Predicting Infection Outbreak Dynamics in Toronto Healthcare Institutions: A Bayesian Approach

Respiratory Outbreaks Are More Likely in Long-Term Care Homes and Retirement Homes, Driven by Seasonal Trends and Causative Agents Like COVID-19 and Influenza

Divya Gupta
December 3, 2024

Abstract

Identifying key trends in infection outbreaks in a city can help policymakers implement better reforms to support citizens and prevent outbreaks. This study uses cleaned datasets retrieved from Open Data Toronto to develop a model that can have insightful implications for the City of Toronto policymakers, especially regarding resource allocation for outbreak prevention. This study aims to help policymakers better understand the trends of infection outbreaks in Toronto,

so that this data can be used to identify problem areas in the city (such as specific times of year). This study examines the factors influencing healthcare outbreaks in Toronto healthcare institutions, such as: type of infection (respiratory vs. gastroenteric), causative agent of infection, setting and month of infection. This study uses Bayesian logistic regression to identify the factors associated with infectious outbreaks and predict the likelihood of a specific outbreak type depending on the aforementioned variables. The findings show that COVID-19 is a significant causative agent for respiratory infections, while norovirus is a causative agent for enteric outbreaks. Moreover, there is a significant increase in the number of outbreaks in winter months and in long term care homes.

1. Introduction

Toronto has multitudes of healthcare institutions such as hospitals, long-term care homes and retirement homes that often house hundreds of people at a time and can serve as places of community for vulnerable population groups. Any healthcare-associated infectious outbreaks in such spaces pose substantial risk for all vulnerable populations in the space. Understanding the varying factors that influence and further an outbreak - such as its causes, duration, and type of outbreak - can help better understand the nature of the outbreak and plan interventions and outbreak prevention measures accordingly. Furthermore, analyzing data about infection outbreak trends in Toronto can help us predict the time, region and duration of the next outbreak occurrence which can be instrumental in helping vulnerable populations such as older people and children. This study aims to answer the research question: *How can Bayesian models help predict infection outbreak behavior, specifically based on the outbreak's healthcare setting, causative agent and month of outbreak?*

The estimand for this analysis is the probability of an infection outbreak being of a specific type (namely respiratory or enteric) as determined by the causative agent, month of outbreak, or healthcare setting.

Developing a robust Bayesian model that can predict outbreaks in the city and develop outbreak management strategies in real time, ensuring better resource allocation and healthcare management. I will be employing Bayesian Logistic Regression to predict the likelihood of respiratory versus enteric outbreaks and how they are influenced based on the healthcare institution, causative agent, duration of previous outbreaks, and time of the year. I will then be employing Bayesian Survival Analysis to model the duration of outbreaks and identify defining factors specifically in prolonged cases (duration of infection > 10 days).

The remainder of the paper is structured as follows. Section 2 details the data, measurement, along with helpful visualisations to understand the data. Section 3 outlines the modeling approach, while Section 4 presents the results. Section 5 discusses implications, limitations, and future directions. This is followed by the Appendix.

2. Data

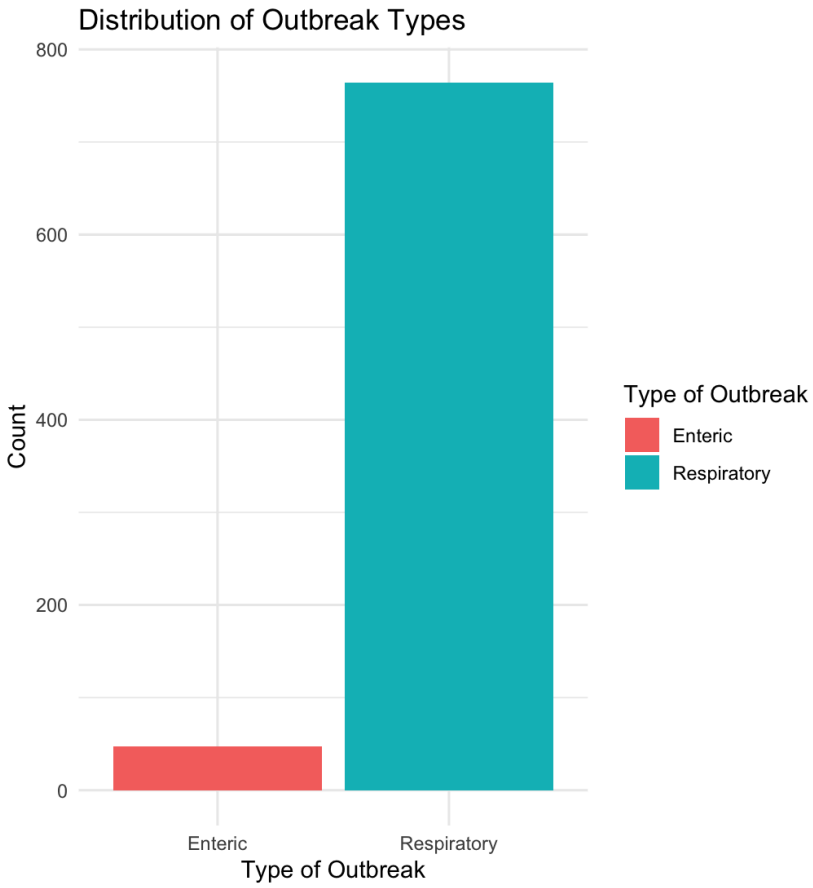
-

2.3 Measurement

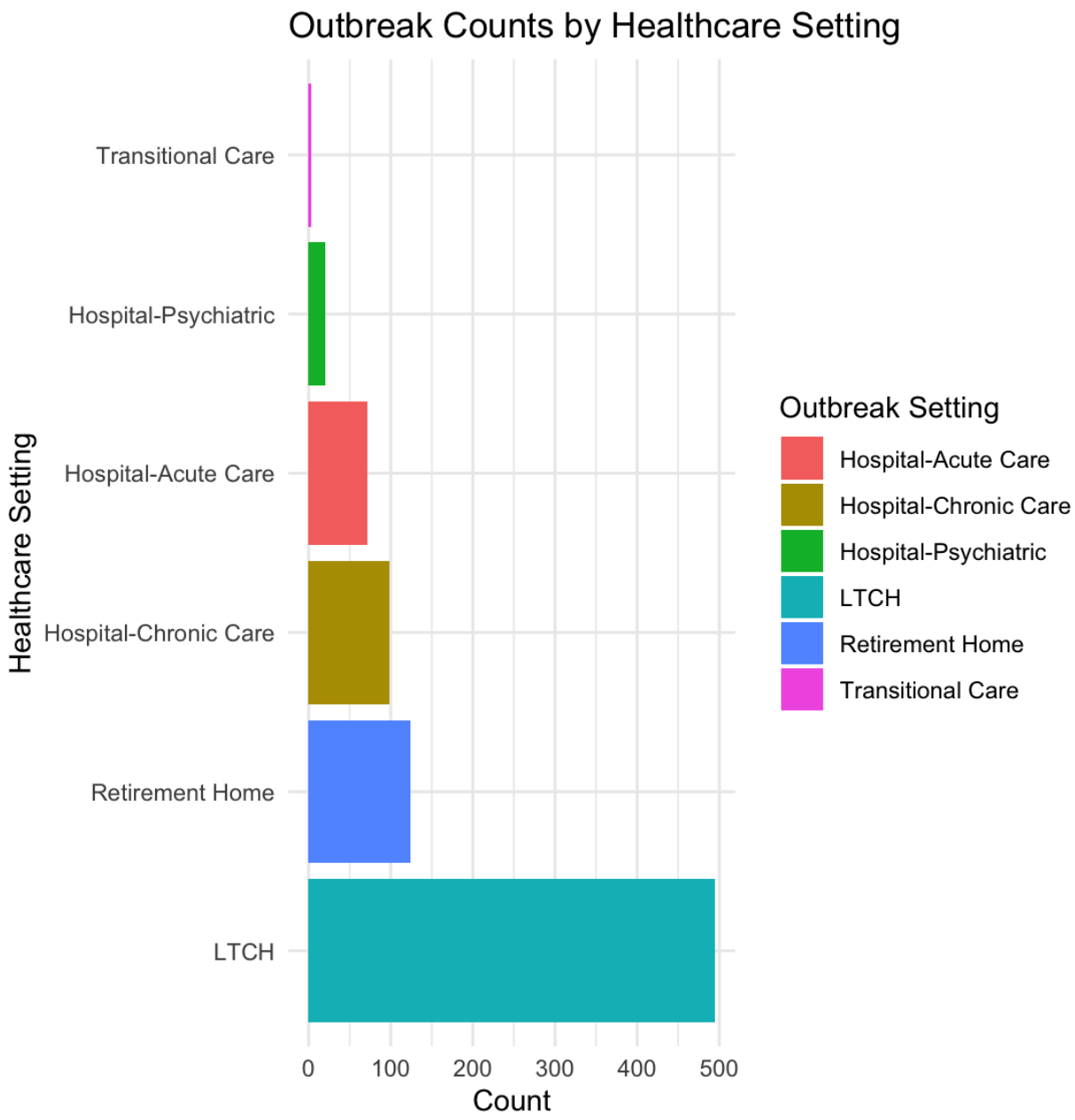
2.4 Exploratory Data Analysis

Initial analyses revealed:

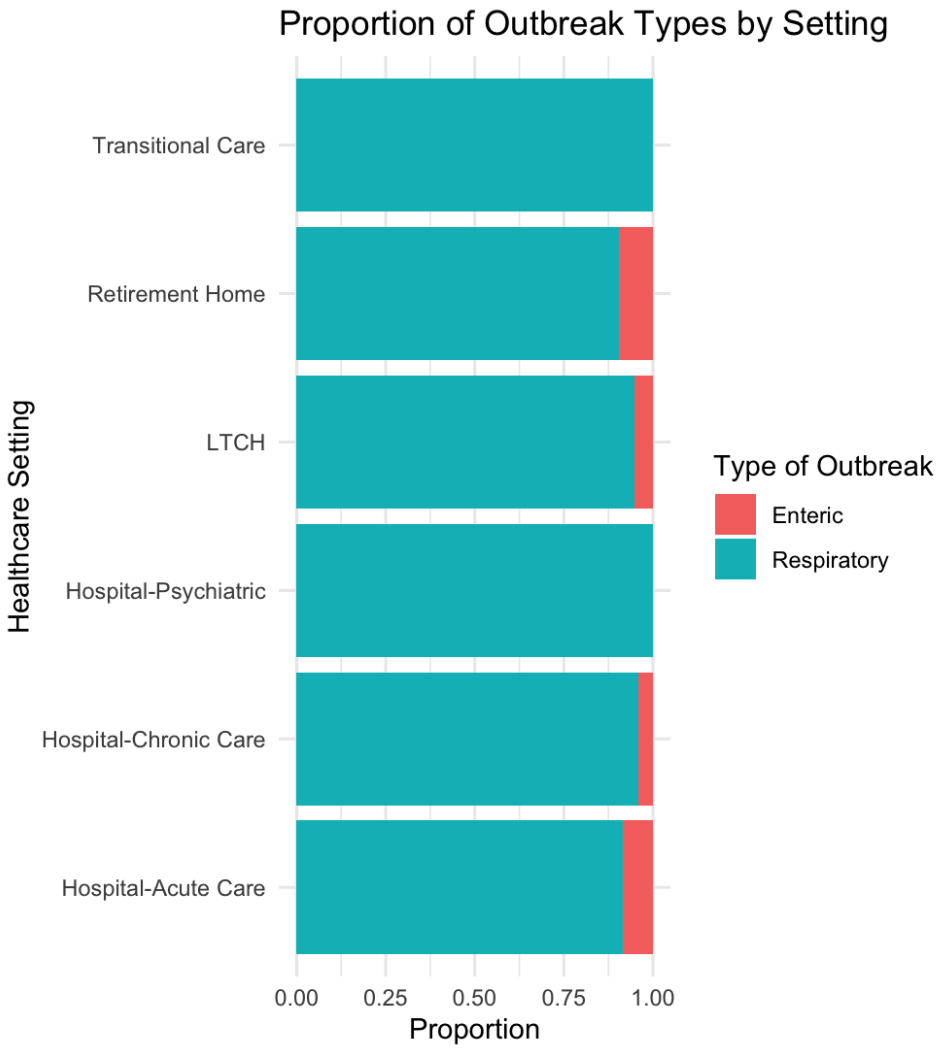
- Respiratory outbreaks were more common than enteric outbreaks. (Figure 1)



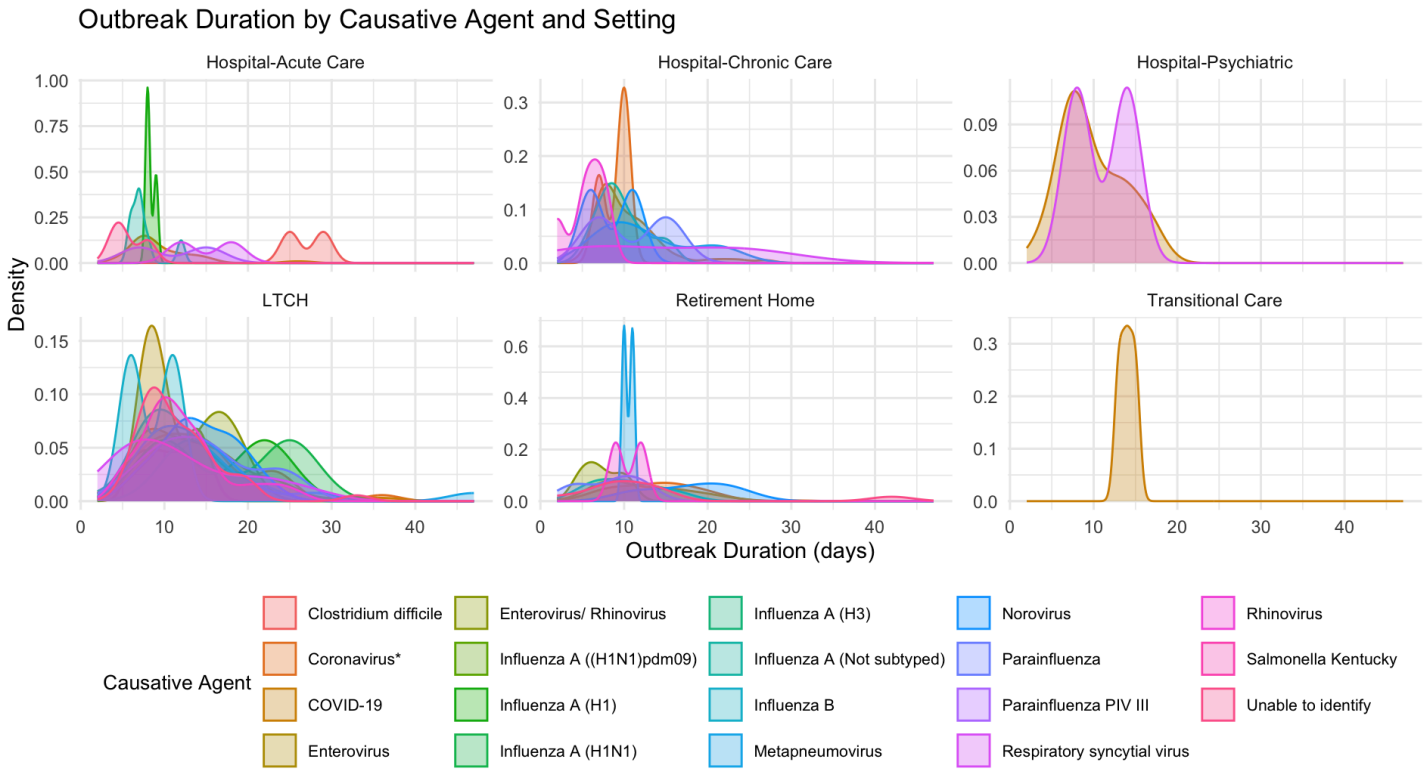
- Long-Term Care Homes experienced the highest number of outbreaks. (Figure 2)
- # Visual 3 - Interaction between Outbreak Type and Setting


Outbreak Counts by Healthcare Setting
Transitional Care
Hospital-Psychiatric
Hospital-Acute Care
Hospital-Chronic Care
Retirement Home
LTCH
Healthcare Setting
0
100
200
300
400
500
Count
Outbreak Setting
Hospital-Acute Care
Hospital-Chronic Care
Hospital-Psychiatric
LTCH
Retirement Home
Transitional Care

-



-
- Duration varied significantly by setting and causative agent. (Figure 4)



3. Model

3.1 Model Set-Up

I employed two Bayesian models:

1. **Logistic Regression**

yi ~Bernoulli(πi),logit(πi)=β0 +β1×Xsetting +β2 ×Xagent +β3 ×Xmonth

Priors:

(β ~ N(0, 2.5)) -> weakly informative normal priors that we can ensure the regularization of the model without imposing strong assumptions.

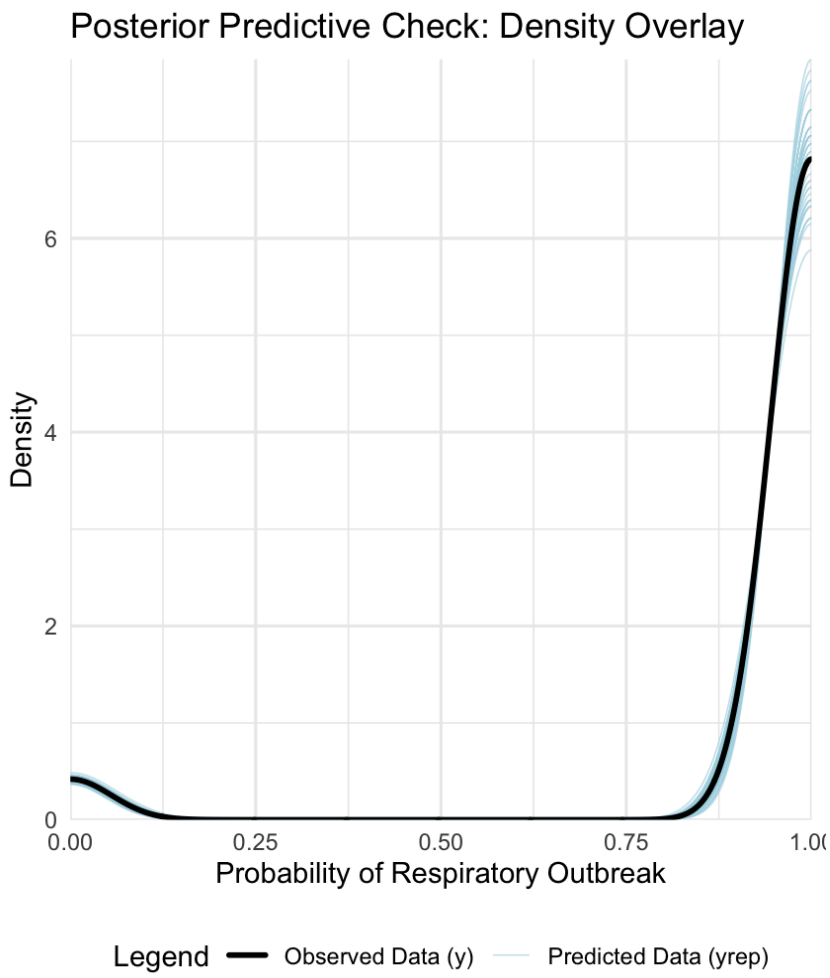
2. **Survival Analysis**

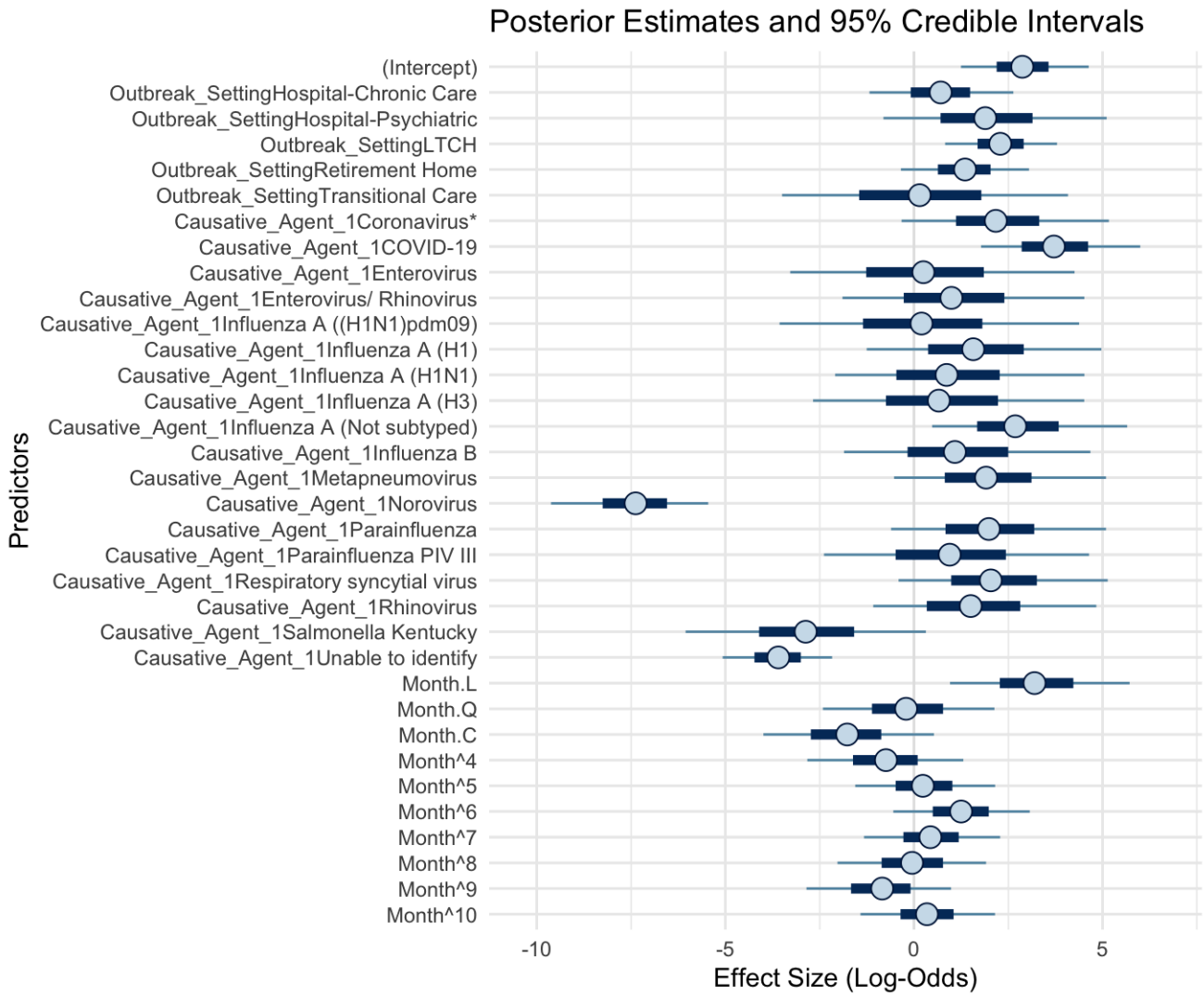
The models were implemented using the **rstanarm** and **brms** packages in R.

3.2 Model Validation

Validation is critical to ensure the models are reliable and provide accurate predictions. The following steps were undertaken to validate the models:

1. **Posterior Predictive Checks**:
    - Compared observed data distributions with data simulated from the posterior predictive distribution.
    - Ensured the models captured key patterns without overfitting.





2. **Cross-Validation**:
    - Employed leave-one-out cross-validation (LOO-CV) to compare models and assess predictive performance.
    - Metrics such as **WAIC** (Widely Applicable Information Criterion) and **LOOIC** (Leave-One-Out Information Criterion) provided a robust basis for model selection.
3. **Sensitivity Analysis**:
    - Tested the robustness of the models by varying priors and including interaction terms.
    - Assessed the impact of removing outliers and reanalyzing the data.
4. **Out-of-Sample Testing**:
    - Reserved a portion of the data for testing, comparing predicted probabilities and durations against actual outcomes.

3.3 Model Justification

The priors used in the model allow the integration of domain knowledge, such as typical outbreak durations or pathogen characteristics. Moreover, real-time updates to the model by acquiring new data will allow for it to continue being relevant to Toronto public health, making it a dynamic model that can help manage outbreaks.

4. Results

4.1 Logistic Regression

Key predictors of outbreak type included:

- Outbreak Setting: Hospitals were more likely to experience respiratory outbreaks.
- Causative Agent-1: Norovirus was strongly associated with enteric outbreaks.

Causative agent COVID-19 (β=3.8\beta = 3.8β=3.8) strongly increases the likelihood of respiratory outbreaks, while Norovirus (β=−7.4\beta = -7.4β=−7.4) is strongly associated with enteric outbreaks.

Long-Term Care Homes (LTCH) (β=2.3\beta = 2.3β=2.3) have a higher likelihood of respiratory outbreaks than hospitals.

Some month-related predictors (e.g., Month.L\text{Month.L}Month.L) show positive associations, suggesting seasonal trends.

Causative Agents Are Key Predictors:

- COVID-19 and other respiratory viruses (e.g., Influenza A) significantly increase the likelihood of respiratory outbreaks.
- Norovirus and Salmonella are strong predictors of enteric outbreaks.

Healthcare Settings Matter:

- Long-Term Care Homes and Retirement Homes are more prone to respiratory outbreaks compared to hospitals.
- This suggests differences in patient demographics, environmental factors, or infection control practices.

Seasonality Is Relevant:

- Month-related predictors indicate seasonal variation in respiratory outbreaks, with winter months likely contributing to higher probabilities.

Model Fit and Validity:

- The model fits well, with R^=1.0\hat{R} = 1.0R^=1.0 and high effective sample sizes. This suggests reliable and converged posterior distributions.

5. Discussion

5.1 Insights

COVID-19 remains a significant driver of respiratory outbreaks, highlighting the importance of ongoing prevention measures.
Norovirus's strong association with enteric outbreaks suggests a need for targeted hygiene protocols.

- Long-Term Care Homes and Retirement Homes could benefit from additional respiratory infection control measures, such as air quality monitoring and staff training.

5.2 Limitations

5.3 Future Directions

Appendix

A. Data Details

B. Model Details

References