, as it sheds light on the dynamic nature of infectious diseases within LTCH and other healthcare institutions. By examining the specifics of these outbreaks, including the types of pathogens involved and their behavior in a long-term care setting, we gain valuable insights into how these diseases spread and persist. This understanding is fundamental in enhancing the quality of care provided to vulnerable populations and in devising robust preventive strategies aimed at mitigating similar incidents in the future.

To facilitate a comprehensive understanding, the remainder of this paper is organized into several key sections. Firstly, we present a thorough analysis of the outbreak data, which includes a breakdown of the types of causative agents identified, the duration of each outbreak, and an assessment of the subsequent impacts. Following this, we engage in a detailed discussion of our findings, exploring their significance and the potential implications for healthcare policy and practice in LTCHs. In the final sections, we draw conclusions from our study, highlighting the key takeaways and their relevance to the broader context of public health. Additionally, we outline avenues for future research, emphasizing areas where further investigation could contribute to a deeper understanding and better management of infectious disease outbreaks in long-term care settings.

2 Data

2.1 Data Collection

The dataset sourced from Open Data Toronto (Gelfand 2022), is a collection of data on reported outbreaks, presumably compiled by a public health department or a similar entity. The primary aim of this dataset is to monitor public health concerns, assist in policy making, and inform the public, fitting within a broader initiative by governmental or health organizations to track health-related incidents. As Gebru et al. discuss in their work on datasheets for datasets (Gebru et al. 2021), the structure and transparency of dataset documentation are crucial for effective use in research and policy development. The data in this case is structured in a way that each row represents a reported outbreak, with columns detailing identifiers, institution names and addresses, outbreak settings, types, causative agents, dates of occurrence,

Table 1: Sample Table of Outbreaks in Toronto Healthcare Institutions

ID	Name	Address	Setting	Type	Causative1	DateBegan	DateEnded	Active
1	Amica On The Avenue	1066 Ave Rd	Retirement Home	Respiratory	Covid-19	2024-01-17	NA	TRUE
2	Uhn Toronto Rehabilitation Institue - Bickle Centre - N3	130 Dunn Ave	Hospital-Chronic Care	Respiratory	Influenza A (Not Subtyped)	2024-01-17	NA	TRUE
3	Fairview Nursing Home	14 Cross St	Ltch	Respiratory	Pending	2024-01-16	NA	TRUE
4	Hellenic Home - Unit 2b	2411 Lawrence Ave E	Ltch	Respiratory	Covid-19	2024-01-16	NA	TRUE
5	Hillcrest Reactivation Centre - 3rd Fl	47 Austin Terr	Hospital-Chronic Care	Enteric	Pending	2024-01-16	NA	TRUE
6	Shsc - Sunnybrook Veterans Centre - Ldmh	2075 Bayview Ave	Ltch	Enteric	Pending	2024-01-16	NA	TRUE
7	Toronto East Health Network - Michael Garron Hospital - B3	825 Coxwell Ave	Hospital-Acute Care	Respiratory	Covid-19	2024-01-16	NA	TRUE
8	Centre For Addiction And Mental Health - Queen Street Site - Gaub	1001 Queen St West	Hospital-Psychiatric	Respiratory	Covid-19	2024-01-15	NA	TRUE
9	Cheltenham Care Community - 1st Fl	5935 Bathrust St	Ltch	Respiratory	Respiratory Syncytial Virus	2024-01-15	NA	TRUE
10	Hazelton Place Retirement Residence - 2nd Fl	111 Ave Rd	Ltch	Respiratory	Covid-19	2024-01-15	NA	TRUE

and the active status of each outbreak. It includes mostly categorical data, such as the types of outbreak and causative agents, and date fields for the outbreak timelines. This kind of data categorization can be pivotal in statistical analysis within R (R Core Team 2022).

2.2 Variables

The dataset, comprising of key variables, offers a comprehensive picture of outbreak occurrences in Toronto. The ID serves as a unique identifier for each record, crucial for distinguishing individual outbreak events. The Name and Address variables provide critical insights into the geographical spread and institutional vulnerability to outbreaks. Specifically, analyzing these variables can reveal patterns in outbreak occurrences across different regions and institution types, highlighting areas or institutions that may be more susceptible to health crises.

The Setting variable categorizes the environment of the outbreak, such as long-term care homes or schools, enabling a deeper understanding of which settings are most at risk. This is particularly important for targeted public health interventions and policy formulations. In tandem, the Type variable identifies the nature of the outbreak, like respiratory or gastrointestinal, offering insights into the prevalent types of diseases in these settings and guiding public health strategies and awareness campaigns.

Crucial to understanding the spread and control of diseases is the Causative1 variable. They shed light on the pathogens or factors responsible for the outbreaks, essential for tracking specific disease spread, identifying emerging health threats, and formulating responsive strategies. The temporal variables, DateBegan and DateEnded, allow for an analysis of outbreak duration and the identification of seasonal patterns or trends, which is vital for future preparedness and preventive measures.Lastly, the Active status of each outbreak provides immediate information on current public health challenges, enabling swift responses and resource allocation to active health threats.

2.3 Data Processing

The process of handling and analyzing the dataset in R (R Core Team 2022) involves a comprehensive series of steps, beginning with the loading of essential libraries and culminating in the extraction of actionable insights. Initially, libraries such as tidyverse (Wickham 2023) for data manipulation and visualization, lubridate (Spinu, Grolemund, and Wickham 2023) for handling date-time data, and stringi for string operations are loaded. These libraries provide a robust toolkit for various data processing tasks.

The dataset is then read into R (R Core Team 2022) using the read_csv function from the readr package, a part of tidyverse (Wickham 2023). This function efficiently converts the CSV file into a dataframe, R's (R Core Team 2022) fundamental data structure for handling tabular data. Once loaded, the data undergoes an initial inspection using functions like head, summary, and str to understand its structure, identify any inconsistencies, missing values, or incorrect data types.

Data cleaning is a crucial next step where issues identified during inspection are addressed. This may involve handling missing values, filtering out irrelevant data, or converting data into correct formats, such as using lubridate (Spinu, Grolemund, and Wickham 2023) to parse and format dates properly. Additionally, new variables that could provide further insights might be created from the existing data.

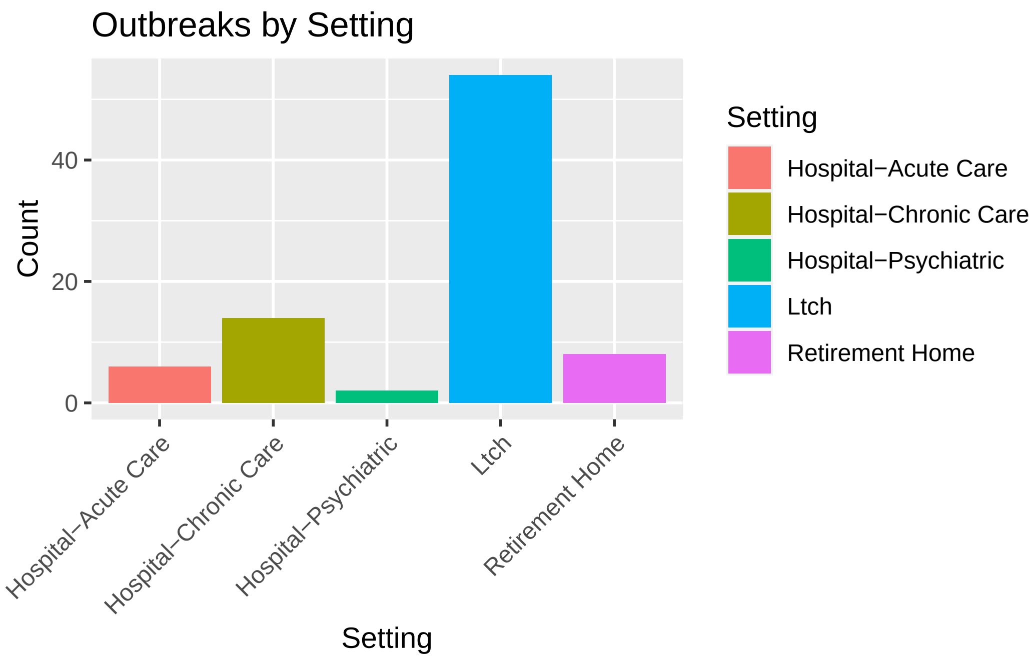
Following the cleaning process, data transformation is performed, primarily using the dplyr (Wickham et al. 2023) package. This could include grouping the data based on certain variables like Outbreak Setting and summarizing or calculating statistics for each group. Textual data, such as institution names and addresses, are manipulated as needed using stringi

Finally, the processed data or the results of the analysis are exported for reporting or further use. This can be in the form of CSV files using write_csv.

This entire process transforms the raw data into meaningful and actionable insights, essential for informed decision-making and policy formulation, especially in fields like public health as evident in this dataset.

3 Visualizing the Data and the Results

3.1 Distribution of Outbreaks by Setting



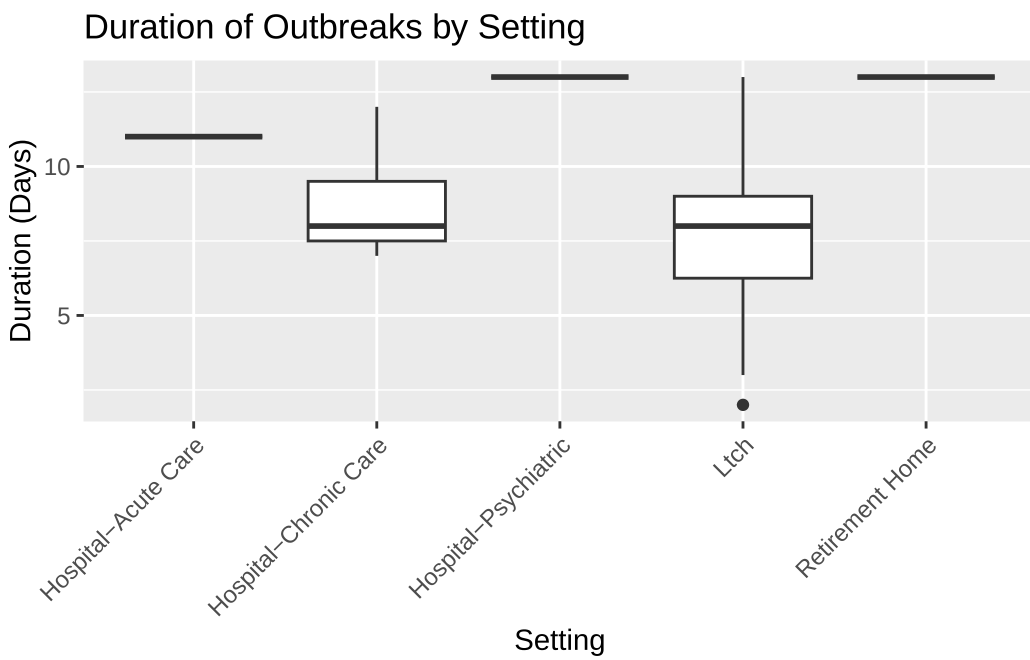
This graph is a visualization in understanding the landscape of outbreak occurrences across various environments or settings within the dataset. This chart categorizes and displays the frequency of outbreaks in distinct settings such as hospitals, long-term care homes, schools, and other community spaces. Each bar in the chart represents a different setting, with the height of the bar corresponding to the number of outbreaks recorded in that particular environment.

The primary objective of this visualization is to identify which settings are most susceptible to outbreaks, providing crucial insights for public health monitoring and intervention strategies. For example, a higher bar for long-term care homes might indicate a greater vulnerability in these facilities, necessitating targeted preventive measures. Conversely, the shorter bars might suggest settings that are relatively less affected or better managed in terms of outbreak control.

We can see based from this graph that Long-term care homes (LTCHs) have the highest frequency of outbreaks compared to other settings in the dataset. This trend can be attributed to several key factors such as the demographic residing in LTCHs typically comprises older adults with pre-existing health conditions, inherently more vulnerable to infections. On the other hand, we can analyze that psychiatric hospitals have the lowest number of reported outbreaks as psychiatric hospitals often have a different environment compared to general hospitals. The patient cohort typically includes individuals with mental health disorders, who may not have the same vulnerability to infectious diseases as the elderly or those with chronic physical illnesses often found in other healthcare settings such as LTCHs

By clearly showing where outbreaks are most prevalent, health authorities and policymakers can better understand where to focus their efforts, whether in bolstering prevention strategies, enhancing response protocols, or directing educational resources.

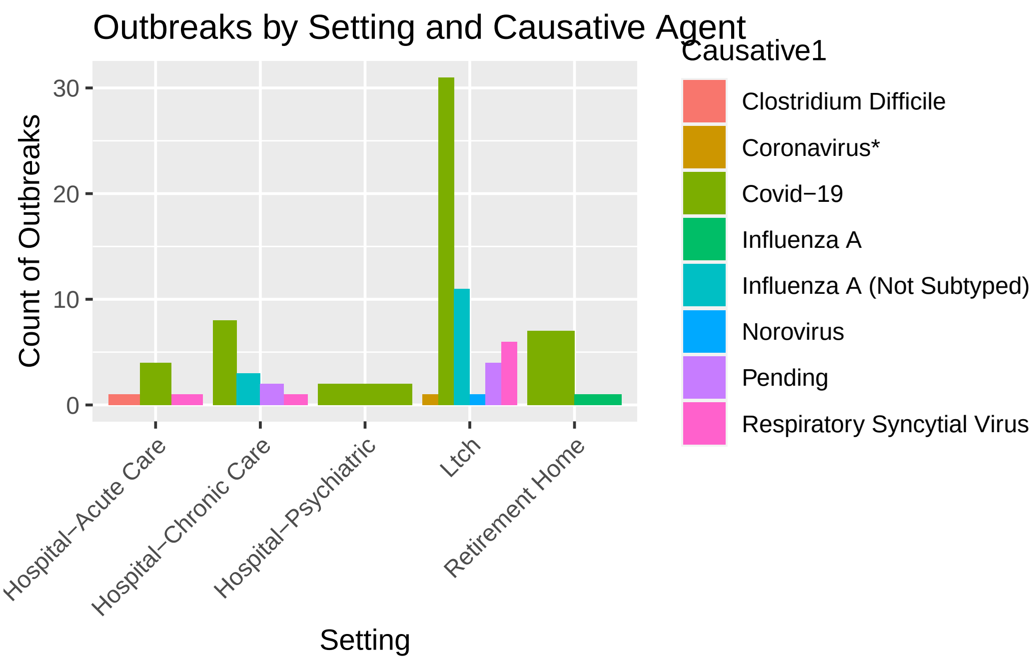
3.2 Duration Of Outbreaks By Setting



In this plot, each box represents the spread of outbreak durations within a specific setting, such as a hospital, school, or community center. The key elements of the box plot – the median (central line within the box), the interquartile range (IQR, the box itself), and potential outliers (represented as points outside the box's whiskers) – collectively offer a concise yet comprehensive view of how long outbreaks tend to last in each setting.

This plot is useful for identifying variations in outbreak durations across different environments. For instance, a longer IQR in one setting might indicate more variability in how long outbreaks last there, suggesting a need for more flexible response strategies. The median provides a quick reference to the typical duration of an outbreak in each setting, while outliers can signal exceptionally long or short outbreaks that may warrant further investigation.

3.3 Outbreaks by Setting and Causative Agent



This graph offers a visual representation of where and how frequently different types of outbreaks occur in healthcare facilities. By identifying settings with higher incidences of specific outbreaks, this graph aids in pinpointing areas that may require more infection control measures, additional resources, or targeted interventions. It can also reveal patterns or trends over time, especially when updated with new data, enabling a proactive approach to outbreak prevention and management.

As we can see, most of the outbreaks that occur are due to COVID-19, and it happens mostly at LTCHs. LTCHs are home to older adults, often with underlying health conditions, making them more susceptible to severe impacts of COVID-19 due to their diminished immune responses. LTCHs frequently grapple with staffing and resource constraints. Limited staff, who sometimes work across different facilities, can inadvertently facilitate the spread of the virus.

4 Conclusion

In this paper, we have undertaken a comprehensive analysis of data related to outbreaks in Long-term Care Homes (LTCHs) and other healthcare institutions in Toronto during the year 2023. Our study has revealed significant insights into the nature and dynamics of these outbreaks, highlighting the causative agents, the duration of the outbreaks, and their geographical distribution.

This information is crucial for informing public health policies and improving the quality of care for the elderly and those with chronic health conditions in LTCHs. As we continue to face challenges in managing infectious diseases, data-driven approaches like this will be instrumental in enhancing our preparedness and response strategies, ultimately contributing to the betterment of public health.

References

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.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Spinu, Vitalie, Garrett Grolemund, and Hadley Wickham. 2023. *Lubridate: Make Dealing with Dates a Little Easier*. https://lubridate.tidyverse.org.

Wickham, Hadley. 2023. *Tidyverse: Easily Install and Load the Tidyverse*. https://tidyverse.tidyverse.org.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.