ATQ: Alarm Time Quality, An Evaluation Metric For Assessing Timely Epidemic Detection Models Within A School Absenteeism-Based Surveillance System

by

Kayla Vanderkruk

A Thesis

presented to

The University of Guelph

In partial fulfilment of requirements

for the degree of

Master of Science

in

Mathematics & Statistics

Guelph, Ontario, Canada

ABSTRACT

ATQ: ALARM TIME QUALITY, AN EVALUATION METRIC FOR ASSESSING TIMELY EPIDEMIC DETECTION MODELS WITHIN A SCHOOL ABSENTEEISM-BASED SURVEILLANCE SYSTEM

Kayla Vanderkruk
University of Guelph, 2021

Advisor: Dr. Zeny Feng
Co-Advisor: Dr. Lorna Deeth

Model-based school absenteeism surveillance systems have been proposed to raise seasonal influenza epidemic alarms. Previous studies used metrics such as false alarm rate (FAR) and accumulated days delayed, for model evaluation and selection, however they were unable to optimize both alarm accuracy and timeliness. In this study, we developed a metric, alarm time quality (ATQ), that simultaneously evaluated both aspects by assessing alarms on a gradient, where alarms raised incrementally before or after an optimal time were informative, but penalized. Summary statistics of ATQ, average alarm time quality (AATQ) and first alarm time quality (FATQ), were used as model selection criterion. Alarms raised by ATQ and FAR-selected logistic regression models were compared. Daily school absenteeism and laboratory-confirmed influenza data collected by Wellington-Dufferin-Guelph Public Health was used for demonstration. A simulation study representative of Wellington-Dufferin-Guelph was conducted for further evaluation. ATQ-selected models were found to raise alarms that were timelier than the FAR-selected model.

Acknowledgements

Firstly, I would like to thank my co-advisors Dr. Zeny Feng and Dr. Lorna Deeth for their constant guidance and support throughout the past two years. I appreciate the time you invested in my growth and education. I would also like to thank Wellington-Dufferin-Guelph Public Health for providing the absenteeism and influenza data sets used in this study. With special thanks to Dr. Lise Trotz-Williams for helping me to understand the data sets provided and current surveillance processes within the region, and for the invaluable feedback. Additionally, I would like to thank Madeline Ward for sharing R code from their recent study.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Seasonal influenza produces a significant burden on the health care system. The World Health Organization (WHO) estimated that between 290,000 and 650,000 deaths globally are attributed to seasonal influenza each year (36). Since influenza is spread through close contact and fomites within the community, early detection at the local level can be conducive in reducing the impact of influenza through prompt implementation of mitigating strategies. Once local public health units are alerted of the upcoming seasonal influenza epidemic, they can communicate behavioural interventions such as frequent hand washing, covering coughs and sneezes, and staying home from work or school while ill to the community to reduce disease transmission (35). The timing of this communication is vital, as the greatest impact will be observed if they are communicated as early as possible.

Syndromic surveillance aims to identify disease outbreaks by using non-conventional indicators that capture behaviours that occur during symptom onset (8; 20), such as non-prescription drug sales (28; 31), ambulance dispatch data (6; 18; 2), emergency department data (4; 22; 9), and web queries (10; 5). Since the focus is on behaviour that occurs prior to laboratory diagnosis, syndromic surveillance systems have an increased sensitivity that can allow them to detect disease outbreaks earlier than monitoring reported cases (8; 20). This enables public health agencies to respond quickly, thus reducing the burden of morbidity and mortality. However, the high sensitivity of syndromic surveillance can be accompanied by a decreased specificity which can be cause for many false alarms (20). Thus, with a syndromic surveillance system, a practical balance between the number of false alarms and the timeliness of true alarms prior to the reporting of laboratory-confirmed cases is needed in order to be beneficial.

1.2 School Absenteeism Influenza Surveillance Models

Schools play an important role in the spread of influenza to the wider community. This is not only because schools are a channel for disease transmission (14), but also a result of elementary school aged children (5-15 years old) having the highest rates of influenza infection compared to all other age groups (7). Since children are likely to stay home when ill, monitoring school absenteeism as a surveillance model can provide early signs of community infection. Additionally, school absenteeism is significantly higher during the seasonal epidemic than compared to the non-epidemic winter season (19), indicating that school absenteeism and influenza activity is correlated (17).

1.3 Wellington-Dufferin-Guelph Influenza Surveillance

Wellington-Dufferin-Guelph (WDG) Public Health (WDGPH) in Ontario, Canada serves the counties of Wellington and Dufferin, as well as the City of Guelph. In 2016, the region had a population of 284,461, covering an area of 4,147 km^2 and resulting in a population density of 68.6 people/km^2(26). The region is predominantly rural, with the exception being the City of Guelph where roughly half of the region's population is located in the city's 87 km^2 boundary(25). The remaining region contains 15 municipalities: ten townships, four towns, and one village.

WDGPH began a school absenteeism-based surveillance program for influenza in 2008. This program relies on elementary and secondary schools within the region to report daily absenteeism to WDGPH using a online form. Since cause of absenteeism is often unavailable to reporting schools, all-cause absenteeism is reported and monitored. When a school reports absenteeism of 10% or more, WDGPH investigates to determine if the high absenteeism is illness related, and advises on mitigating measures if necessary. This 10% threshold method is similar to other school absenteeism-based surveillance programs (16; 13; 23; 15), however it has been found to be ineffective for some schools with high baseline absenteeism levels and provided inadequate lead time for public health officials to implement mitigation strategies (16). A recent study indicated that model-based approaches can improve alarm accuracy when compared to the 10% threshold method (32).

1.4 Evaluation Metrics

With an increasing interest in using statistical models for detecting outbreaks of infectious diseases, a method to evaluate surveillance models is needed. Measures such as sensitivity and specificity have commonly been used to evaluate the performance of surveillance systems, and can be combined into a single metric characterized by the area under the receiver operating characteristic (ROC) curve. However these measures do not capture the timeliness of an alarm, which is crucial for surveillance. Kleinman and Abrams extended the ROC curve to include timeliness (11), but this extension assumes the epidemic has already begun, and measures time between the beginning of the epidemic and the first alarm. However, our interest lies in producing an alarm before an epidemic would be declared based on reported laboratory-confirmed influenza cases, enabling timely interventions by public health agencies.

Two metrics proposed by Ward *et al.* separately evaluated the accuracy and timeliness of alarms raised prior to the start of an epidemic (32). However, it is hard to obtain a model that minimizes both metrics simultaneously, and therefore, one metric must be prioritized. This limits either the accuracy or timeliness of alarms produced by selected models. Therefore, due to the lack of appropriate metrics, our primary focus in this thesis is to develop a metric to evaluate the accuracy and timeliness of alarms raised prior to the start of an epidemic.

1.5 Thesis Layout

This thesis proposes a new metric to evaluate accuracy and timeliness of alarms raised by a school absenteeism surveillance prediction model, and uses it to aid model selection. Chapter 2 includes a manuscript to be submitted to *Epidemics*, a journal on infectious disease dynamics. This chapter describes the data sets provided by WDGPH, and the applied statistical models that will be used as a syndromic surveillance system. It introduces our metric that evaluates alarms raised by a surveillance model, and how this metric can be used as a model selection criterion. This chapter then describes our simulation study and its components, used for further investigation of our evaluation metric. Finally in this chapter, alarms raised by models selected by our proposed metric will be compared to that of alarms raised by current model selection criterion. Chapter 3 summarizes the findings of this thesis, and outlines future work relating to the production of local epidemic alarms for influenza surveillance based on school catchment areas.

Chapter 2

ATQ: Alarm time quality, an evaluation metric for assessing timely epidemic detection models within a school absenteeism-based surveillance system

2.1 Abstract

Background: Wellington-Dufferin-Guelph Public Health (WDGPH) has conducted an absenteeism-based influenza surveillance program in the WDG region of Ontario, Canada since 2008. A 10% absenteeism threshold for any given school within the region raises an alarm for implementation of mitigating measures. A recent study indicated that model-based alternatives provided improved alarms, however evaluation was primarily based on alarm accuracy, and failed to optimize timeliness. We developed a new metric, alarm time quality (ATQ), that simultaneously evaluated epidemic alarm accuracy and timeliness.

Methods: Our general approach assessed alarms on a gradient, where alarms raised incrementally before or after the optimal day were considered informative, but penalized. Summary statistics of ATQ, average alarm time quality (AATQ) and first alarm time quality (FATQ), were used for model evaluation and selection criterion. Alarms raised by ATQ and FAR-selected models were compared. Daily aggregated elementary school absenteeism and laboratory-confirmed influenza case data collected by WDGPH was used for demonstration. We also conducted a simulation study representative of WDG that was used for further evaluation of the proposed metric. Distributed lag logistic regression models were used as epidemic detection models.

Results: The FATQ-selected model raised acceptable first alarms most frequently, while the AATQ-selected model raised first alarms within the ideal range most frequently.

Conclusions: ATQ-selected models raised alarms that outperformed that of the previously proposed FAR for the region serviced by WDGPH.

2.2 Introduction

Seasonal influenza epidemics cause significant morbidity and mortality each year (18), with the duration and severity of influenza epidemics varying year-to-year. Early epidemic detection at the regional level can be conducive in reducing the impact of influenza by prompt implementation of mitigating strategies. Syndromic surveillance methods focus on behaviours occurring during symptom onset, such as non-prescription drug sales (19; 22), absenteeism from work or school, and web queries (6; 3). These methods have an increased sensitivity that can allow them to identify disease outbreaks, or the start of a seasonal epidemic, earlier than typical surveillance methods that monitor laboratory-reported cases which are often lagged (5; 12). Using syndromic surveillance methods make it possible for public health agencies to respond quickly, thus reducing the burden of morbidity and mortality.

Children typically have the highest influenza infection rates (4), and since children are encouraged to stay home when ill, monitoring school absenteeism as a surveillance model can provide early signs of community infection (23; 9). The school absenteeism-based surveillance program in Wellington-Dufferin-Guelph Public Health (WDGPH) began in 2008. In this program, schools within the Upper Grand District School Board (UGDSB) voluntarily report daily absenteeism to WDGPH using an online form. When student absenteeism surpasses 10% in a given school, WDGPH contacts the school to determine if the high absenteeism is illness related, and advises on mitigating measures if necessary. Other school absenteeism-based surveillance programs used similar absenteeism threshold methods (11; 9; 14; 10), however it was found to be ineffective for schools with high baseline absenteeism levels and provided inadequate lead time for public health officials to implement mitigation strategies (11). Model-based alternatives to the 10% threshold method have recently been studied, and were found to improve epidemic alarm accuracy (23). These models are found to be limited as they primarily optimized alarm accuracy using the false alarm rate. However, such criteria did not consider the timeliness of the alarms and thus could lead to an alarm raised too late to provide sufficient time for implementing interventions (23).

This paper proposes a novel metric, Alarm Time Quality (ATQ), to evaluate alarms raised by a school absenteeism surveillance prediction model. This metric considers and balances both the accuracy and timeliness of a raised alarm. Summary statistics of the ATQ serve as a model selection criterion when multiple statistical models are considered, and to determine tuning parameters of a given statistical model for raising an epidemic alarm. To assess the ATQ metric, school absenteeism and influenza data from WDGPH are used to fit models and compare alarms raised when using the ATQ and the False Alarm Rate (FAR) metric proposed by Ward *et al.*, as model selection criterion (23). Furthermore, these proposed surveillance methods have not been evaluated using a computational simulation study. One main reason for that is the lack of software or computer programming packages for simulating the absenteeism and influenza data that mimic what may be observed in a given community. Therefore, a computer-based simulation study was conducted to evaluate the proposed metric.

2.3 Methods

This section is composed of four parts: (1) description of WDGPH regional school absenteeism and influenza data, (2) statistical models for raising an alarm, (3) ATQ and model evaluation, and (4) simulation study. We start this section with a description of the data sets provided by WDGPH, and the applied statistical models that will be used as a syndromic surveillance system. Next we introduce our novel metric, ATQ, that is used to evaluate the performance of alarms raised through different statistical models for identifying the start of an epidemic. The ATQ metric and considered statistical models will be applied to the regional data collected by WDGPH. Additionally, the simulation study developed for generating annual region wide epidemics included in the study will be described. This simulated data set will be used to assess the performance of alarms raised by statistical models, when these models are selected by the proposed evaluation metrics.

2.3.1 Influenza and Absenteeism Data

Two data sets were provided by WDGPH for this study: (1) absenteeism data for schools within the UGDSB, and (2) laboratory-confirmed influenza cases within WDG. The absenteeism data is used in this study to raise alarms for the beginning of an influenza epidemic, whilst the influenza data is used to define the annual influenza epidemic start date. The two data sets work together to establish an epidemic detection model.

Influenza case and school absenteeism data provided by WDGPH was available from

January 2008 to June 2018. In Canada, a school year runs from early September to late June. The partial school year from January to June 2008 was omitted from this analysis as the influenza epidemic for this year could have begun prior to January. Additionally, as the data from the 2009-2010 school year reflected the H1N1 pandemic that occurred over that time period, this school year was excluded from the study. Therefore the study period included eight complete school years, from September 2008 to June 2018 with the 2009-2010 school year excluded. All data analyses and data visualizations were performed using R version 3.5 (13).

Absenteeism Data

Schools within the UGDSB were asked to report daily student absence data to WDGPH using an online form. The absenteeism data obtained from WDGPH contained information on school type (elementary or secondary), anonymized school catchment area identifier, school student population size, and the number of absent students each school day for the participating schools. Daily all-cause absenteeism percentages were calculated by dividing the number of absent students in a school by the school population size. Data was not collected on days students did not have to attend school, such as weekends, statutory and school board holidays, and professional activity (PA) days. A total of 88 elementary schools and 14 secondary schools reported to WDGPH during the study period. Because attendance reporting was voluntary, not all schools within the UGDSB region participated, and participating schools may not have consistently reported daily absences. In a previous study, models that exclusively used elementary school absenteeism data consistently raised influenza alarms with higher accuracy than those that used secondary school absenteeism, both exclusively or in combination with elementary school absenteeism (23). Therefore, the current analysis focused solely on elementary school absenteeism, and removed the secondary school data from the analysis. For the remainder of the study, the terms “school” and “schools” refer solely to elementary schools.

Since absenteeism was manually entered, it was prone to data entry errors such as mistypes. For example, a school with a population of 500 may have accidentally reported 5000 students on a given day. Thus, observations where an elementary school reported a population of fewer than 45 or greater than 820 students, the smallest and largest consistently reported elementary school population sizes respectively, were removed from the data set. Similarly, observations in which a school reported a daily absenteeism of 50% or more were removed for suspicion of data entry error. Additionally, if a school reported absenteeism on fewer than five days over the study period, they were removed. If a school had submitted duplicate reports on a given day, the maximum absenteeism percent for the day was used. Mean daily all-cause absenteeism percentage was calculated for each day using all reporting elementary schools.

Influenza data

The influenza data set provided by WDGPH detailed de-identified influenza cases reported to the public health unit during the study period. The influenza case information included the report date (the date that WDGPH was notified of the case), laboratory results classification, and age group. Laboratory-confirmed cases of all age groups were included in the study. Influenza cases were aggregated daily to provide a count of the confirmed cases, and a binary variable was created to indicate if at least one case occurred on a given day.

For this study, the reference date was defined as the date of the second laboratory-confirmed influenza case within a seven day period for the first time within an influenza season, where an influenza season was defined to begin annually on September 1. The reference day was used to indicate the start of the seasonal epidemic within the community, as opposed to the occasional cases that occur during an influenza season that might not necessarily be an indication of the start of the influenza epidemic.

2.3.2 Epidemic Detection Models

The current practice for WDGPH is to raise an alarm when school absenteeism surpasses 10% in any given school, as an indication of possible increased transmission of communicable diseases (including influenza) within the given school. Since this study focused on region-wide alarms, aggregated absenteeism data was used as opposed to individual school absenteeism data. Therefore, for the purposes of this study we considered a region-wide epidemic alarm to be raised when the mean daily absenteeism surpassed 10%.

Model-based alternatives have recently been investigated as a potential improvement upon the current 10% threshold method (23). In particular, distributed lag logistic regression models with fixed effects for seasonality terms and a random effect for school year consistently performed well in a previous study (23), lending itself to be the primary model of interest.

Following the notation of Ward *et al.* (23), the region-wide seasonal mixed effects logistic regression model was given by:

$$\begin{aligned}\text{logit}(\pi_{tj}) =& \beta_0 + \beta_1 x_{tj} + \beta_2 x_{(t-1)j} + ... + \beta_{l+1} x_{(t-l)j} \\ &+ \beta_{l+2} \sin_j(\frac{2\pi t^*}{T^*}) + \beta_{l+3} \cos_j(\frac{2\pi t^*}{T^*}) + \gamma_j,\end{aligned} \tag{2.1}$$

where the outcome of interest, π_{tj}, was the probability of at least one case occurring on day t in school year j. The main predictor variables, $x_{(t-i)j}$, were the lagged mean absenteeism percentages among elementary schools on the $t-i$th day with absenteeism data available before day of school year j, where lag time is from 0 up to l. The trigonometric functions captured the seasonality effect of influenza, where t^* represented the calendar day of the year on which x_{tj} was observed, and T^* equaled 365.25. The random effect for the jth school year was represented by γ_j, and was assumed to follow a $N(0, \tau^2)$ distribution, accounting for intracorrelation among daily absenteeism and influenza observations within a given school year, but varied over different school years.

Additional models were explored to include indicator covariates for the day of the week (DOW), Monday through Friday, to account for their effects on absenteeism. For all versions of the seasonal mixed effects models, lag values between 1 and 15 were considered. The `glmr` function from the `lme4` package in R was used to fit the aforementioned models (1; 13).

Data from the first full school year available was used for training, and therefore was not used in model evaluation. Each school year was evaluated using models that had been fit using data from all prior years and September of the current year. Data from October to the reference date of the year of interest was used to evaluate model performance. The deletion method was used to deal with missing school absenteeism values (23).

2.3.3 Evaluation Metrics

A model was considered to raise an alarm on day t of school year j if the predicted probability of at least one laboratory-confirmed influenza case, π_{tj}, was greater than a threshold θ. Various threshold values between 0.10 and 0.60 were tested, using 0.05 increments. The FAR and ATQ-based evaluation metrics were used to select an optimal threshold and lag value. All alarms raised after the reference date and before the beginning of the following school year were ignored.

Existing Evaluation Metrics

The study by Ward *et al.* proposed two metrics, false alarm rate (FAR) and accumulated days delayed (ADD), to evaluate the accuracy and timeliness, respectively, of alarms raised by a prediction model (23). Within these two metrics, a true alarm was defined as an alarm that was raised within the 15 calendar day period prior to and including the reference date (23). Alarms raised prior to this 15 day period are considered false alarms.

The FAR for year j was defined as:

$$FAR_j = \begin{cases} \frac{n_f}{n_f+1} & \text{if a true alarm was raised} \\ 1 & \text{if no true alarms were raised,} \end{cases} \tag{2.2}$$

where n_f was the number of false alarms raised in school year j (23). The FAR produced a minimum value of 0 if no false alarms and at least one true alarm was raised in a given year, and a maximum value of 1 if no true alarms were raised in a year. FAR values close to 0 indicated high alarm accuracy, and values of 1 or close to 1 indicated no true alarms, or large amounts of false alarms.

The ADD for year j was defined as:

$$ADD_j = \begin{cases} \tau_{optimal} - \tau_{observed} & \text{if a true alarm was raised} \\ \tau_{max} & \text{if no true alarms were raised,} \end{cases} \tag{2.3}$$

where $\tau_{optimal}$ was the optimal number of calendar days of advance notice before the reference date (in this case 14 days), and $\tau_{observed}$ was the number of calendar days before the reference date that the first true alarm is raised in year j (23). If a model raised no true alarms in year j, a large value was assigned, τ_{max}, representing the number of days between the first day of the school year absenteeism data was available and the reference date (23). The ADD produced its optimal value of 0 when a true alarm was raised 14 days prior to the reference date. For true alarms, the ADD increased in value as an alarm was raised closer to the reference date.

Both the FAR_j and ADD_j were calculated for each school year j, and then were averaged over all school years. Models that produced the smallest average FAR and ADD values were favoured, however it was hard to obtain a model that minimized both the FAR and ADD simultaneously. Due to this limitation, one metric was prioritized. In this case, model accuracy was selected, therefore the model that produced the smallest FAR was selected, and the ADD was used as a tie breaker (23).

The two metrics proposed by Ward *et al.* evaluated the accuracy and timeliness of alarms separately (23). As a result of this, models were selected based solely on the accuracy criterion (23), failing to optimize the timeliness of alarms. Additionally, Ward *et al.* defined true alarms to be strictly between the reference date and 14 days prior, and defined the optimal alarm to be exactly 14 days prior to the reference date (23). Consequently, an alarm raised one day earlier than the optimal alarm day is considered to be a false alarm, which makes the FAR and ADD metrics appear rigid in definition. These limitations motivated the formulation of a novel metric, Alarm Time Quality (ATQ), which provides a gradient approach to optimizing models based on alarm timeliness and accuracy.

Alarm Time Quality

Our proposed ATQ metric was based on the principles that: (1) an optimal alarm is raised 14 days prior to the reference date, (2) it is preferred that an alarm is raised during the ideal time interval of one to two weeks prior to the start of the influenza epidemic, and (3) alarms raised marginally before or after the optimal alarm time are informative but should be penalized in comparison to the optimal alarm time and the ideal time interval.

The ATQ for alarm i raised in year j was defined as:

$$ATQ_{ij} = \begin{cases} \left(\frac{14-\tau_{ij}}{21}\right)^4 & \text{if } \tau_{ij} \leq 14 \\ \left(\frac{14-\tau_{ij}}{21}\right)^2 & \text{if } 14 < \tau_{ij} \leq 35 \\ 1 & \text{if } \tau_{ij} > 35 \end{cases}, \qquad (2.4)$$

where τ_{ij} was the number of calendar days before the reference date that alarm i was raised in year j. In lieu of strict true and false alarm definitions, a power function was used in the

ATQ that was contingent on when an alarm was raised in relation to the optimal alarm. The power function penalizes alarms raised prior to the optimal day more than equidistant alarms raised after the optimal day. This provided a gradient scale that assessed the accuracy of each individual alarm raised by a model. The numerator of the ATQ is comparable to the ADD metric defined by Ward *et al.*, calculating the distance in days from the optimal alarm. Since the reference date was different every year, the maximum possible value of the numerator changed every year. Thus, we restricted the ATQ to be a value between 0 and 1, by normalizing on a three week (21 day) period from the optimal alarm day. Alarms raised prior to the three week period from the optimal alarm (35 days prior to the reference date), were considered too early for public health to effectively implement mitigation strategies, and were assigned the maximum value of 1. Thus the ATQ yielded its minimum value of 0 when an alarm was raised on the optimal day, and increased as an alarm was raised earlier or later than the optimal day up to a maximum value of 1. Visually, the ATQ embodied these principles by creating a valley shaped curve to assess an alarm (Figure 2.1), with the smallest values achieved for alarms raised 7-14 days prior to the reference date, and higher values observed as alarms were raised further from the optimal alarm time.

Additional values for the denominator and power functions in the ATQ were considered, to identify a definition of the ATQ that provided adequate penalization for suboptimal alarms. The denominator values of 14, 21, 30, and τ_{max} were examined, where τ_{max} is the number of days between September 1 of the given school year and the reference date. The denominator value of 14 was deemed to penalize alarms that occurred between the reference date and a few days prior too heavily. Conversely, denominator values of 30 and τ_{max} did not penalize alarms raised prior to the optimal alarm date enough. A balance was found with the choice of 21 for the denominator.

Similarly, the power function values of 2, 3, 4, and 5 for alarms raised after the optimal alarm date (i.e. $\tau_{ij} \leq 14$) were examined. Power values of 2 and 3 were found to increase the ATQ value too rapidly after the optimal alarm date, which penalized alarms within the ideal range too severely, whereas the power value of 5 did not sufficiently penalize alarms that occurred a few days prior to the reference date. The selected exponent value of 4 minimized the value of the ATQ during the ideal range of 7-14 days prior to the reference date, and sufficiently penalized alarms raised outside of that range. For alarms raised prior to the optimal alarm date (i.e. $14 < \tau_{ij} \leq 35$), power function values of 1, 1.5, 2, and 3 were examined. Powers of 1 and 1.5 appeared to penalize values too linearly, especially near the optimal alarm date, where a slow gradient was preferred for increasing the penalization. Conversely, the power value of 3 had a gradient that increased too slowly, which falsely suggested that alarms raised 21 days prior to the reference date were within the ideal range. Therefore, a power of 2 was selected because it had an ideal gradient.

Model Evaluation

The ATQ evaluates the quality of a single alarm raised. In order to evaluate how a statistical model performs while taking into consideration all alarms raised over several years, different summary statistics of ATQs were considered. We considered the average alarm time quality (AATQ), and first alarm time quality (FATQ), as well as their weighted counter parts. Each metric was used as model selection criterion, to choose model lag and threshold parameters.

Average Alarm Time Quality

The first method, AATQ, was the average ATQ values. However, due to the varying number of alarms raised each year, particularly years where no alarms were raised, the AATQ was not an average of all alarms raised, but rather an average of ATQ yearly averages. Therefore the AATQ was calculated in a three step process:

1. Calculate the ATQ value for every alarm raised using equation 2.4.

2. For each year, j, take the mean ATQ value to produce an $AATQ_j$ value for every year, assigning a value of 1 if no alarms were raised. That is:

$$AATQ_j = \begin{cases} \frac{\sum_{i=1}^{n_j} ATQ_{ij}}{n_j} & \text{if alarm raised in school year } j \\ 1 & \text{if no alarms were raised in school year } j, \end{cases} \tag{2.5}$$

where n_j is the number of alarms raised in year j.

3. Take the mean of all yearly $AATQ_j$ values, producing a final AATQ value for the model and its corresponding lag and threshold parameters. This is given by:

$$AATQ = \frac{\sum_{j=1}^{J} AATQ_j}{J}, \tag{2.6}$$

where J is the number of years evaluated.

First Alarm Time Quality

In practice, public health units would be likely to react and consider the implementation of behavioural intervention strategies when the first alarm is raised. Therefore it is crucial to determine the quality of the first raised alarm. Based on this, the FATQ was derived, which evaluated the first alarm raised in each year. Subsequent alarms raised after the first alarm do not affect the FATQ value. The FATQ was calculated for each model using the following steps:

1. For each year, j, calculate the ATQ value for the first alarm raised, denoted ATQ_{1j}, using equation 2.4. If no alarms were raised in a given year, assign that year a value

of 1. The $FATQ_j$ is therefore defined as:

$$FATQ_j = \begin{cases} ATQ_{1j} & \text{if alarm raised in school year } j \\ 1 & \text{if no alarms were raised in school year } j \end{cases} \tag{2.7}$$

2. Take the mean of all $FATQ_j$ values, producing the final FATQ value for the model and its corresponding lag and threshold parameters. This is expressed by:

$$FATQ = \frac{\sum\limits_{j=1}^{J} FATQ_j}{J}, \tag{2.8}$$

where J is the number of years evaluated.

Weighted AATQ and FATQ

In both of the AATQ and FATQ metrics, all prediction years were weighted equally. However since a prediction model to raise alarms for each year was trained using all data from preceding years, later school years would have larger training data sets to fit a prediction model (and thus result in better model fitting) than earlier years. Therefore, predictions in later school years could yield better predictive results than earlier school years' predictions. To account for this, weighted versions of the AATQ and FATQ were developed. The weight applied to each years' prediction was calculated based on the number of full years used in its training set, divided by the total number of training years in every prediction years' model. Therefore, the weight applied to year j predictions was defined as:

$$W_j = \frac{\text{Number of Training Years in Prediction Model}_j}{\sum_{i=1}^{n} \text{Number of Training Years in Prediction Model}_i}, \tag{2.9}$$

where j indexed the year, and n was the number of prediction years in the study. This weight was applied in steps 3 and 2 of the AATQ and FATQ algorithms respectively, to calculate the the Weighted Average Alarm Time Quality (WAATQ) and Weighted First Alarm Time Quality (WFATQ) metrics. The weighted metrics were given by:

$$WAATQ = \sum_{j=1}^{J} W_j AATQ_j, \tag{2.10}$$

and

$$WFATQ = \sum_{j=1}^{J} W_j FATQ_j \ . \tag{2.11}$$

For all evaluation metrics, the model that produced the smallest value was selected. The alarms raised by the models with the smallest values of the AATQ, FATQ, and their weighted counterparts were compared graphically to the alarms raised from the model selected using the FAR proposed by Ward *et al.* (23). Additionally, to aid the evaluation of timing of alarms over all years, alarms were categorized as follows: too late (alarm raised 0-3 days prior to the reference date), slightly late (alarm raised 4-6 days prior to the reference date), ideal (alarm raised 7-14 days prior to the reference date), slightly early (alarm raised 15-21 days prior to the reference date), and too early (alarm raised more than 21 days prior to the reference date). For the purposes of this analysis, an acceptable alarm was defined as an alarm raised between 4-21 days prior to the reference date, encompassing alarms that were categorized as slightly late, ideal, and slightly early. This range was considered to be acceptable as it provided sufficient time for public health officials to implement mitigation strategies prior to the start of the epidemic, while not being so early in the season that residents may not feel cause to follow recommendations.

2.3.4 Simulation Study

The purpose of the simulation study was to verify that ATQ-based metrics selected models that raised high quality alarms, and to compare alarms raised by ATQ and FAR-selected models. Our simulation model consisted of three main sequential parts: 1) a population of individuals was generated; 2) an influenza epidemic was simulated over the population for a given year; 3) a random probabilistic model was applied to the simulated population and epidemic to generate influenza and simulated school absenteeism data. The following sections detail each stage of the data simulation.

Population Simulation

To mimic a realistic population, our simulation model was developed based on demographics of WDG region in Ontario, Canada, and estimated parameters such as the number of elementary schools and school size. An 80x80 square was used to represent the WDG region, and was divided into 16 equally sized square subregions to allow for heterogeneous population densities across the region. Each subregion had a population size proportional to the number of schools within the subregion, and its student population sizes. To generate the number of elementary schools within each subregion, a Gamma(α, β) distribution was considered. The α and β parameters were estimated by fitting a gamma distribution to the number of elementary schools within each catchment area in the UGDSB, using the absenteeism data provided by WDGPH. The number of elementary schools within each subregion was randomly drawn from the resultant Gamma(4.313, 3.027) distribution and rounded to the nearest integer. Similarly, a Gamma(α, β) distribution was used to generate the student population sizes for each elementary school. Parameters were estimated by fitting a gamma distribution to the elementary school student population sizes using the absenteeism data set provided by WDGPH. The student population size of each elementary school was randomly drawn from the resultant Gamma(5.274, 0.014) distribution and rounded to the nearest integer. Elementary schools were randomly assigned to subregions, based on the number of schools within each subregion.

Since our study focused on elementary school absenteeism, our simulated population consisted of two subpopulations: (1) households with at least one elementary school aged child, (2) households without members of elementary school age. Given the simulated elementary school population sizes, the population was created using the demographics of WDG reported in the WDGPH's 2016 Census profile (17). The census profile provided information on household size, lone and coupled parents, households with and without children, the distribution of number of children, and age category distribution. Household structures were generated using this information, where we assumed that a household has a maximum size of five people, and a maximum of three children within a household. Since not all children within a household are elementary school aged, an approximation of the proportion of children that are elementary school aged was made based on the census age categories (17). Therefore the probability a simulated child within a household attended elementary school was proportional to the population under 20 years old, that was aged 5-14 years old. Subpopulation 1 was randomly assigned to an elementary school such that the children within the same household were assigned to the same school. Based on the elementary school assignment, households within subpopulation 1 were assigned to the schools' corresponding subregion. Subpopulation 2 was randomly assigned to subregions, such that a subregions' population was proportional to that of school population sizes within the given subregion. Locations of each household were generated by complete spatial randomness within the subregions' 20x20 boundaries. Due to the computational cost associated with epidemic simulation, approximately a quarter of the WDG population was simulated for illustrative purposes.

Epidemic Simulation

Individual level models (ILM), as outlined by Deardon *et al.* (2), have been used to model and simulate epidemic data. ILMs allow for individual level effects and can be used for spatial or contact-based infections. For this simulation study, a homogeneous spatial ILM was used within a susceptible, infectious, and removed (SIR) framework. Under the ILM framework, the probability of a susceptible individual i becoming infected by an infectious individual in the time interval $[t, t+1)$, $P(i,t)$, was given by:

$$P(i,t) = 1 - \exp(-\alpha \sum_{j \in I(t)} d_{ij}^{-\beta}) \qquad \alpha, \beta > 0, i \in S(t), \tag{2.12}$$

where $S(t)$ and $I(t)$ denoted the set of all susceptible and infectious individuals at time t respectively; α was the infectivity parameter for contracting the disease; $d_{ij}^{-\beta}$ quantified the risk of infection of individual i depending on the distance from infectious individual j, with d_{ij} being the Euclidean distance between individual i and infectious individual j, and β was the spatial parameter for the geometric rate of decay. Additionally, the infectious period was set to 4 days (8).

To simulate influenza epidemics in our population, the susceptibility factor and the spatial parameter were selected as $\alpha = 0.0019$, and $\beta = 3$ respectively. These values were chosen because they created a reasonable spread of infection across the population, in which the resulting infection would not spread too quickly and would typically infect 3-11% of the population (20). Epidemic curves were visually inspected to determine if an influenza epidemic grew and decayed reasonably, where unreasonable epidemics were discarded. The choice of parameter values depend on the population size, density, and the nature of the disease, thus the chosen parameter values are only suitable for this specific study. The `epidata` function in the `EpiILM` R package was used to simulate epidemics with a spatial ILM within a SIR frame work (24).

To mimic the fluctuating start of the annual influenza epidemic, the epidemic start time on which initial infection times were based, was drawn from a normal distribution with a mean and standard deviation of 45 and 15 days respectively. Without loss of generality, we set September 1 to be day 1 ($t = 1$), thus a mean of 45 days after day 1 would be October 16, which was similar to the mean of the occurrence time of the first influenza case observed in the WDGPH data. Start times of less than 20 days were reassigned to 20 days, since the influenza epidemic was unlikely to begin that early on in a year. Epidemics were initiated by randomly infecting two individuals from each of the 16 subregions, with a random infection time that is within 14 days of the epidemic start time. The maximum epidemic length was set to $t = 270$ to resemble the seasonal epidemic ending by the end of May, however since we were primarily interested in the start of the epidemic the length of the epidemic was not critical.

Laboratory Case Confirmation and Absenteeism System

Influenza and absenteeism data was generated based on probabilistic models, because not all infected individuals seek medical attention, and not all student absences are due to illness. We assumed that 2% of infected individuals received influenza laboratory confirmation based on the proportion of laboratory-confirmed influenza cases in WDG, under the assumption that 3-11% of the WDG 2016 census population was infected with influenza each year (20; 17). Since the infection period was 4 days long, there was a 0.5% chance per day that an infected individual would receive laboratory confirmation. Simulated laboratory-confirmed influenza cases were aggregated daily to provide a count of the confirmed cases, and a binary variable was created to indicate if at least one case occurred on a given day. Reference dates for each simulated epidemic were calculated as described in Section 2.3.1.

In this study, we only considered two scenarios in which a child is absent from school.

The first scenario was that a child did not have influenza and is absent due to other reasons. The second scenario was that a child had influenza and is absent. It required two different probabilistic models to simulate absent students under the two scenarios. Since students are unlikely to be absent due to influenza illness early in the school year, a baseline proportion of absenteeism under scenario 1 was estimated using absenteeism data provided by WDGPH from all September months. Our estimated baseline proportion of a student being absent from school given they are not infected with influenza was 2% per day. When a student was infected by influenza, we assumed that there was a 95% chance they were absent from school each day until recovered. Using the simulated data absenteeism counts, daily absenteeism percentage was recorded for each school. Mean daily all-cause absenteeism percentage across all schools was used to produce data sets for statistical modelling, as described in Section 2.3.1.

The simulation study was simpler than real epidemics. Therefore, a simpler statistical model without school year random effects and DOW indicators was considered for raising alarms in the simulated data set. A logistic regression model with lagged absenteeism and fixed seasonal terms was fit to the simulated data given by:

$$\begin{aligned}\text{logit}(\pi_{tj}) =&\beta_0 + \beta_1 x_{tj} + \beta_2 x_{(t-1)j} + ... + \beta_{l+1} x_{(t-l)j} \\ &+ \beta_{l+2} \sin_j(\frac{2\pi t^*}{T^*}) + \beta_{l+3} \cos_j(\frac{2\pi t^*}{T^*}).\end{aligned} \tag{2.13}$$

All covariates are as described in Section 2.3.2.

A total of ten replications, each consisting of ten annual epidemics, were simulated. For each replication, logistic regression models described by equation 2.13 were fit to the simulated data of each year, and lag and threshold parameters were selected for each replication based on the optimization of the evaluation metrics. Differences between the alarms raised by

ATQ and FAR-selected models were assessed across the ten replications, aided by graphical comparisons and alarm categorizations as described in Section 2.3.3.

2.4 Results

2.4.1 Preliminary Data Analysis of WDGPH Data

Average elementary school absenteeism and laboratory-confirmed influenza cases for WDG over the study period are shown in Figure 2.2. Among 88 elementary schools, on a given day there were between 0 to 40 elementary schools reporting absenteeism to WDGPH, with a median of 10 elementary schools. Throughout the study period, only nine elementary schools reported absenteeism on more than 50% of the available school days. In total, 1,697 out of 1,746 school days had recorded absenteeism data in the study period, based on an assumed 194 day school year. Across all study years, epidemic reference dates ranged from late October to late January, and epidemic reference dates most frequently occurred in December (Table 2.1). There was a mean daily all-cause absenteeism of 5.94% [95% CI = (5.58%, 6.31%)] for elementary schools (23). The Spearman correlation between laboratory-confirmed influenza case counts and mean elementary school absenteeism was 0.371 (23), and cross-correlation was highest when there was a six day lag between elementary school absenteeism and laboratory-confirmed influenza case counts (0.405) (23).

2.4.2 Comparison of Epidemic Alarms Raised by Evaluation Metric-Selected Models for WDGPH Data

When using the FAR, the selected model was a seasonal mixed model with lag time $l = 11$ and threshold $\theta = 0.25$ (Table 2.2). Figure 2.3 shows the timing of alarms raised by this FAR-selected model, as well as the ATQ-selected models, relative to daily absenteeism and influenza cases in WDG. The FAR-selected model raised alarms for six out of eight years (Figure 2.4). Within these six years, alarms were raised in an acceptable range in three years (2013-14, 2014-15, and 2016-17). For the remaining three years (2010-11, 2011-12, and 2017-18), alarms under this model were not raised until less than four days prior to the reference date, leaving little time for public health officials to respond in order to reduce the burden of influenza on the public.

The seasonal mixed model with lag time $l = 15$ and threshold $\theta = 0.15$ was selected by the AATQ, as well as the WAATQ (Table 2.2). This model raises alarms for seven out of eight years (Figure 2.4). Within these seven years, for four years (2010-11, 2013-14, 2014-15, and 2017-18), alarms were raised within the acceptable range, providing an adequate amount of time for public health officials to respond and promote community mitigation strategies. Additionally, the AATQ-selected model outperformed the FAR-selected model for the 2010-11, 2015-16, and 2017-16 school years, by raising more timely alarms that were closer to the optimal alarm day and raised an alarm in one year in which the the FAR-selected model failed to do so. The FAR-selected model outperformed the AATQ-selected model in 2013-14 and 2014-15 by producing an alarm closer to the optimal alarm day, however the AATQ-selected model still produced acceptable alarms within these years. There were two years in which the AATQ-selected model raised alarms too early (2011-12 and 2016-17). Of these two years, the FAR-selected model raised alarms within an ideal range in one year, but raised an alarm too late in the other year.

In comparison, when using the FATQ, the selected model was the seasonal mixed model with lag $l = 15$ and threshold $\theta = 0.25$ (Table 2.2). This model raised alarms for six out of eight years (Figure 2.5). In four of these years (2011-12, 2013-14, 2014-15, and 2016-17), the first alarm was raised within a week of the optimal alarm date, providing sufficient advance notice for public health officials to react. None of the alarms raised by the FATQ-selected model were too early. However, the FATQ-selected model raised alarms that were too late in two school years, 2010-11 and 2017-18. In comparison with the FAR-selected model, the FATQ-selected model improved the timing of alarms in the two years (2011-12 and 2017-18). The FATQ-selected model also provided the same amount of advance notice as the FAR-selected model in four years (2010-11, 2013-14, 2014-15, 2016-17), however the FATQ-selected model produced more alarms in two of those years, providing more confidence to public health officials of an upcoming epidemic. There were no years in which the FAR-selected model raised a more timely alarm than the FATQ-selected model.

When using the WFATQ metric, the selected model was the seasonal mixed DOW model with lag time $l = 11$ and threshold $\theta = 0.25$ (Table 2.2). Similar to that of the FATQ-selected model, the WFATQ-selected model raised alarms for six out of the eight years (Figure 2.6). In the last four school years, the WFATQ-selected model raised alarms with timing equal to or better than that of alarms raised by the FATQ-selected model.

For the 2012-13 school year, no alarms were raised by any of the selected models. The 2012-13 school year had the earliest epidemic reference date on October 26, 2012 (Table 2.1), and there were only two preceding school years available to train the model for raising an alarm for that year. The combination of an atypical epidemic reference date and limited data in which to train a predictive model, may be the reason no selected model was able to raise an alarm prior to the 2012-13 influenza epidemic.

An overall comparison of the timing of the first alarm raised by the FAR, AATQ, FATQ and 10% threshold approach is illustrated in Figure 2.7. The 10% absenteeism threshold method only raised a first alarm within the acceptable range for one school year. For the other seven years, the 10% absenteeism threshold either failed to raise an alarm or raised the first alarm too late or too early. Overall, the AATQ and FATQ-selected models raised the most first alarms that were within the acceptable range. While the FAR and FATQ-selected models tied for the most amount of first alarms that were raised within the ideal range. Alarms raised by the WAATQ and WFATQ-selected models were not included in this comparison, since by design they do not perform as well as the unweighted metrics for earlier school years but may outperform the unweighted metrics in the later school years.

2.4.3 Simulation Study Results

A comparison of the alarms raised under the FAR and AATQ-selected models for one simulated replication are displayed in Figure 2.8. For this particular simulated replication, the AATQ and FATQ metrics selected the same model, however this was not a consistent occurrence across all ten simulated replications.

In school years 4, 5, 6 and 10 of the illustrated replication, alarms raised by the AATQ-selected model were closer to the optimal alarm date than those raised by the FAR-selected model (Figure 2.8). However, in school year 3, the FAR-selected model raised a more optimal alarm than the AATQ-selected model. In school year 8, the AATQ-selected model raised its first alarm on the same day as the FAR-selected model, providing the same amount of advance notice of the upcoming epidemic. No alarms were raised by either of the selected models in two school years, and alarms were raised too early for both the FAR and AATQ-selected models in school year 2.

Overall trends of the simulation study found that the AATQ and FATQ-selected models raised timelier first alarms when compared to the FAR-selected model, more often. The alarms raised by the AATQ-selected model provided notice equal to that of the FAR-selected model nearly a third of the time, whereas the FATQ-selected model provided equal notice four tenths of the time. Additionally, the FAR-selected model failed to raise an alarm more times than all ATQ-selected models (Figure 2.9, Table 2.4). Generally, school year 2 either had poor alarms or no alarms raised, and the mean and median metric values were higher than all other school years across all model selection metrics (Table 2.5). This was likely due to the fact that there was only one school year of training data, as the alarm quality improved for later years in which there were more preceding school years included in the training data.

The timing of first alarms raised in each school year for all ten simulated replications is presented in Figure 2.9 and Table 2.4, however direct comparisons between the weighted and unweighted ATQ-selected models is not suggested since the weighted metrics do not perform as well as the unweighted metrics for earlier school years. No alarms were raised under the 10% absenteeism threshold method for any of the ten simulated replications. While ATQ-based and FAR-selected models raised their first alarm within the ideal range more frequently than in any other alarm time category, the AATQ-selected model raised the most first alarms within the ideal range. However, the AATQ-selected model also raised the first alarm too early the most frequently. The FAR-selected model raised more first alarms within the slightly late time range, followed by the FATQ and the AATQ-selected models. The FATQ-selected model raised the most alarms within the slightly early time range, followed by the AATQ and the FAR-selected models. Although alarms raised in the slightly early and slightly late time range are not ideal, they are still useful alarms for public health officials to react in order to reduce the spread of influenza within the community. Overall, the FATQ-selected model had the highest proportion of first alarms raised within the acceptable time range (53%). The proportion of first alarms raised within the acceptable time range was comparable between the AATQ and FAR-selected models, at 45% and 47% respectively, however the AATQ-selected model had the most first alarms raised within in the ideal time range (27%).

To provide a more direct comparison of the unweighted and weighted ATQ-selected models, the last 5 years of each of the 10 replications are compared in Table 2.5. The WAATQ had a lower mean and median AATQ_j value in 3 out of the last 5 years of the simulation study, indicating that the WAATQ-selected model had improved alarms in the latter years. The WFATQ-selected model had a lower mean FATQ_j in 3 out of the last 5 years, but it only had a lower median FATQ_j in 2 out of the last 5 years of the simulation study when compared to that of the FATQ-selected model. Both the WAATQ and WFATQ outperformed their unweighted counterparts in the last year of the simulation study.

2.5 Discussion

In this study, we proposed a new metric, ATQ, to simultaneously evaluate the accuracy and timeliness of alarms raised by a school absenteeism-based syndromic surveillance system. Summary statistics of the ATQ, AATQ and FATQ, were used as model selection criterion. These metrics supported the selection of a model that raised an alarm for an approaching seasonal influenza epidemic in a timely manner, and was shown to generally select a model that raised more quality alarms than the FAR-selected model or the 10% absenteeism threshold method.

Overall, the model selected by the FATQ raised acceptable first alarms more frequently than the other metrics. However, the AATQ-selected model raised a first alarm within the ideal time range of 7-14 days prior to the reference date the most frequently. Despite the AATQ-selected model having the most occurrences of alarms being raised too early in the influenza season, it still raised a comparable number of acceptable first alarms to that of the FAR-selected model while attaining more ideal alarms. Thus the FATQ and AATQ-selected models outperformed the FAR-selected model and the 10% absenteeism threshold method currently in place. For the later school years' predictions, the AATQ and FATQ metrics were slightly improved by the inclusion of a weight function in their calculations. Our results were consistent between the WDGPH and simulation studies.

Measures such as sensitivity and specificity have commonly been used to evaluate the performance of surveillance systems (21). These measures are usually combined into a single metric characterized by the area under the receiver operating characteristic (ROC) curve, to be used for comparisons across models. However these measures do not capture the timeliness of an alarm which is crucial for surveillance (21). and Abrams extended the ROC curve to include timeliness (7), but this extension assumes the epidemic has already begun, and measures time between the beginning of the epidemic and the first alarm. In our study, we were interested in raising an alarm prior to when the epidemic would be declared based on reported laboratory-confirmed influenza cases, enabling timely interventions by public health agencies. To the best of our knowledge, there are no other metrics that evaluate the accuracy and timeliness of alarms raised prior to the start of an epidemic.

While the ATQ metrics outlined in this study were successful in selecting an epidemic detection model that raised more high quality alarms than the FAR-selected model, there are limitations. Since the AATQ averages all alarms within a year, in the extreme case where an alarm was raised everyday until the epidemic reference date in a given year, the AATQ would still produce a value less than 1 (the maximum value of AATQ) for that year despite the poor specificity of this epidemic detection model. The FATQ does not have the same limitation since its calculation uses the first alarm raised in a given year, and therefore would give its maximum value of 1. In addition, the AATQ would favour an epidemic detection model that raises one alarm at the optimal time over any other model, including a hypothetical model that raises an alarm every day during the ideal range. The hypothetical model that raises alarms every day within the ideal range has a higher confidence, but because every suboptimal alarm has a non-zero ATQ, no matter how small, the AATQ increases, thus favouring the model that produced one alarm at the optimal time. However, based on our results this is unlikely since subsequent alarms often follow shortly after the initial alarm.

Absenteeism data provided by WDGPH was limited by the voluntary completion of the online form by schools in the UGDSB. Only 5 out of 88 participating elementary schools consistently reported absenteeism information over the entire study period, leaving a substantial amount of missing data. Since daily absenteeism was averaged across all reporting schools, alarms would be contingent on the location of the reporting schools and may not be reflective of the whole region. For example, if initial influenza activity occurred in Wellington county, but the only reporting schools are within Guelph, the resultant alarm may be too late for the whole region. To facilitate school absenteeism-based syndromic surveillance, as of January 2020 it is mandatory that schools within the UGDSB report daily absenteeism to WDGPH. Mandatory absence reporting will reduce the bias introduced by few consistently reporting schools, and create alarms more reflective of the entire region. Additionally, mandatory absence reporting will provide more information to characterize baseline absenteeism, and thus models will have a higher sensitivity to the change in influenza-specific absenteeism patterns. Our simulation study was based on a scenario with no missing data, more closely mimicking a system with mandatory absenteeism reporting, although in reality, school absenteeism can only be collected on school days and thus will always be missing weekends and holidays. Despite this discrepancy, both the WDGPH study and simulation study led to similar results.

The WDG region is predominantly rural, with the exception being the City of Guelph, a relatively small area of WDG where approximately half of the region's total population is located (17; 16). Due to the rurality of most of the region, alarms raised by the influenza surveillance system may only pertain to a specific area of the greater region. Thus, if spatial parameters are incorporated into the epidemic detection models, localized alarms can be raised. While spatial locations of reporting schools were not provided for confidentiality reasons, school catchment area identifiers could be obtained without loss of anonymity, and may provide a suitable spatial approximation for localized epidemic detection improvement. In this study, there was insufficient data to study catchment area alarms due to the small number of consistently reporting elementary schools, however future work could incorporate catchment areas into the simulation study to evaluate the performance of localized alarms.

The simulated absenteeism and laboratory confirmation system might not be able to fully capture the dynamics of a real influenza epidemic. For example, studies have shown that students are more likely to be absent on Mondays and Fridays, and that school absence episodes usually last a single day (15). Additionally, influenza transmission varies between weekdays and weekends (25). Thus, developing a more refined simulation model to account for the variation due to the day of week may help to generate data closer to that of a real influenza epidemic. Finally, our simulation study was based on the collective estimates of population parameters, such as the absenteeism rate and the rate at which individuals seek medical attention, that were specific to the population in the WDG region. This simulation model can be generalized to other regions by using the collective estimates of population parameters from the respective regions. Future work could include the sensitivity analysis of the proposed metrics based on simulation studies on populations from different public health regions.

Based on the results of this study, we believe that the ATQ metric is suitable to be used to evaluate accuracy and timeliness of a raised alarm by a given surveillance system. We recommend selecting a model that minimizes the FATQ for school absenteeism-based influenza surveillance in WDG. This approach raised the first alarm within the acceptable time range the most frequently in both a real data and simulation study. It is suggested that the selected model, and its parameters, be updated annually to incorporate yearly influenza and school absenteeism data.

Bibliography

[1] Bates, D., Mächler, M., Bolker, B., and Walker, S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*, 1 (2015), 1–48.

[2] Deardon, R., Brooks, S. P., Grenfell, B. T., Keeling, M. J., Tildesley, M. J., Savill, N. J., Shaw, D. J., and Woolhouse, M. E. Inference for individual-level models of infectious diseases in large populations. *Statistica Sinica 20*, 1 (2010), 239.

[3] Eysenbach, G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. In *AMIA Annual Symposium Proceedings* (2006), vol. 2006, American Medical Informatics Association, p. 244.

[4] Hayward, A. C., Fragaszy, E. B., Bermingham, A., Wang, L., Copas, A., Edmunds, W. J., Ferguson, N., Goonetilleke, N., Harvey, G., Kovar, J., et al. Comparative community burden and severity of seasonal and pandemic influenza: results of the Flu Watch cohort study. *The Lancet Respiratory Medicine 2*, 6 (2014), 445–454.

[5] Henning, K. J. What is syndromic surveillance? *Morbidity and Mortality Weekly Report* (2004), 7–11.

[6] Hulth, A., Rydevik, G., and Linde, A. Web queries as a source for syndromic surveillance. *PloS One 4*, 2 (2009), e4378.

[7] Kleinman, K. P., and Abrams, A. M. Assessing surveillance using sensitivity, specificity and timeliness. *Statistical Methods in Medical Research 15*, 5 (2006), 445–464.

[8] Lau, L. L., Cowling, B. J., Fang, V. J., Chan, K.-H., Lau, E. H., Lipsitch, M., Cheng, C. K., Houck, P. M., Uyeki, T. M., Peiris, J. M., et al. Viral shedding and clinical illness in naturally acquired influenza virus infections. *The Journal of Infectious Diseases 201*, 10 (2010), 1509–1516.

[9] Lenaway, D. D., and Ambler, A. Evaluation of a school-based influenza surveillance system. *Public Health Reports 110*, 3 (1995), 333.

[10] Mann, P., O'Connell, E., Zhang, G., Llau, A., Rico, E., and Leguen, F. C. Alert system to detect possible school-based outbreaks of influenza-like illness. *Emerging Infectious Diseases 17*, 2 (2011), 262.

[11] Mogto, C. A. K., De Serres, G., Fradet, M. D., Lebel, G., Toutant, S., Gilca, R., Ouakki, M., Janjua, N. Z., and Skowronski, D. M. School absenteeism as an adjunct surveillance indicator: experience during the second wave of the 2009 H1N1 pandemic in Quebec, Canada. *PLoS One 7*, 3 (2012), e34084.

[12] Provincial Infectious Diseases Advisory Committee. Syndromic surveillance discussion paper. *Ontario Agency for Health Protection and Promotion* (2012).

[13] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

[14] Sasaki, A., Hoen, A. G., Ozonoff, A., Suzuki, H., Tanabe, N., Seki, N., Saito, R., and Brownstein, J. S. Evidence-based tool for triggering school closures during influenza outbreaks, Japan. *Emerging Infectious Diseases 15*, 11 (2009), 1841.

[15] Schmidt, W., Pebody, R., and Mangtani, P. School absence data for influenza surveillance: a pilot study in the United Kingdom. *Eurosurveillance 15*, 3 (2010), 19467.

[16] Statistics Canada. Census profile, 2016 census, Guelph, Ontario. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E, 2017. Accessed: 2020-09-22.

[17] Statistics Canada. Census profile, 2016 census, Wellington-Dufferin-Guelph Health Unit, Ontario. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E, 2017. Accessed: 2020-09-22.

[18] Thompson, W. W., Shay, D. K., Weintraub, E., Brammer, L., Cox, N., Anderson, L. J., and Fukuda, K. Mortality associated with influenza and respiratory syncytial virus in the United States. *Journal of the American Medical Association 289*, 2 (2003), 179–186.

[19] Todd, S., Diggle, P. J., White, P. J., Fearne, A., and Read, J. M. The spatiotemporal association of non-prescription retail sales with cases during the 2009 influenza pandemic in Great Britain. *BMJ Open 4*, 4 (2014).

[20] Tokars, J. I., Olsen, S. J., and Reed, C. Seasonal incidence of symptomatic influenza in the United States. *Clinical Infectious Diseases 66*, 10 (2018), 1511–1518.

[21] Unkel, S., Farrington, C. P., Garthwaite, P. H., Robertson, C., and Andrews, N. Statistical methods for the prospective detection of infectious disease outbreaks: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 175*, 1 (2012), 49–82.

[22] Vergu, E., Grais, R. F., Sarter, H., Fagot, J.-P., Lambert, B., Valleron, A.-J., and Flahault, A. Medication sales and syndromic surveillance, France. *Emerging Infectious Diseases 12*, 3 (2006), 416.

[23] Ward, M. A., Stanley, A., Deeth, L. E., Deardon, R., Feng, Z., and Trotz-Williams, L. A. Methods for detecting seasonal influenza epidemics using a school absenteeism surveillance system. *BMC Public Health 19*, 1 (2019), 1232.

[24] Warriyar, K. V. V., Almutiry, W., and Deardon, R. *EpiILM: Spatial and Network Based Individual Level Models for Epidemics*, 2020. R package version 1.5.

[25] Willem, L., Van Kerckhove, K., Chao, D. L., Hens, N., and Beutels, P. A nice day for an infection? Weather conditions and social contact patterns relevant to influenza transmission. *PloS One 7*, 11 (2012), e48695.

2.6 Tables

Table 2.1: Reference dates representing the beginning of each seasonal influenza epidemic. A reference date is defined as the date of the second laboratory-confirmed influenza case within a seven day period for the first time within an influenza season.

School Year	Reference Date
2008-09	January 20, 2009
2010-11	December 14, 2010
2011-12	January 9, 2012
2012-13	October 26, 2012
2013-14	November 27, 2013
2014-15	December 8, 2014
2015-16	November 17, 2015
2016-17	December 15, 2016
2017-18	December 6, 2017

Table 2.2: Models selected by each metric, and its lag and threshold values. Bold values indicate the minimum value of the given metric.

Model	Parameters	FAR	AATQ	FATQ	WAATQ	WFATQ
Seasonal Mixed	$l = 11, \theta = 0.25$	**0.3125**	0.3545	0.3208	0.3512	0.3123
Seasonal Mixed	$l = 15, \theta = 0.15$	0.6043	**0.2171**	0.3312	**0.1694**	0.2912
Seasonal Mixed	$l = 15, \theta = 0.25$	0.4583	0.3186	**0.2923**	0.3224	0.2912
Seasonal Mixed, DOW	$l = 11, \theta = 0.25$	0.3750	0.3351	0.3021	0.3227	**0.2747**

Table 2.3: Frequency of first alarms that occurred in each time range based on WDGPH data, under the selected models.

Optimized Metric	**No Alarm**	**Too late 0-3 days before ref.**	**Slightly late 4-6 days before ref.**	**Ideal 7-14 days before ref.**	**Slightly early 15-21 days before ref.**	**Too early 22+ days before ref.**
10% Threshold	5	2	0	0	0	1
FAR	2	3	0	2	1	0
AATQ	1	1	1	0	3	2
FATQ	2	2	0	2	2	0
WAATQ	1	1	1	0	3	2
WFATQ	2	2	0	3	1	0

Table 2.4: Proportion of first alarms that occurred in each time range over 10 replications of 9 prediction years, under the selected models.

Model	**No alarm**	**Too late 0-3 days before ref.**	**Slightly late 4-6 days before ref.**	**Ideal 7-14 days before ref.**	**Slightly early 15-21 days before ref.**	**Too early 22+ days before ref.**
10% Threshold	1.00	0.00	0.00	0.00	0.00	0.00
FAR	0.21	0.19	0.13	0.24	0.09	0.13
AATQ	0.18	0.17	0.08	0.27	0.11	0.20
FATQ	0.19	0.14	0.12	0.24	0.17	0.13
WAATQ	0.18	0.16	0.11	0.20	0.13	0.22
WFATQ	0.20	0.18	0.13	0.19	0.13	0.17

Table 2.5: Evaluation metric measures of variability for each year of the 10 replications under various models.

Model	**Measure**	**Year 2**	**Year 3**	**Year 4**	**Year 5**	**Year 6**	**Year 7**	**Year 8**	**Year 9**	**Year10**
$\mathbf{FAR}_j$	Mean	0.6973	0.2250	0.2300	0.5875	0.3625	0.3911	0.2500	0.3500	0.3395
	Median	0.8782	0.000	0.000	0.9373	0.000	0.0000	0.0000	0.0000	0.0000
	Variance	0.1590	0.1451	0.1512	0.2562	0.2276	0.2552	0.1806	0.2250	0.2104
$\mathbf{AATQ}_j$	Mean	0.4218	0.1656	0.1707	0.3894	0.2898	0.2250	0.2671	0.3649	0.1335
	Median	0.2557	0.0572	0.0682	0.1486	0.0830	0.1394	0.0829	0.1139	0.1140
	Variance	0.1740	0.0890	0.0876	0.1793	0.1466	0.0801	0.1522	0.1933	0.0115
$\mathbf{WAATQ}_j$	Mean	0.4310	0.3623	0.1761	0.3996	0.1995	0.3058	0.0780	0.3809	0.1254
	Median	0.2899	0.0908	0.0853	0.1613	0.0830	0.1170	0.0698	0.1485	0.0889
	Variance	0.1790	0.1959	0.0867	0.1729	0.0857	0.1404	0.0032	0.1839	0.0134
$\mathbf{FATQ}_j$	Mean	0.5699	0.2607	0.1428	0.5086	0.3322	0.3375	0.1510	0.3340	0.2749
	Median	0.7222	0.0230	0.0274	0.4408	0.0567	0.0790	0.0283	0.0724	0.0753
	Variance	0.2207	0.1760	0.0946	0.2211	0.2141	0.2127	0.0939	0.2134	0.1576
$\mathbf{WFATQ}_j$	Mean	0.5852	0.4114	0.3669	0.4679	0.2046	0.2583	0.1886	0.3534	0.1760
	Median	0.7222	0.1302	0.0389	0.3285	0.0465	0.0910	0.1111	0.1290	0.0350
	Variance	0.2035	0.2257	0.2216	0.1851	0.1166	0.1570	0.0891	0.2029	0.0981

2.7 Figures

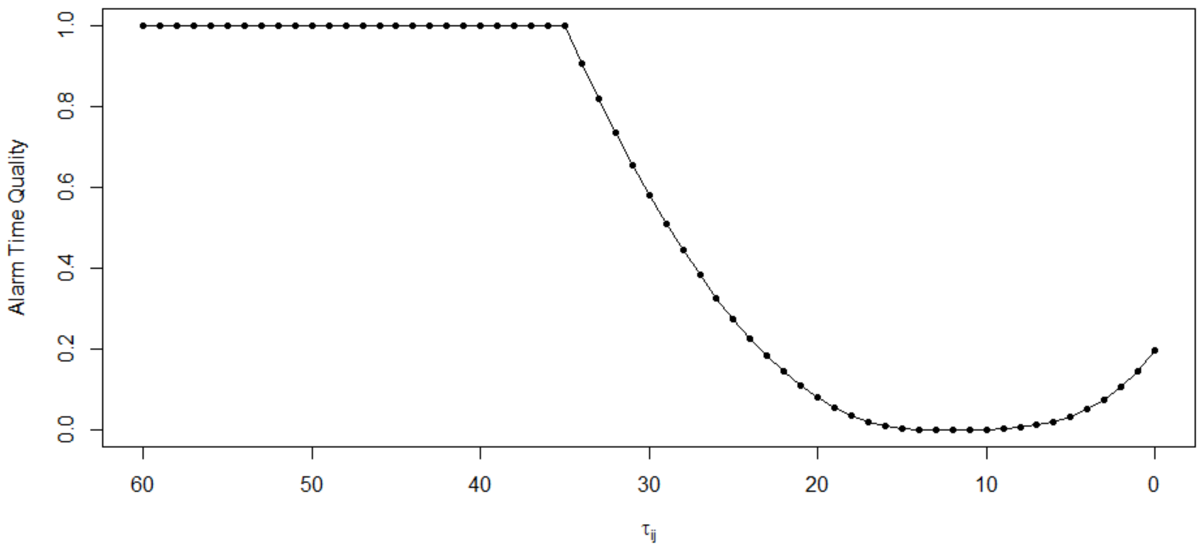

Figure 2.1: Alarm time quality values corresponding to the number of days prior to the reference date ($\tau_{ij} = 0$), that an alarm was raised

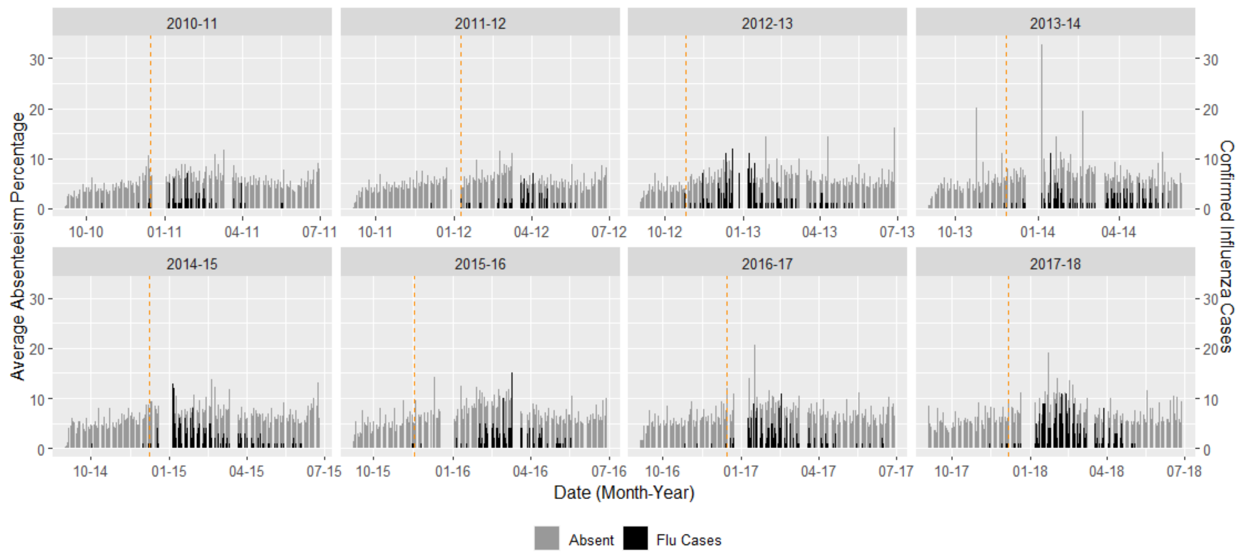

Figure 2.2: Daily average elementary school absenteeism and laboratory-confirmed influenza cases in WDG. Each panel represents a school year. Yearly reference dates are represented with the dashed orange lines.

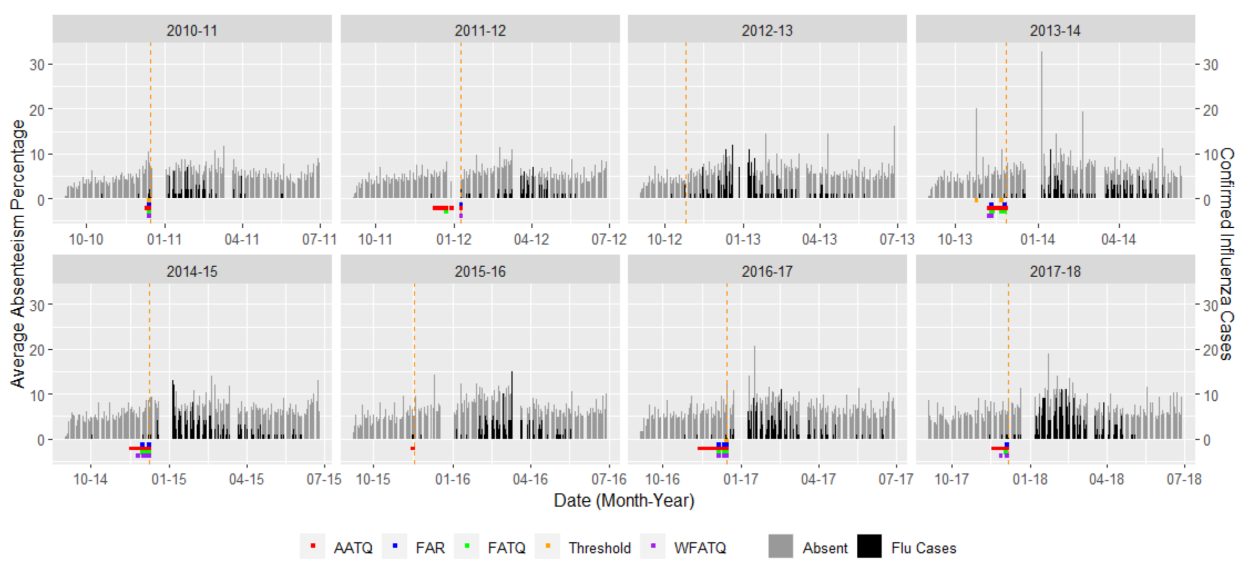

Figure 2.3: Alarms raised by the model selected by each evaluation metric for the WDGPH data. Coloured squares below the X-axis represent the date an alarm was raised by each of the selected models. Daily average absenteeism is plotted as grey bars, laboratory-confirmed influenza case counts are overlaid with black bars, and the epidemic reference day is indicated by the dashed orange lines.

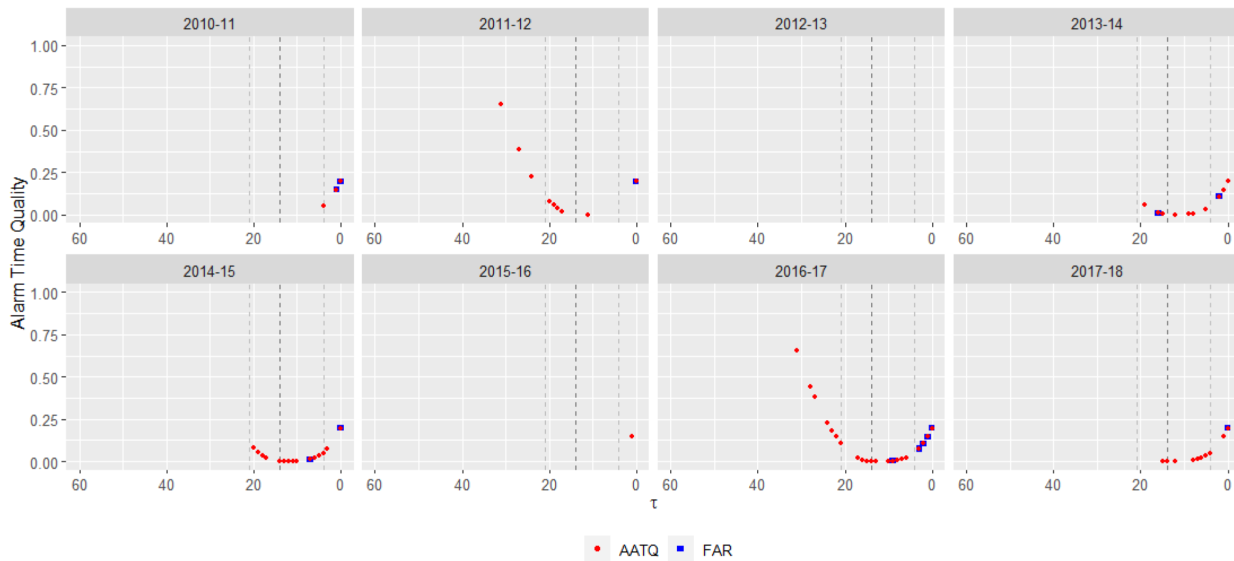

Figure 2.4: Comparison of the timing of alarms raised by the AATQ and FAR-selected models for the WDGPH data, relative to the reference date ($\tau = 0$). Each school year is represented in its own panel. The vertical black dashed line represents the optimal alarm time ($\tau = 14$). The vertical grey dashed lines represent the boundaries for acceptable alarms ($\tau = 21$ and $\tau = 4$).

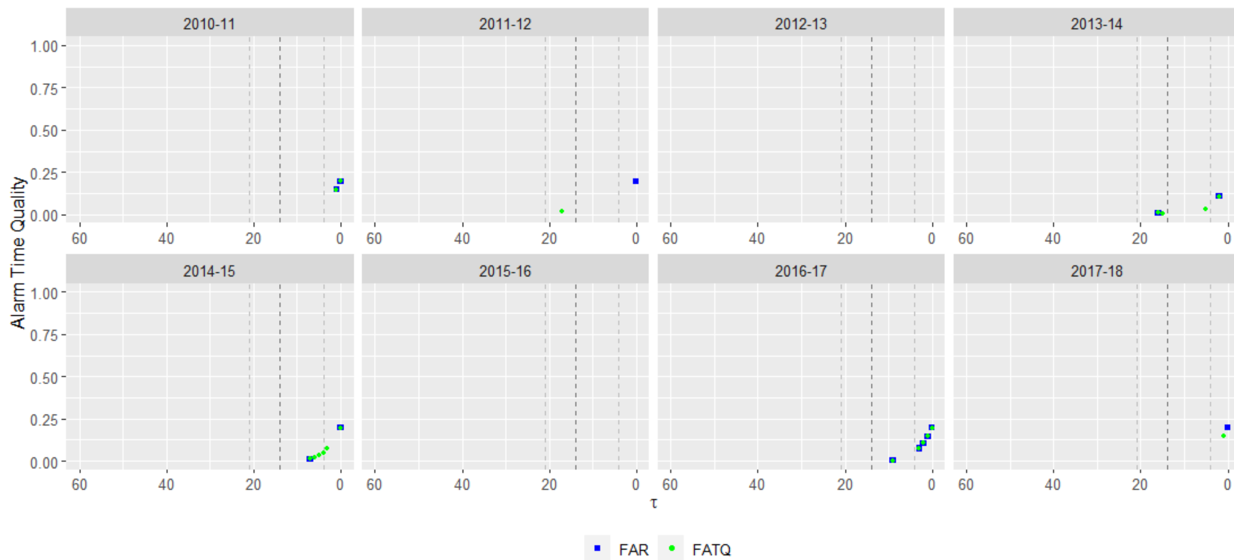

Figure 2.5: Comparison of the timing of alarms raised by the FATQ and FAR-selected models for the WDGPH data, relative to the reference date ($\tau = 0$). Each school year is represented in its own panel. The vertical black dashed line represents the optimal alarm time ($\tau = 14$). The vertical grey dashed lines represent the boundaries for acceptable alarms ($\tau = 21$ and $\tau = 4$).

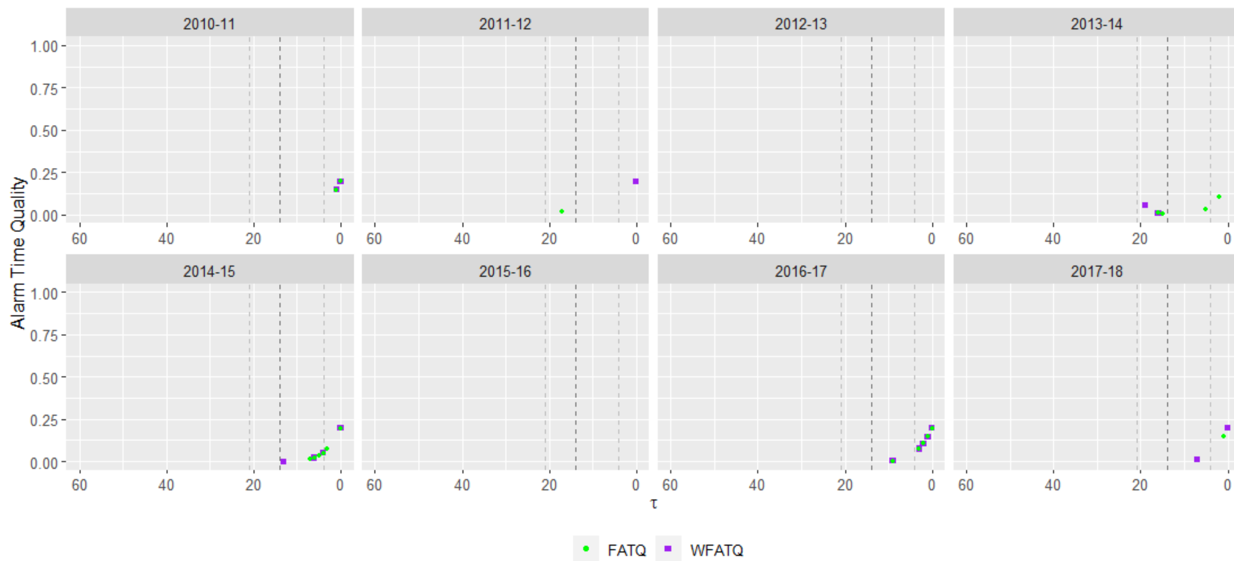

Figure 2.6: Comparison of the timing of alarms raised by the AATQ and WAATQ-selected models for the WDGPH data, relative to the reference date ($\tau = 0$). Each school year is represented in its own panel. The vertical black dashed line represents the optimal alarm time ($\tau = 14$). The vertical grey dashed lines represent the boundaries for acceptable alarms ($\tau = 21$ and $\tau = 4$).

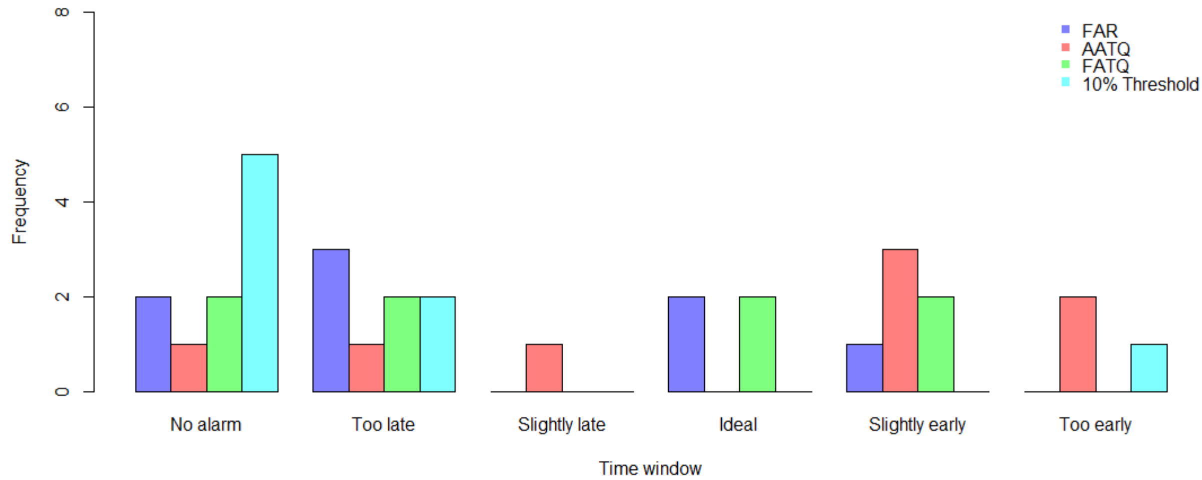

Figure 2.7: Frequency of first alarms that occurred in each time range for the WDGPH data, under each selected model. Alarms are categorized as follows: too late (alarm raised 0-3 days prior to the reference date), slightly late (alarm raised 4-6 days prior to the reference date), ideal (alarm raised 7-14 days prior to the reference date), slightly early (alarm raised 15-21 days prior to the reference date), and too early (alarm raised more than 21 days prior to the reference date). An acceptable alarm is an alarm categorized as slightly late, ideal, or slightly early.

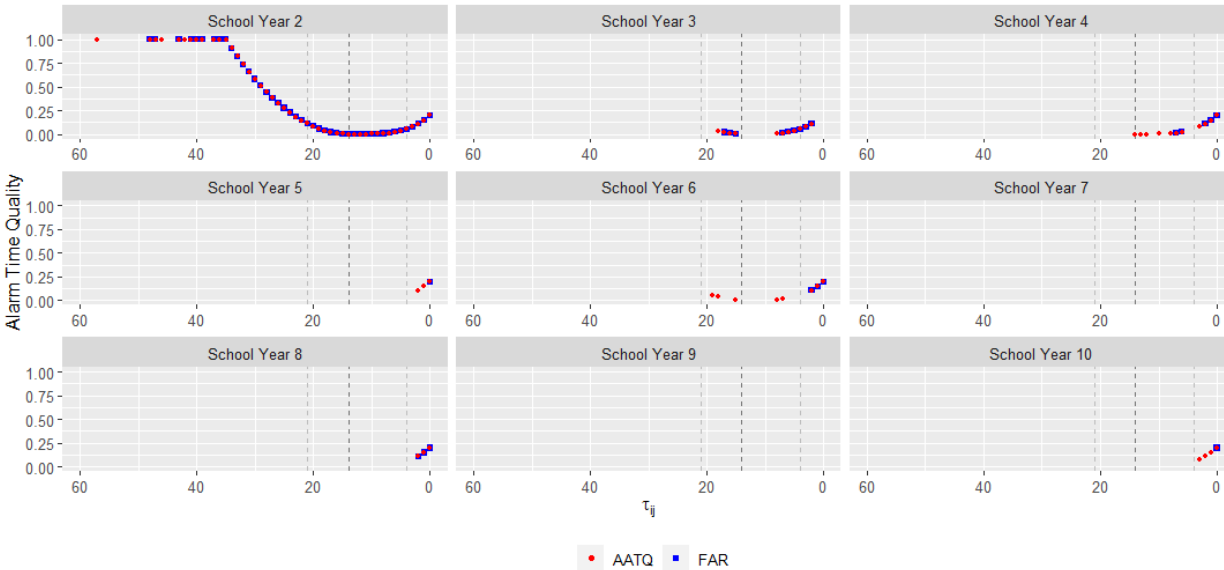

Figure 2.8: Comparison of the timing of alarms raised by the AATQ and FAR-selected models for one replication of the simulation study, relative to the reference date ($\tau = 0$). Each school year is represented in its own panel. The vertical black dashed line represents the optimal alarm time ($\tau = 14$). The vertical grey dashed lines represent the boundaries for acceptable alarms ($\tau = 21$ and $\tau = 4$).

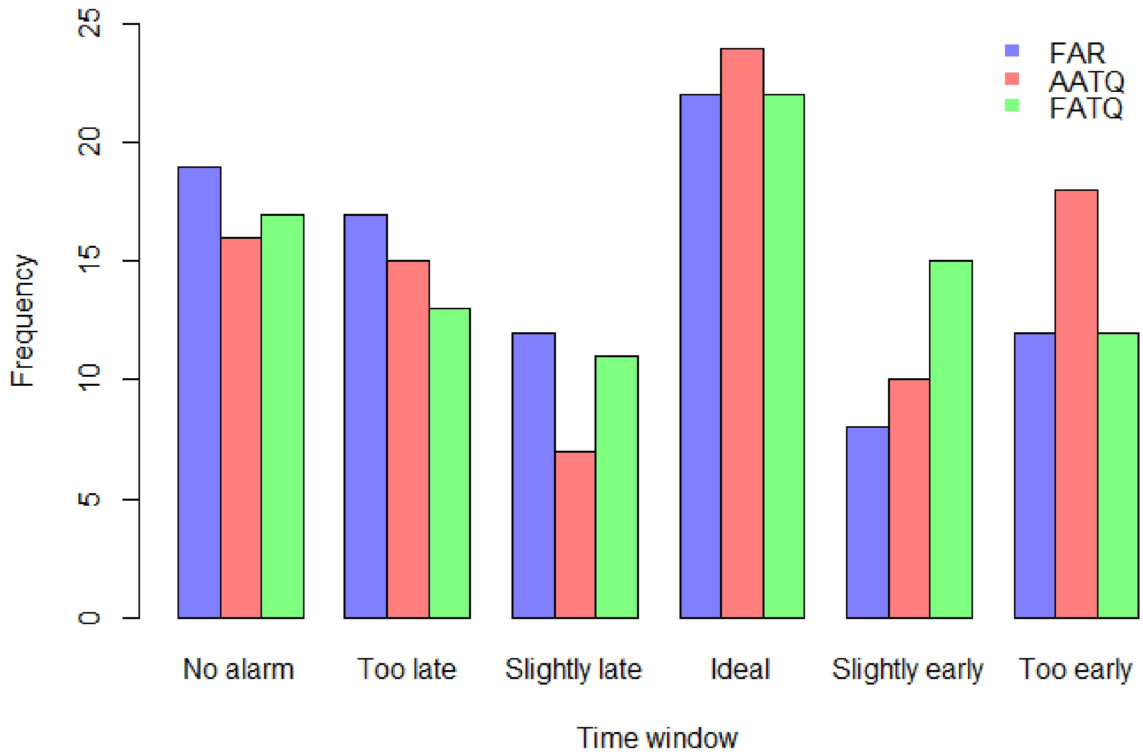

Figure 2.9: Frequency of first alarms that occurred in each time range over 10 replications of 9 prediction years, under each selected model. Alarms are categorized as follows: too late (alarm raised 0-3 days prior to the reference date), slightly late (alarm raised 4-6 days prior to the reference date), ideal (alarm raised 7-14 days prior to the reference date), slightly early (alarm raised 15-21 days prior to the reference date), and too early (alarm raised more than 21 days prior to the reference date). An acceptable alarm is an alarm categorized as slightly late, ideal, or slightly early.

Chapter 3

Further Work and Conclusion

3.1 Catchment Area Analysis

Initial efforts have been made to assess the use of school catchment areas as a spatial approximation within epidemic detection models, and the quality of the resultant local alarms. The data provided by WDGPH used UGDSB elementary public school boundaries to map participating schools and influenza cases to 40 distinct catchment areas. However, dividing the relatively small region of WDG region into 40 distinct areas created challenges.

The first notable challenge was low case counts in many of the catchment areas. Having few cases in catchment areas resulted in difficulties defining appropriate reference dates for each catchment area and year. Using the current reference date definition, 65% of catchment areas within a given year did not define a reference date. Of the school catchment areas that did not define a reference date, the time interval between laboratory-confirmed influenza cases was sometimes longer than the seven days required of the current reference date definition. Conversely, of the catchment areas that were able to define a reference date within a year, some reference dates were near the end of the seasonal epidemic, in months such as April and May. In an attempt to define a more appropriate reference date for catchment areas, we considered the following dates: the date of the first laboratory-confirmed influenza case within a catchment area that occurred on or after the region-wide reference date; and the date of the second laboratory-confirmed influenza case within a catchment area that was within a ten day period for the first time in an influenza season. However, due to the low case counts, some catchment areas remained without a reference date under the alternative definitions.

In an attempt to accommodate the lack of reference dates for some catchment areas, the AATQ_j and FATQ_j metrics were updated accordingly. In the event that no reference date was defined for a catchment area in a given year, and no alarm was raised, an AATQ_j and FATQ_j value of 0 was assigned to reward the absence of false alarms. Conversely, if there was no reference date defined for a catchment area in a given year, and an alarm was raised, an AATQ_j and FATQ_j value of 1 was assigned to penalize false alarms. If there was a reference date within a catchment area for a given year, metric values were calculated as per usual. The AATQ and FATQ calculations were adjusted to include the mean value of each catchment areas' alarms, or first alarms, within a given year. The seasonal logistic regression model used for raising alarms was modified to include catchment area indicator variables.

Another challenge using the WDGPH catchment areas, was the scarcity of reported absenteeism data. Of the 40 catchment areas, only five catchment areas reported absenteeism on more than 40 school days each school year. This rendered the continued analysis of the WDGPH catchment area data infeasible, however the simulation study in Chapter 2 included avenues to allow for catchment area analysis. Preliminary results indicated that very few catchment area alarms were raised. This was likely due to the imbalance of catchment areas without reference dates, favouring models that did not raise alarms for the years that catchment areas did not have reference dates. Additional analyses explored dividing the catchment areas into low and high population density categories, and model each category separately. Some improvements were seen with respect the amount of alarms being raised for the high density catchments, however the region-wide model still outperformed the catchment area model with respect to producing timely alarms.

Since WDG is a relatively small public health unit, with a largely rural population, dividing the region into smaller areas for local alarms may not be beneficial. A catchment area approach may be better suited for a more densely populated region such as Toronto, Ontario, Canada. Future work could include collaboration with larger public health units, and simulation studies on a set of diverse populations, to evaluate the performance of catchment area models against region wide models. However, further investigation on appropriate catchment area reference dates is recommended prior to implementation.

3.2 Conclusion

In conclusion, this thesis developed the ATQ metric to evaluate accuracy and timeliness of alarms raised, and was used as a model selection criterion through metric minimization. The FATQ-selected model proved to be able to raise the first alarm within the acceptable time range the most frequently, and outperformed the other considered metrics. These results were supported by our simulation study. Therefore, it is recommended to select a model that minimizes the FATQ for school absenteeism-based influenza surveillance in WDG. The increased notice of an upcoming influenza epidemic that the FATQ-selected model can provide, will be beneficial in reducing the impact of seasonal influenza on the community. The developed simulation study can also be used by other public health units considering the implementation of a school absenteeism-based influenza surveillance system.

Bibliography

[1] Bates, D., Mächler, M., Bolker, B., and Walker, S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*, 1 (2015), 1–48.

[2] Coory, M., Kelly, H., and Tippett, V. Assessment of ambulance dispatch data for surveillance of influenza-like illness in Melbourne, Australia. *Public Health 123*, 2 (2009), 163–168.

[3] Deardon, R., Brooks, S. P., Grenfell, B. T., Keeling, M. J., Tildesley, M. J., Savill, N. J., Shaw, D. J., and Woolhouse, M. E. Inference for individual-level models of infectious diseases in large populations. *Statistica Sinica 20*, 1 (2010), 239.

[4] Dijk, A. v., Aramini, J., Edge, G., and Moore, K. M. Real-time surveillance for respiratory disease outbreaks, Ontario, Canada. *Emerging Infectious Diseases 15*, 5 (2009), 799.

[5] Eysenbach, G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. In *AMIA Annual Symposium Proceedings* (2006), vol. 2006, American Medical Informatics Association, p. 244.

[6] Greenko, J., Mostashari, F., Fine, A., and Layton, M. Clinical evaluation of the Emergency Medical Services (EMS) ambulance dispatch-based syndromic surveillance system, New York City. *Journal of Urban Health 80*, 1 (2003), i50–i56.

[7] Hayward, A. C., Fragaszy, E. B., Bermingham, A., Wang, L., Copas, A., Edmunds, W. J., Ferguson, N., Goonetilleke, N., Harvey, G., Kovar, J., et al. Comparative community burden and severity of seasonal and pandemic in-

fluenza: results of the Flu Watch cohort study. *The Lancet Respiratory Medicine 2*, 6 (2014), 445–454.

[8] Henning, K. J. What is syndromic surveillance? *Morbidity and Mortality Weekly Report* (2004), 7–11.

[9] Hiller, K. M., Stoneking, L., Min, A., and Rhodes, S. M. Syndromic surveillance for influenza in the emergency department–a systematic review. *PloS One 8*, 9 (2013), e73832.

[10] Hulth, A., Rydevik, G., and Linde, A. Web queries as a source for syndromic surveillance. *PloS One 4*, 2 (2009), e4378.

[11] Kleinman, K. P., and Abrams, A. M. Assessing surveillance using sensitivity, specificity and timeliness. *Statistical Methods in Medical Research 15*, 5 (2006), 445–464.

[12] Lau, L. L., Cowling, B. J., Fang, V. J., Chan, K.-H., Lau, E. H., Lipsitch, M., Cheng, C. K., Houck, P. M., Uyeki, T. M., Peiris, J. M., et al. Viral shedding and clinical illness in naturally acquired influenza virus infections. *The Journal of Infectious Diseases 201*, 10 (2010), 1509–1516.

[13] Lenaway, D. D., and Ambler, A. Evaluation of a school-based influenza surveillance system. *Public Health Reports 110*, 3 (1995), 333.

[14] Longini Jr, I. M., Koopman, J. S., Monto, A. S., and Fox, J. P. Estimating household and community transmission parameters for influenza. *American Journal of Epidemiology 115*, 5 (1982), 736–751.

[15] Mann, P., O'Connell, E., Zhang, G., Llau, A., Rico, E., and Leguen, F. C. Alert system to detect possible school-based outbreaks of influenza-like illness. *Emerging Infectious Diseases 17*, 2 (2011), 262.

[16] Mogto, C. A. K., De Serres, G., Fradet, M. D., Lebel, G., Toutant, S., Gilca, R., Ouakki, M., Janjua, N. Z., and Skowronski, D. M. School absenteeism as an adjunct surveillance indicator: experience during the second wave of the 2009 H1N1 pandemic in Quebec, Canada. *PLoS One 7*, 3 (2012), e34084.

[17] Mook, P., Joseph, C., Gates, P., and Phin, N. Pilot scheme for monitoring sickness absence in schools during the 2006/07 winter in England: can these data be used as a proxy for influenza activity? *Eurosurveillance 12*, 12 (2007), 11–12.

[18] Mostashari, F., Fine, A., Das, D., Adams, J., and Layton, M. Use of ambulance dispatch data as an early warning system for communitywide influenzalike illness, New York City. *Journal of Urban Health 80*, 1 (2003), i43–i49.

[19] Neuzil, K. M., Hohlbein, C., and Zhu, Y. Illness among schoolchildren during influenza season: effect on school absenteeism, parental absenteeism from work, and secondary illness in families. *Archives of Pediatrics & Adolescent Medicine 156*, 10 (2002), 986–991.

[20] Provincial Infectious Diseases Advisory Committee. Syndromic surveillance discussion paper. *Ontario Agency for Health Protection and Promotion* (2012).

[21] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2018.

[22] Rosenkötter, N., Ziemann, A., Riesgo, L. G.-C., Gillet, J. B., Vergeiner, G., Krafft, T., and Brand, H. Validity and timeliness of syndromic influenza

surveillance during the autumn/winter wave of A (H1N1) influenza 2009: results of emergency medical dispatch, ambulance and emergency department data from three European regions. *BMC Public Health 13*, 1 (2013), 905.

[23] Sasaki, A., Hoen, A. G., Ozonoff, A., Suzuki, H., Tanabe, N., Seki, N., Saito, R., and Brownstein, J. S. Evidence-based tool for triggering school closures during influenza outbreaks, Japan. *Emerging Infectious Diseases 15*, 11 (2009), 1841.

[24] Schmidt, W., Pebody, R., and Mangtani, P. School absence data for influenza surveillance: a pilot study in the United Kingdom. *Eurosurveillance 15*, 3 (2010), 19467.

[25] Statistics Canada. Census profile, 2016 census, Guelph, Ontario. `https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E`, 2017. Accessed: 2020-09-22.

[26] Statistics Canada. Census profile, 2016 census, Wellington-Dufferin-Guelph Health Unit, Ontario. `https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E`, 2017. Accessed: 2020-09-22.

[27] Thompson, W. W., Shay, D. K., Weintraub, E., Brammer, L., Cox, N., Anderson, L. J., and Fukuda, K. Mortality associated with influenza and respiratory syncytial virus in the United States. *Journal of the American Medical Association 289*, 2 (2003), 179–186.

[28] Todd, S., Diggle, P. J., White, P. J., Fearne, A., and Read, J. M. The spatiotemporal association of non-prescription retail sales with cases during the 2009 influenza pandemic in Great Britain. *BMJ Open 4*, 4 (2014).

[29] Tokars, J. I., Olsen, S. J., and Reed, C. Seasonal incidence of symptomatic influenza in the United States. *Clinical Infectious Diseases 66*, 10 (2018), 1511–1518.

[30] Unkel, S., Farrington, C. P., Garthwaite, P. H., Robertson, C., and Andrews, N. Statistical methods for the prospective detection of infectious disease outbreaks: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 175*, 1 (2012), 49–82.

[31] Vergu, E., Grais, R. F., Sarter, H., Fagot, J.-P., Lambert, B., Valleron, A.-J., and Flahault, A. Medication sales and syndromic surveillance, France. *Emerging Infectious Diseases 12*, 3 (2006), 416.

[32] Ward, M. A., Stanley, A., Deeth, L. E., Deardon, R., Feng, Z., and Trotz-Williams, L. A. Methods for detecting seasonal influenza epidemics using a school absenteeism surveillance system. *BMC Public Health 19*, 1 (2019), 1232.

[33] Warriyar, K. V. V., Almutiry, W., and Deardon, R. *EpiILM: Spatial and Network Based Individual Level Models for Epidemics*, 2020. R package version 1.5.

[34] Willem, L., Van Kerckhove, K., Chao, D. L., Hens, N., and Beutels, P. A nice day for an infection? Weather conditions and social contact patterns relevant to influenza transmission. *PloS One 7*, 11 (2012), e48695.

[35] World Health Organization. Behavioral interventions for reducing the transmission and impact of influenza A (H1N1) virus: a framework for communication strategies. `https://www.who.int/csr/resources/publications/swineflu/framework_20090626_en.pdf?ua=1`, 2009. Accessed: 2020-07-27.

[36] World Health Organization. Up to 650 000 people die of respiratory diseases linked to seasonal flu each year, 2017. Accessed: 2020-12-17.

Appendix

Complete R code for the simulation study is publicly available here: https://github.com/vanderkk/School_Abstenteeism_Based_Influenza_Surveillance_Simulation_Study