Analyzing the Factors Influencing Outbreak Duration in Toronto Healthcare Institutions*

Kevin Cai

December 1, 2024

This study investigates the factors influencing the duration of infectious disease outbreaks in healthcare institutions in Toronto, focusing on the role of outbreak setting, causative agent, and seasonal variations. Using a Bayesian negative binomial regression model, the analysis identifies significant drivers of outbreak duration and highlights the importance of targeted preparedness in settings like Long-Term Care Homes (LTCHs) and Retirement Homes. The findings suggest that COVID-19 has a substantial impact on outbreak duration compared to other respiratory viruses. This research underscores the need for robust healthcare infrastructure and adaptive strategies to manage future healthcare outbreaks effectively.

Table of contents

*Code and data are available at: https://github.com/kevicai/toronto-healthcare-outbreak-prediction.

1 Introduction

Infectious disease outbreaks in healthcare settings present significant challenges, particularly in hospitals, long-term care homes, and retirement homes, where vulnerable populations are at highest risk (Toronto Public Health). These environments house individuals who are more susceptible to infections, including the elderly, individuals with compromised immune systems, and those with underlying health conditions. In Toronto, public health authorities are required to respond to and manage outbreaks of gastroenteric and respiratory illnesses in these institutions (Toronto Public Health). Efficient management of such outbreaks is essential to prevent widespread illness and ensure the safety of patients, staff, and visitors. Understanding the factors that influence the duration of these outbreaks is important for enhancing preparedness, minimizing response times, and optimizing resource allocation.

This study estimates the duration of infectious disease outbreaks in Toronto healthcare facilities. Specifically, it examines how factors such as the type of healthcare setting, the causative agent, and the month when the outbreak began influence the number of days an outbreak

lasts. By analyzing these relationships, the study evaluates how changes in these predictors affect outbreak duration.

The duration of an outbreak is influenced by various factors including the healthcare setting, the causative agent, and the timing of the outbreak. In nursing homes, for example, the close living conditions, shared spaces, and frequent interactions between residents, staff, and visitors contribute to the spread of infections. In hospitals, the high turnover of patients and healthcare workers, coupled with the need for intensive medical care, can also worsen the problem. Additionally, the characteristics of specific pathogens can influence transmission dynamics and outbreak duration. For instance, the concentration of SARS-CoV-2 in patient secretions peaks approximately 10 days after symptom onset, influencing transmission patterns and the length of outbreaks in healthcare facilities (McDonald). Understanding why some outbreaks persist longer than others requires identifying the drivers of these durations. While previous studies have examined factors affecting outbreak durations, more research is needed, especially within the context of Toronto's healthcare system. Toronto, with its large and diverse population, faces unique challenges in managing healthcare-associated outbreaks. To analyze these relationships, the study uses a Bayesian negative binomial regression model, suitable for count data and predicting the duration of outbreaks. The findings aim to inform strategies for better outbreak management, resource allocation, and preparedness in healthcare facilities.

The remainder of this paper is structured as follows: Section 2 covers the dataset used for analysis, Section 3 describes the model setup and methodology, Section 4 presents the results and their interpretation, and Section 5 discusses the findings and future research directions.

2 Data

2.1 Overview

To examine the effects of factors in healthcare facilities in Toronto on outbreak duration, this report uses the Outbreaks in Toronto Healthcare Institutions dataset, which contains data from January 2016 to November 2024. The dataset is provided by Toronto Public Health, through City of Toronto Open Data Portal (Toronto Public Health 2024). Since no other datasets specific to healthcare institutions in Toronto are available, no additional data sources were considered. The dataset tracks reported outbreaks of gastroenteric and respiratory illnesses in Toronto healthcare institutions and contains detailed information on outbreak settings, causative agents, and outbreak durations. Following the principles from Telling Stories with Data (Alexander 2024), we examine how the characteristics of outbreaks, such as the type of healthcare institution, the causative agent, and the month the outbreak began, influence their durations. A sample of the cleaned dataset is shown in Table 1.

Table 1: Sample of Cleaned Outbreaks in Toronto Healthcare Institution Data

Outbreak Setting	Causative Agent	Month	Outbreak Duration
LTCH	Influenza	Dec	20
Hospital-Acute Care	Norovirus	Dec	5
LTCH	Respiratory syncytial virus	Dec	14
LTCH	Metapneumovirus	Dec	21
Retirement Home	Influenza	Dec	21

There is 5387 observations in the orginal dataset and 1119 observations were removed that contained missing, invalid, or irrelivant data of the variables we're interested in. The data was first downloaded using `Python` (Van Rossum and Drake 2009) and cleaned with the `pandas` package (team 2020). The cleaning process involved converting dates to a standardized date-time format, creating a "duration" variable representing the length of each outbreak, and extracting the month of the outbreak's start. Irrelevant columns were removed, and variables were renamed for clarity. Causative agents were grouped into broader categories, and rows with missing or invalid data were removed, including those with unidentifiable causative agents or certain outbreak settings. The final dataset was saved for further analysis.

`R` (R Core Team 2023) is used for the generation of figures, graphs, and tables throughout this paper. Specifically, the `rstanarm` package (Goodrich et al. 2024) was employed to fit the model. For data manipulation, the `dplyr` package (Wickham et al. 2023) was utilized to clean and transform the data efficiently. The `caret` package (Kuhn and Max 2008) was used for model training, while `modelsummary` (Arel-Bundock 2022) was used to produce concise tables summarizing the model output. The `loo` package (Vehtari et al. 2024) was used to perform leave-one-out cross-validation, which helped assess the model's predictive performance. Finally, the package `ggplot2` is used to generate graphics and figures for this analysis. The starter code and the data analysis techniques used are from Telling Stories with Data (Alexander (2024)).

2.2 Measurement

The data was primarily collected through mandatory reporting by healthcare institutions to Toronto Public Health under the Ontario Health Protection and Promotion Act (HPPA). Reports of suspected or confirmed outbreaks include both gastroenteric and respiratory illnesses. These reports are based on active monitoring by institutional staff, who observe and document signs and symptoms such as nausea, vomiting, fever, cough, or sore throat.

Some details, such as the causative agent group, may initially be unconfirmed and later identified through laboratory tests or clinical evaluations. However, these identifications are not always definitive. For instance, "Coronavirus*" in the dataset refers to seasonal coronaviruses, which are commonly implicated in respiratory outbreaks, and does not include COVID-19.

The unit of measurement for outbreak duration is in days. Other data fields, such as outbreak setting and causative agent group, are categorical features without numerical units. The dataset is updated weekly, ensuring it reflects the most recent outbreak data available.

2.3 Outcome variable

2.3.1 Duration

The Duration variable is numerical and indicates the total number of days each outbreak lasted. This reflects the severity and magnitude of the outbreak. It is constructed from the dataset by calculating the difference between the outbreak start and end dates.

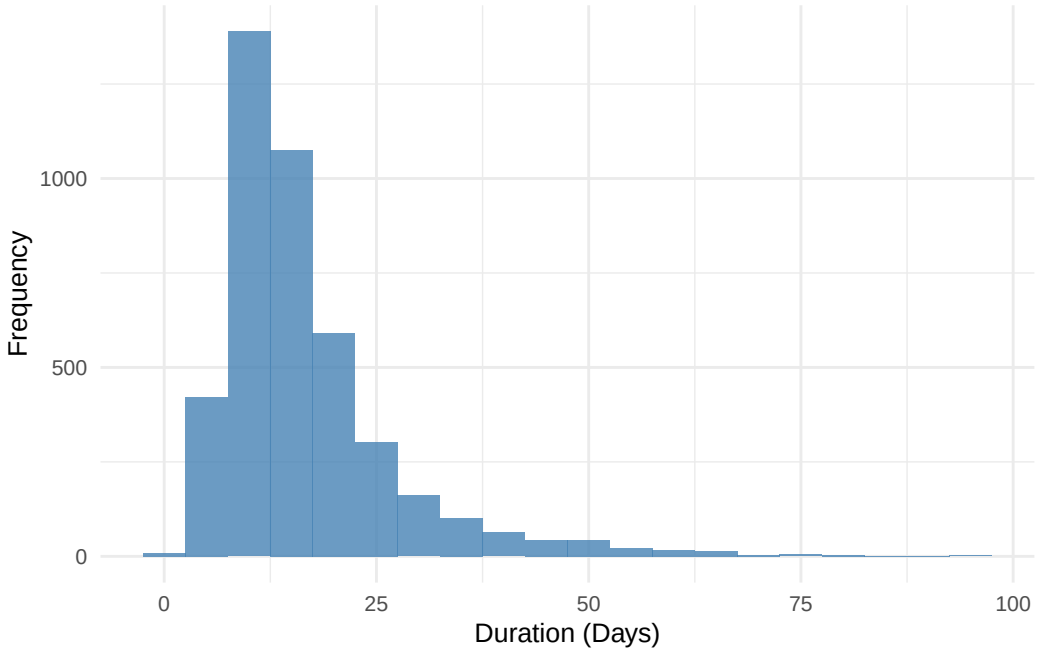


Figure 1: Distribution of Outbreak Duration

Table 2: Summary of Outbreak Duration: Mean and Variance

Statistic	Value
Mean Duration	16.57873
Variance	110.89162

Longer outbreak durations may indicate challenges in containment, possibly influenced by the Outbreak Setting and Causative Agent.

2.4 Predictor variables

2.4.1 Outbreak setting

The Outbreak Setting variable is categorical and identifies the type of healthcare institution where the outbreak occurred, such as hospitals, long-term care homes (LTCH), or retirement homes. It provides insights into the environments most affected by outbreaks.

Figure 2 illustrates the count of outbreaks across different settings in the dataset.

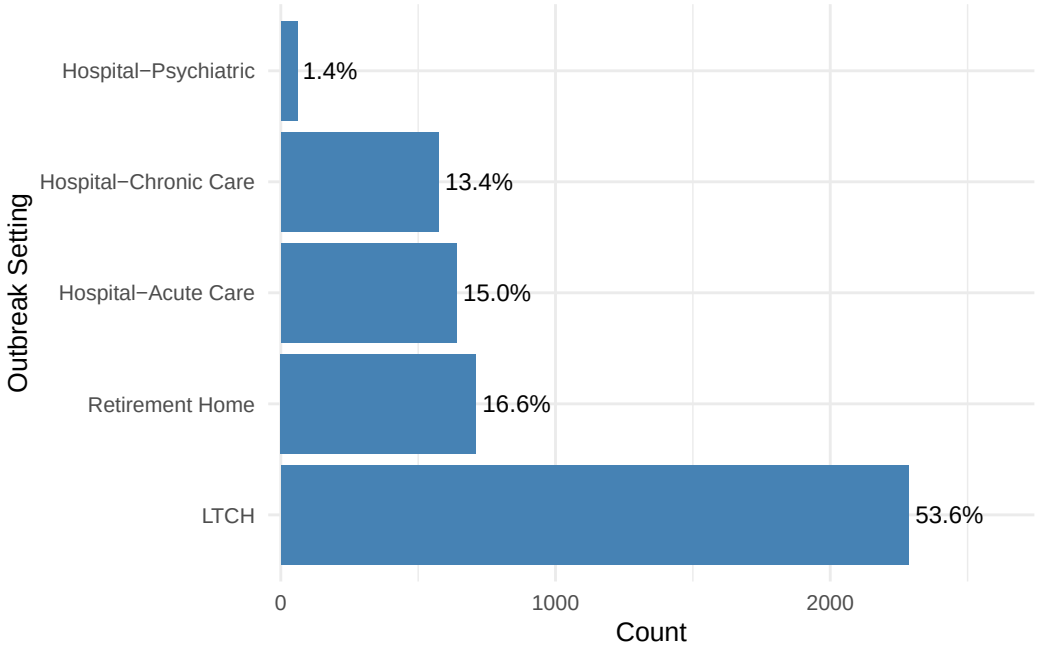


Figure 2: Outbreak occurrence in healthcare settings

LTCH (Long-Term Care Homes) accounts for a significant portion of outbreaks, likely due to the vulnerability of their populations. Comparing the frequency of outbreaks across settings can reveal risk patterns.

2.4.2 Causative agent

The Causative Agent variable is categorical and reflects the infectious agents responsible for outbreaks. While the original dataset contains 55 agents, they are grouped into seven broader categories to simplify the analysis and enhance interpretability.

Figure 3 illustrates the count and percentage distribution of causative agents in the dataset.

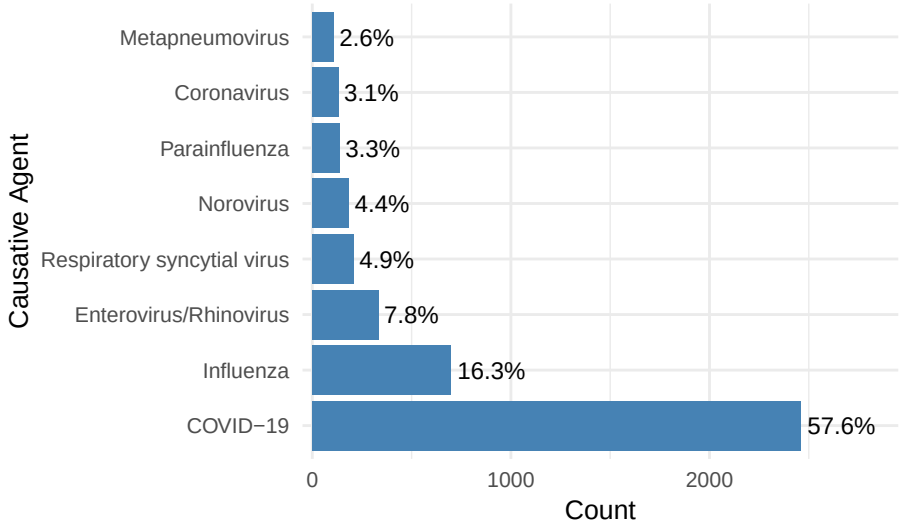


Figure 3: Outbreak causative agent count and percentage

2.4.3 Month outbreak began

The Month variable is categorical and records the calendar month when each outbreak started as a name (e.g., Jan, Feb). It reflects seasonal trends and potential patterns in infection rates. This variable is extracted from the date where each outbreak began from the original dataset and converted from a number to the corresponding month name.

Figure 4 shows the the occurance of outbreaks in each month, with winter months having siginifiantly more outbreaks compared to other months. This suggest that seasons have effects on outbreak occurances.

Figure 5 the boxplot visualizes the distribution of outbreak durations for each month. The duration of months January to November outbreaks appears similar, while December has a noticeable increase in duration compared to other months.

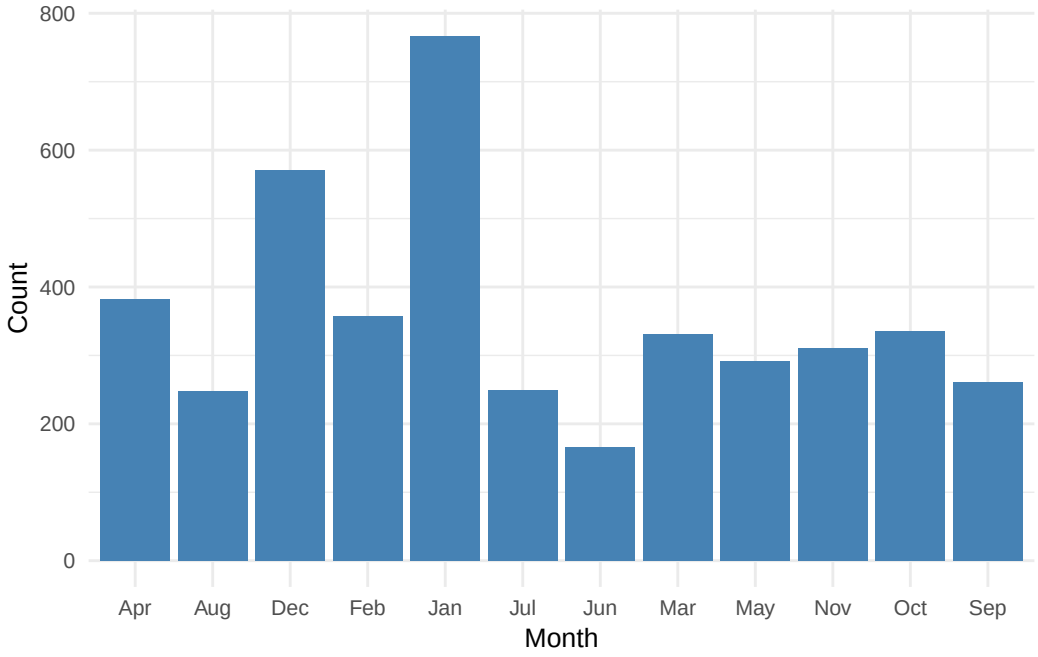


Figure 4: Seasonal trends in outbreak occurrence and percentage

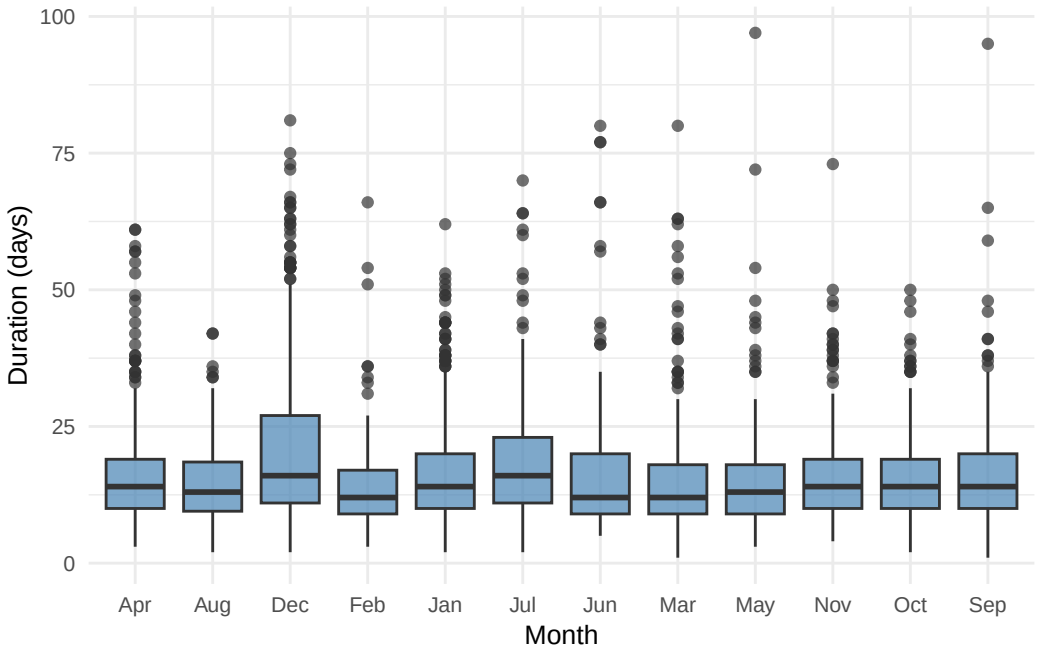


Figure 5: Duration of outbreaks across different months

3 Model

3.1 Model overview

To better understand the factors influencing the duration of outbreaks in Toronto healthcare facilities, a statistical model was developed using the negative binomial regression framework. This model was chosen because the outcome variable of interest, outbreak duration, is a count variable with evidence of overdispersion—where the variance exceeds the mean (Alexander 2024). Additionally, this model was Bayesian, meaning the parameters were treated as random variables with prior probability distributions reflecting initial beliefs about their values before considering the data.

3.2 Model setup

The setup for the Bayesian negative binomial regression model used in this analysis is as follows:

$$y_i|\lambda_i \sim \text{Negative Binomial}(\lambda_i, \phi) \tag{1}$$
$$\log(\lambda_i) = \beta_0 + \beta_1 \times \text{outbreak\_setting}_i + \beta_2 \times \text{causative\_agent}_i + \beta_3 \times \text{month}_i \tag{2}$$
$$\beta_0 \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta_1 \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\beta_2 \sim \text{Normal}(0, 2.5) \tag{5}$$
$$\beta_3 \sim \text{Normal}(0, 2.5) \tag{6}$$
$$\phi \sim \text{Exponential}(1) \tag{7}$$

In the above model:

- λ_i is the expected duration of outbreak i, modeled through a log link.
- β_0 is the intercept term.
- β_1 is the coefficient for the **outbreak setting**.
- β_2 is the coefficient for the **causative agent**.
- β_3 is the coefficient for the **month** when the outbreak started.
- ϕ is the **dispersion parameter** that controls the degree of overdispersion in the negative binomial distribution.

All coefficients ($\beta_0, \beta_1, \beta_2, \beta_3$) are assigned the prior of Normal(0, 2.5), which is the `rstanarm` package's default priors. The choice of the prior is sufficient for the model in this analysis and is a common non-informative prior that allows for reasonable variability (Alexander 2024). The dispersion parameter ϕ is assigned an **Exponential(1)** prior, which reflects the observation

that the variance is greater than the mean. Since we have categorical predictor variables, these priors allow the coefficients to adjust based on the data.

3.3 Model selection

Both negative binomial model and Poisson model for the dataset was constructed using the `rstanarm` package (Goodrich et al. 2024) and R (R Core Team 2023). But the negative binomial model was chosen over the Poisson model for several reasons. First, as shown in **?@tbl-modelresults**, the variance of the outcome variable, duration, is significantly higher than the mean, indicating overdispersion. The Poisson model assumes equal mean and variance, which is not suitable in this case. The negative binomial model relaxes this assumption, allowing for overdispersion and providing a better fit for the data (Alexander 2024).Additionally, the Leave-One-Out Cross Validation (LOO-CV) results in Table 4 show that the negative binomial model has a higher ELPD (Expected Log Pointwise Predictive Density) compared to the Poisson model. The ELPD is a metric that measures the model's predictive performance, with higher values indicating a better fit to the data (Alexander 2024). The fact that the negative binomial model outperforms the Poisson model in this regard suggests that it is more effective at capturing the underlying patterns of the outbreak duration data.

Other regression models like logistic regression were not chosen because logistic regression is designed for modeling binary outcomes. Since our outcome variable, duration, is a continuous count variable representing the number of days an outbreak lasts, logistic regression is not appropriate because it cannot model continuous or count data. Linear regression was also not chosen because Poission and negative binomial distributions are more suitable for modeling count data like outbreak duration in days, where as linear regression is more suitible for continuous data.

3.4 Model diagnostics and validation

We conducted several key validation checks to assess its predictive performance and overall adequacy. Aside from using LOO Cross Validation technique, we also calculated the Mean Absolute Error (MAE) for both models as a metric to assess the predictive performance of the Negative Binomial model over the Poisson model. To ensure the model doesn't over fit the training data, we first split the data into training and test sets. The data was randomly divided using the `caret` package (Kuhn and Max 2008), with 80% used for model training and the remaining 20% reserved for testing. We used both models to predict the outcome variable (outbreak duration) on the test set and compared the predicted values to the actual values from the test set to compute the MAE for each model.

Table 3: Comparison of Mean Absolute Error (MAE) for Poisson and Negative Binomial Models

Model	MAE
Poisson Model	6.53
Negative Binomial Model	6.52

From Table 3, the MAE for the Poisson model is 6.53, while the MAE for the Negative Binomial model is 6.52. The difference between the MAEs is minimal, suggesting that both models perform similarly in terms of prediction accuracy. However, the Negative Binomial model may still be preferred as it accounts for overdispersion, which is more appropriate for count data. The MAE of 6.52 means that, on average, the predicted outbreak duration from the Negative Binomial model deviates from the actual duration by approximately 6.52 days. In other words, for any given outbreak in the test data, the model's prediction of the outbreak's duration is off by around 6.5 days, either overestimating or underestimating the actual duration.

In Figure 8, the MCMC algorithm is also used to check for potential issues with the model. The trace plot is constructed using a subset of categorical values from each predictor variable, as there are many values for each category. As seen in the trace plot in Figure 8 (a), the lines bounce around but remain horizontal, with a nice overlap between the chains. This indicates that the chains have effectively explored the posterior distribution (Alexander 2024). The Rhat plot evaluates whether the chains have converged to a common distribution. As seen in the Rhat plot in Figure 8 (b), the values are close to 1 and fall below 1.1, suggesting no problems with convergence (Alexander 2024). Therefore, the model appears to be properly converged, and we do not need to remove or modify predictors, adjust the priors, or re-run the model.

4 Results

Our results are summarized in Table 5 using the modelsummary package (Arel-Bundock 2022), where the model is tidied with the broom.mixed package (Bolker and Robinson 2024) and the output table is scaled down with the kableExtra package (Zhu 2024). In the summary the model automatically selects the first category of each factor based on alphabetical orders of the category names. The intercept represents that the baseline outbreak duration is 2.257 days when all predictors are at their reference levels. For the outbreak setting, the reference category is "Hospital-Acute Care", all other outbreak settings are compared to "Hospital-Acute Care" to evaluate their influence on outbreak duration. The reference category is "Coronavirus" for the causative agent, the reference category is "April" for the month. By comparing the results, we can observe how each variable influences the duration differently.

Based on the 90% credibility interval plot for the model in Figure 6, we can analyze the influence of each predictor on outbreak duration: Intercept: The intercept coefficient estimate lies

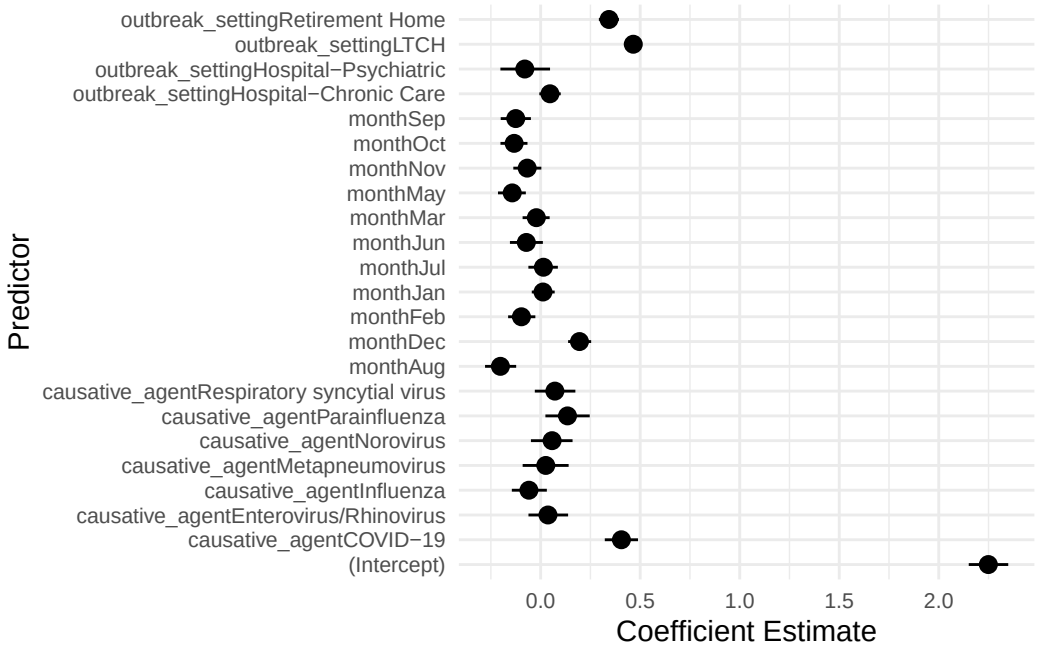


Figure 6: 90 percent credibility interval for coefficients

above 0, indicating that the baseline outbreak duration (when all predictors are at their reference levels) is positively associated with longer outbreak durations. The confidence intervals for most variables are relatively small, with the largest interval reaching around 0.25. A wider interval suggests greater uncertainty and higher variation in the estimate for that predictor. Conversely, narrower intervals indicate more precise estimates of the effect of the predictor on outbreak duration.

Based on Figure 6 and Table 5, we can see that for outbreak settings, the model suggests that outbreaks in Retirement Homes and Long-Term Care Homes (LTCHs) tend to last longer than other settings. This supports the expectation that these facilities, housing more vulnerable demographics, are more prone to prolonged outbreaks. On the other hand, Hospital-Psychiatric outbreaks tend to be shorter, as indicated by the negative coefficient. This supports the variability in duration across different healthcare settings. Regarding the causative agents, Respiratory Syncytial Virus (RSV), Parainfluenza, Norovirus, Enterovirus/Rhinovirus, and COVID-19, all show positive coefficient estimates, with COVID-19 being much larger compared to all other agents. This suggests that outbreaks caused by these agents last longer than those caused by Coronavirus (the reference group). In contrast, Influenza outbreaks are expected to last shorter, as indicated by its negative coefficient, while Metapneumovirus shows a marginal increase in duration, though its coefficient is very close to 0. Finally, the month of the outbreak also plays a role in its duration. The model suggests that outbreaks occurring in August and January are generally shorter compared to April (the reference month), while

December outbreaks tend to last longer, which may suggest seasonal variation in outbreak duration. Overall, the 90% credibility intervals for the coefficients highlight the significance of the predictors in influencing outbreak duration, with some factors such as the type of healthcare setting and causative agent having a more substantial effect, while others, like month of occurrence, show moderate influences.

5 Discussion

5.1 Long-term care homes and retirement homes are most susceptible to outbreaks

The type of healthcare setting where an outbreak occurs significantly impacts its duration. According to the model's results, outbreaks in Long-Term Care Homes (LTCHs) and Retirement Homes are more likely to last longer compared to other settings. This finding aligns with the vulnerability of populations in these settings, particularly the elderly, who may have weakened immune systems or multiple preexisting conditions. This observation is consistent with a population-based study of Retirement Homes (RHs) in Ontario, which reported that RHs are highly susceptible to SARS-CoV-2 outbreaks due to home-level and community-level factors. Larger resident capacities, co-location with Long-Term Care facilities, and offering multiple services onsite were identified as significant contributors to outbreak risk. These factors suggest increased staff movement and staff-resident interactions may exacerbate outbreak dynamics in these settings (Costa). Furthermore, the study observed that RHs in regions with higher community-level ethnic concentration or increased regional COVID-19 incidence were at elevated risk of outbreaks, although the sensitivity to these factors diminished in Wave 2 due to improved preventative measures, such as mandatory surveillance testing and stricter infection control practices (Costa).

The facilities of Long-Term Care Homes and Retirement Homes often lack the resources and infrastructure of hospital settings, which may contribute to longer outbreak durations in these environments. Previous research highlights staffing shortages in Long-Term Care Homes further worsen these challenges. The Long-Term Care workforce, has faced retention and recruitment issues for years. The COVID-19 pandemic intensified these problems, leading to burnout and workforce attrition despite wage increases and hiring incentives (Heiks). Additionally, limited training and career advancement opportunities for direct care aides further strain care quality. These staffing issues, combined with high operational costs, hinder the ability of LTCH facilities to manage outbreaks effectively (Heiks). Addressing these systemic challenges is critical to improving outbreak management in LTCH and Retirement Homes.

5.2 The Impact of COVID-19 vs regular Coronavirus

The analysis shows that COVID-19 had a significant impact on the duration of outbreaks in healthcare settings, especially when compared to other respiratory pathogens like the common coronavirus. COVID-19 outbreaks were found to last much longer than those caused by regular coronaviruses, a finding consistent with the global experience during the pandemic. Several factors contributed to this prolonged duration, including the high transmissibility of the virus, its asymptomatic spread, and the challenges of managing outbreaks in healthcare settings housing vulnerable populations (Costa). Additionally, the virus's evolving nature, with new variants emerging over time, further complicated efforts to contain it, requiring healthcare facilities to frequently adjust infection control measures.

While the model suggests that regular coronaviruses did not have a similar effect on outbreak duration, as seen in Table 5, the coefficient is close to that of other causative agents. This indicates that once the COVID-19 pandemic is under control, coronaviruses may not pose the same level of challenge to healthcare systems. The findings highlight the importance of ongoing preparedness for future global health crises, where the emergence of new infectious diseases or highly transmissible variants could once again stretch healthcare resources. Continued investment in healthcare infrastructure, surveillance, and rapid response capabilities will be essential to minimize the potential impact of future pandemics on healthcare facilities, ensuring they are well-equipped to manage outbreaks of varying durations.

5.3 The Effect of Winter on Outbreak Duration

The model suggests that outbreaks occurring in winter, particularly in December, tend to last longer compared to other months. This seasonal variation is likely due to higher rates of respiratory illnesses such as influenza and RSV during colder months, combined with holiday-related gatherings and staff shortages. The strain on healthcare facilities, particularly those housing vulnerable populations, can further extend outbreak durations. These findings highlight the need for targeted preparedness during winter, including increased staffing, early vaccinations, and robust infection control measures to minimize the impact of seasonal outbreaks.

5.4 Weaknesses and Next Steps

One limitation of this analysis is the reliance on available outbreak data, which may not fully capture all outbreaks in healthcare settings. For instance, some incidents might have gone unreported, or data may have been misclassified, which could introduce bias or reduce the generalizability of the findings. Additionally, the model does not account for interactions between different factors, such as the combined effects of outbreak setting and causative agent, which could provide more nuanced insights. The assumptions made in the negative binomial regression model also rely on the availability of complete and accurate data, which may not always be the case, particularly with retrospective datasets. Furthermore, although the Mean

Absolute Error (MAE) of 6.52 days is relatively small, it still represents a notable deviation from the actual outbreak duration, suggesting that further investigation and model refinement are needed to improve accuracy.

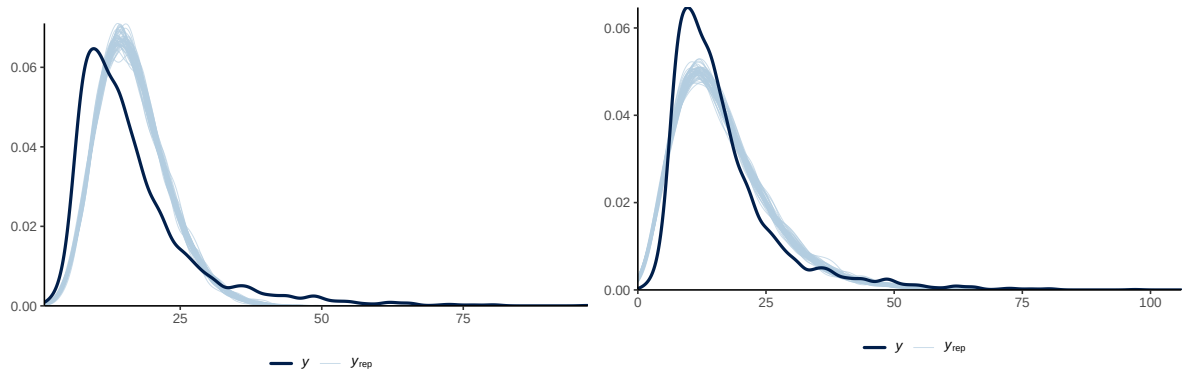
Future research could enhance this study by incorporating more detailed data, including healthcare interventions and patient information. Exploring the interaction between predictors like outbreak setting, causative agent, and seasonal trends could provide deeper insights into outbreak dynamics. The model also does not incorporate the potential influence of COVID-19 vaccinations on outbreak duration, an important factor that could have significantly altered the course of outbreaks, particularly in healthcare settings. Further research could explore the role of vaccination coverage and its interaction with other variables to improve predictive accuracy.

6 Appendix {sec-appendix}

6.1 Model details

6.1.1 Posterior predictive check

In Figure 7, using code adapted from Alexander (2024), posterior prediction checks were performed for both the Poisson model and the negative binomial model. The figure show how well the model is able to predict the observed outcomes.



(a) Poisson model posterior prediction check

(b) Negative binomial model posterior prediction check

Figure 7: Comparing posterior prediction checks for the Poisson model and the negative binomial model

Table 4: Comparing LOO for Poisson and negative binomial models

	elpd_diff	se_diff
neg_binomial_model	0.0	0.0
poisson_model	-3234.5	188.8

6.1.2 Leave-One-Out (LOO) Cross Validation (CV) Comparison

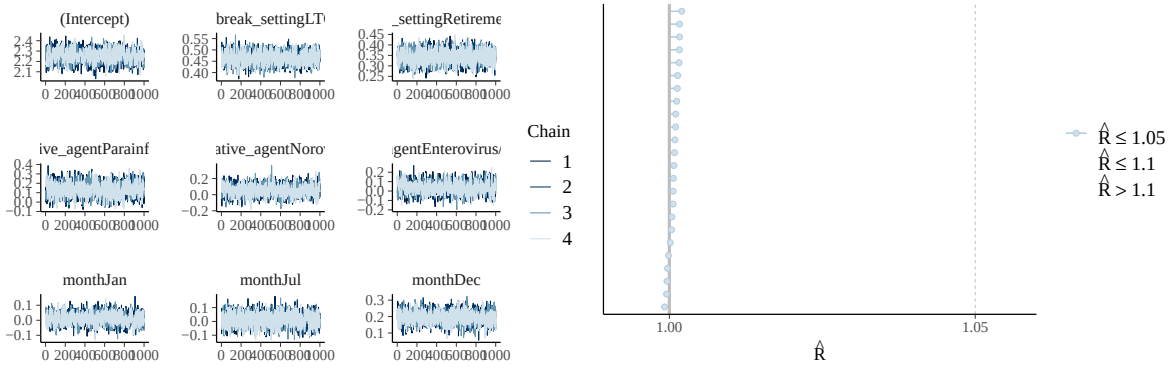
In Table 4, we compare LOO performance of the Poission model against the negative binomial model based on the expected log pointwise predictive density (ELPD) and find that the negative binomial model has a higher ELPD value.

6.1.3 Model summary

Table 5 presents a summary of the model used in the analysis, which includes the intercept and the coefficients for predictor variables, and the model fitting process.

6.1.4 Diagnostics

Figure 8 presents the diagnostic plots for the MCMC algorithm used to estimate the parameters of our model. These plots are essential for assessing the convergence of the sampling process and ensuring the reliability of the Bayesian estimates.



(a) Trace plot (b) Rhat plot

Figure 8: Checking the convergence of the MCMC algorithm

Table 5: Explanatory model of outbreak duration for Model (1), the negative binomial model

	(1)
(Intercept)	2.249
	(0.060)
outbreak_settingHospital-Chronic Care	0.047
	(0.033)
outbreak_settingHospital-Psychiatric	−0.080
	(0.075)
outbreak_settingLTCH	0.465
	(0.026)
outbreak_settingRetirement Home	0.343
	(0.031)
causative_agentCOVID-19	0.406
	(0.051)
causative_agentEnterovirus/Rhinovirus	0.037
	(0.058)
causative_agentInfluenza	−0.059
	(0.054)
causative_agentMetapneumovirus	0.026
	(0.070)
causative_agentNorovirus	0.057
	(0.063)
causative_agentParainfluenza	0.135
	(0.068)
causative_agentRespiratory syncytial virus	0.071
	(0.062)
monthAug	−0.202
	(0.047)
monthDec	0.195
	(0.035)
monthFeb	−0.097
	(0.042)
monthJan	0.012
	(0.035)
monthJul	0.014
	(0.046)
monthJun	−0.072
	(0.050)
monthMar	−0.022
	(0.040)
monthMay	−0.143
	(0.040)
monthNov	−0.068
	(0.041)
monthOct	−0.133
	(0.041)
monthSep	−0.125
	(0.047)
Num.Obs.	3416
algorithm	sampling
pss	4000

6.2 Idealized Survey

6.2.1 Introduction

Thank you for participating in this survey! We are collecting data from healthcare workers to better understand infectious disease outbreaks in healthcare institutions. Your responses will help inform strategies for improving outbreak management and preparedness.

The survey is expected to take approximately 10 minutes to complete. All responses will be kept confidential and used solely for research purposes.

If you have any questions, please contact:

- **Survey Coordinator**: Kevin Cai
- **Email**: kev.cai@mail.utoronto.ca

6.2.2 Survey Questions

6.2.2.1 Section 1: Institutional Information

1. **Institution Name**
   Open-ended response:
   _____
2. **Institution Address**
   Open-ended response:
   _____
3. **Outbreak Setting** (Select one):
   - Hospital - Acute Care
   - Hospital - Chronic Care
   - Hospital - Psychiatric
   - Long-Term Care Home (LTCH)
   - Retirement Home
   - Other: _____

6.2.2.2 Section 2: Outbreak Information

4. **Type of Outbreak** (Select one):
    - Respiratory
    - Enteric
    - Other: _____
5. **Primary Causative Agent** (Select one):
    - COVID-19
    - Influenza
    - Norovirus
    - Respiratory syncytial virus
    - Parainfluenza
    - Enterovirus/Rhinovirus
    - Metapneumovirus
    - Coronavirus (non-COVID)
    - Other: _____
6. **Secondary Causative Agent (if applicable)** (Select one or leave blank):
    - COVID-19
    - Influenza
    - Norovirus
    - Respiratory syncytial virus
    - Parainfluenza
    - Enterovirus/Rhinovirus
    - Metapneumovirus

- Coronavirus (non-COVID)
- Other: _____

6.2.2.3 Section 3: Timeline and Status

7. **Date Outbreak Began**
   Date picker: _____
8. **Date Outbreak Was Declared Over**
   Date picker: _____
9. **Is the Outbreak Currently Active?** (Select one):
   - Yes
   - No

6.2.2.4 Section 4: Outbreak Duration and Impact

10. **Duration of Outbreak (in days)**
    Open-ended response:
    _____
11. **Did the outbreak require external resources or assistance (e.g., public health interventions)?**
    - Yes
    - No
12. **What were the primary challenges faced during this outbreak?** (Select all that apply):
    - Staffing shortages
    - Supply chain issues (e.g., PPE, medications)
    - Limited isolation facilities
    - Insufficient training or protocols
    - Other: _____

13. **What mitigation measures were most effective in managing this outbreak?** (Select all that apply):

    - Enhanced cleaning and disinfection
    - Improved PPE availability
    - Isolation/quarantine measures
    - Staff education and training
    - Rapid diagnostic testing
    - Other: _____

6.2.2.5 Section 5: Feedback and Additional Information

14. **In your opinion, what could have been done differently to reduce the duration of this outbreak?**
    Open-ended response:
    _____

15. **Do you have additional comments or observations about this outbreak?**
    Open-ended response:
    _____

6.2.3 Thank You

Thank you for completing this survey! Your responses will contribute to improving outbreak management strategies in healthcare institutions.

If you would like to receive updates about this research or a copy of the final report, please provide your email address below (optional):
Email Address:

_____

References

Alexander, Rohan. 2024. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Bolker, Ben, and David Robinson. 2024. *Broom.mixed: Tidying Methods for Mixed Models*. https://CRAN.R-project.org/package=broom.mixed.

Costa, Andrew P. "Risk Factors for Outbreaks of COVID-19 at Retirement Homes in Canada – SHEA."

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2024. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Heiks, Cheryl. "Long Term Care and Skilled Nursing Facilities."

Kuhn, and Max. 2008. "Building Predictive Models in r Using the Caret Package." *Journal of Statistical Software* 28 (5): 1–26. https://doi.org/10.18637/jss.v028.i05.

McDonald, L Clifford. "SARS in Healthcare Facilities, Toronto and Taiwan — Pmc.ncbi.nlm.nih.gov."

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

team, The pandas development. 2020. "Pandas-Dev/Pandas: Pandas." Zenodo. https://doi.org/10.5281/zenodo.3509134.

Toronto Public Health. "Active Outbreaks in Toronto Healthcare Institutions."

———. 2024. *Outbreaks in Toronto Healthcare Institutions*. https://open.toronto.ca/dataset/outbreaks-in-toronto-healthcare-institutions/.

Van Rossum, Guido, and Fred L. Drake. 2009. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Vehtari, Aki, Jonah Gabry, Måns Magnusson, Yuling Yao, Paul-Christian Bürkner, Topi Paananen, and Andrew Gelman. 2024. "Loo: Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian Models." https://mc-stan.org/loo/.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.