Predicting Community COVID-19 Public Health Needs Through Wastewater Based Epidemiology, Maine USA

Tyler Zanon MPH
University of New England
Faculty Advisor: Dr. Andrew Pritchard

Introduction:

Wastewater Based Epidemiology (WBE) is an expanding field with increasingly important public health applications.[1] WBE historically relied on environmental scientist achievements for relevancy but has experienced increased use in sewage monitoring for pathogens – including SARS-CoV-2 during the recent Coronavirus disease 2019 (COVID-19) pandemic.[1] COVID-19's high transmissibility resulted in rapid global spread to over 20 million confirmed cases and roughly 733,000 deaths within the first six months after China's lockdown on January 23, 2020.[2] Healthcare systems around the world were overburdened and, at times, incapable of sustaining their prevention, diagnosis, or treatment asks of their populations.[3] The pandemic's unique challenges posed problematic for the slow and cumbersome supply chain that is necessary for successful diagnostic testing.[1] In parallel, countries around the world turned to WBE to supplement traditional efforts via community level screening.[1, 3-8]

Applications of WBE, however, do not end with correlations between observed wastewater viral protein concentrations and community case rates. Various predictive models have been published to support public health estimations related to COVID-19.[9-12] Each method leverages easily captured WBE data to make meaningful public health predictive factors such as SARS-CoV-2 cases, hospitalizations, or deaths.[9-12] The existing models range in complexity, supporting a variety of community needs. Some methods leveraged linear models with relevant corrective factors while others employed SEIR (susceptible, exposed, infected, recovered) simulations to map changes in critical health population proportions. Each model has their own potential pros and cons. The complex SEIR models leverage differential equations that can help understand disease spread over time based on known disease factors. However, this method

could prove mathematically challenging and limit its generalized dissemination and use potential. Alternatively, linear models are limited by their data. If populations experience non-normal distributions of outcomes, the model cannot be applied to them.

In the United States, predictive model research related to COVID-19 varies in their practice and application. Phan et al surveyed wastewater plants serving roughly 2.3 million individuals in Massachusetts, pulling samples from October 02,2020 to January 25, 2021.[10] McMahan et al worked with Clemson University to build predictive models for a population of ~25,000 students broken into three smaller geographical areas.[9]

The challenges with COVID-19 predictive models are rooted in the evolving variation in the disease itself. SARS-CoV-2 is capable of rapid mutation that results in variant strains capable of their own unique disease states.[13] These rapid developments hinder the generalizability of previously published literature due to newly identified incubation periods, infection durations, recovery times, or mortality rates. For example, Phan's monitoring through January of 2021 may have resulted in their capture of data related to COVID-19's Omicron variant, which was designated in November of 2021, while McMahan's sampling would not. Since the pandemic began, the Center for Disease Control and Prevention (CDC) has designated dozens of different COVID-19 lineages of interest.[13] Omicron alone has 17 different lineages identified.[13] Variant contributions to the overall community disease state may vary drastically depending on the geographical location of interest.[13]

There are still many gaps related to WBE based predictive models and COVID-19 disease screening. For starters, geography and timing are crucial to understanding COVID-19 spread and impact. Continued research on varied populations is needed to grow the generalizability of

available predictive models. During their APE, the student determined statistically significant correlations between wastewater concentrations of SARS-CoV-2 RNA and community case rates (Spearman Correlation Coefficient 0.77, p-value of 3.97E-11). Leveraging these predictive models across new communities will support the growing evidence related to their effective utilization for COVID-19 screening. Each involved community have robust sociodemographic data.[14] These data further potential generalizability conclusions for COVID-19 WBE based predictive models.

Additionally, available literature concentrates on limited data collection timepoints that minimizes their tracing of multiple COVID-19 variants and sub-variants.[9-13] The communities involved in this research have collected samples from December 2020 through August 2021. This date range greatly broadens the capture of various COVID-19 variants according to the CDC's viral proportion calculations.[13] Lastly, recommended protocols, including incubation times, infectivity windows, and recovery times have evolved rapidly throughout the pandemic.[15] The new models developed in this research can understand whether previously published measures are still appropriate or whether recent developments have changed their applicability in predicting community cases.

This research aims to determine the applicability of various models in respect to their application to four Maine based communities.

Methods

Tables 1, 2, and 3 provide an overview of the communities involved in the test populations.

Table 1 - Community Population and Housing Characteristics[14]			
City	Population	Households	Persons per Households
Brunswick	21,836	8295	2.28
East End (Portland)	68,313	30796	2.10
Westbrook/Gorham	38,998	14444	2.47
Yarmouth	8,997	3247	2.59

East End (Portland) and Westbrook/Gorham were the two most heavily populated communities with Yarmouth being roughly 7.6x and 4.3x smaller than each respectively.

Table 2 - Community Racial Demographics[14]					
City	White	Black	Hispanic	Asian	AI/AN*
Brunswick	91.90%	1.90%	3.70%	2.10%	0.90%
East End (Portland)	83.60%	8.60%	2.70%	3.90%	0.30%
Westbrook/Gorham	91.59%	2.85%	1.34%	1.50%	0.35%
Yarmouth	93.50%	0.70%	0.90%	2.80%	0.90%

*American Indian or Alaska Native

Table 3 - Community Age Distributions (Years)[14]				
City	< 5	5 to 17	18 to 64	65+
Brunswick	5.70%	17.10%	56.00%	21.20%
East End (Portland)	4.80%	15.40%	64.40%	15.40%
Westbrook/Gorham	5.83%	20.09%	58.15%	15.93%
Yarmouth	3.60%	23.40%	53.40%	19.60%

Each community is predominantly White, a range of 83.60%-93.50%, with varying degrees of smaller Black, Hispanic, Asian, and AI/AN populations. Similarly, most individuals in all communities are between the ages of 18 and 64 years old, typically working age populations.

Qualified community stakeholders collected 24-hour composite samples of influent water at each respective wastewater treatment plant.[16] The samples were measured for SARS-CoV-2 RNA concentration and wastewater flow rate.[16] The CDC recommended N1 and N2 genetic markers to be used during SARS-CoV-2 RNA quantification, as directed by the SARS-CoV-2 Reverse Transcription Polymerase Chain Reaction (RT-PCR) test (INDEXX Laboratories,

Inc. Westbrook, ME).[16] This test quantifies viral marker concentrations that enable understanding. External standard curves were utilized to calibrate the genetic marker concentrations in triplicate.[16] Viral recovery was measured via a surrogate virus, bovine respiratory syncytial virus (BRSV), spiked into wastewater samples prior to processing the samples.[16] This recovery value was factored into subsequent viral load calculations as appropriate. Each site collected and shared their data with the researcher for compilation and further analysis. Table 4 outlines the number of samples collected in each community.

Table 4: Dataset Summary			
Independent Variables	**Dependent Variables**	**City**	**Sample Size**
Wastewater Flow Rate Liters per day Viral RNA Copies per Liter (CPL) Weekly Reported COVID-19 Clinical Cases	Predicted COVID-19 Cases per Week	Brunswick	10
		East End (Portland)	50
		Westbrook/Gorham	50
		Yarmouth	34

The final data set was impacted by various missing datapoints from each community. Missing information included flowrates, community reported case counts, and BRSV recovery. When possible, missing data was remedied through stakeholder communication however the outcome did not always mitigate the lack of source data. BRSV recovery data was not available for analysis prior to December 15, 2020. Additionally, case count data was not available for East End (Portland) on 7/20/2021 and 8/10/2021 and for Westbrook/Gorham on 8/10/2021. A full compilation of missing data is available in Table 8 in the appendix. Lastly, when SARS-CoV-2 concentrations were below the limit of detection, 0.76 copies/mL, half of this concentration was used for calculations and analysis per the CDC's recommendations.[16] A list of non-detect samples can be found in Table 9 of the appendix.

The researcher uploaded the data sets into RStudio leveraging R version 2022.7.0+548 for data cleaning, visualization, and analysis.[17] The data was cleaned leveraging necessary

RStudio packages and further information can be observed in the associated R-markdown file or Table 10.[18-25]

Community case rates were converted from numerical outputs per city during the preceding sampling week to rates per 10,000 individuals. Populations were extracted from available census data and rates were calculated per below[20]:

Equation 1: Case Counts to Incidence per 10,000

$$\frac{Reported\ Case\ Counts}{10{,}000} = \left(\frac{Reported\ Case\ Counts}{City\ Population}\right) * 10{,}000$$

External literature was reviewed to identify available best fit predictive models for analysis. A summary of referenced articles is included below:

Table 5 – Referenced Predictive Models

Article	Year of Publication	Study Location	Population Studied	Statistical Method
McMahan et al[9]	2021	South Carolina, USA	~25,000 College Students	Simple Prediction Model & SEIR Prediction Model
Phan et al[10]	2022	Massachusetts, USA	2.3 Million Cross Community Residents	SEIR Prediction Model
Zhu et al[11]	2022	Japan	1 Million Urban Community Residents	Correlation Testing + Linear Prediction Models
Zhou et al[12]	2020	Michigan, USA	~630,000 Detroit Residents	Linear Regression with ARIMA Modeling

The variables of interest, SARS-CoV-2 RNA Copies per liter and reported or predicted COVID-19 community case rates, are numerical (continuous or integer). Other variables relevant for statistical analysis include wastewater flow rate in liters per day (numerical – continuous),

grams per day of fecal production per person (numerical – continuous), maximum rate of viral shedding per gram of feces per day (numerical – continuous), and location (categorical).[9]

A three-tiered approach was taken in this research. Firstly, the student leveraged the simplified predictive model put forth by McMahan et al[9] to determine its fit for the test communities. Differences between the predicted and observed cases were calculated and tested for statistical significance leveraging a Wilcoxon Sign Rank test.

The simplified prediction model can be observed below:

Equation Two: Simple Predictive Model

$$Jt = \frac{(Q * V)}{(A * B)}$$

Table 6: Simplified Model Breakdown		
Variable	**Meaning**	**Unit of Measure**
Jt	Predicted number of positive COVID-19 cases	Cases per week
Q	Average wastewater flow rate	Liters per day
V	Wastewater sampled SARS-CoV-2 RNA copies	Copies per liter
A	Rate of feces production per person	Grams per day
B	Maximum rate of viral shed	RNA copies per gram of feces per day

The variables Q and V were isolated from the observed data reported by community stakeholders, whereas variables A and B were constants identified in the research performed by McMahan et al.[9] The predicted cases were subtracted from the reported cases to identify the variability between the model and the reported results. The sum differences were tested via the Wilcoxon Sign Rank test per the following hypotheses and a 95% confidence interval:

- H_o → There is no statistically significant difference between the reported and predicted community COVID-19 case rates per week.

- H_a → There is a statistically significant difference between the reported and predicted community COVID-19 case rates per week.

The data was further analyzed via the same methods following the staggering of observed community cases by one week in comparison to their wastewater SARS-CoV-2 RNA concentrations and flow rate data to see if the model's fit was affected.

Secondly, the researchers pursued the acceptability of linear regression models for the sample communities.[11, 12] During analysis, the relevant variables were transformed logarithmically to normalize their distributions and support the use of a linear regression model. The researchers compared adjusted R^2 to determine model fit for community case rate predictions. A 95% confidence interval was associated with the following hypotheses:

- H_o → A linear model is not statistically significantly effective in predicting Log_{10}(community case rates) using Log_{10}(predicted case counts).
- H_a → A linear model is not statistically significantly effective in predicting Log_{10}(community case rates) using Log_{10}(predicted case counts).

The data was analyzed via the same methods following the staggering of observed community cases by one week in comparison to their wastewater SARS-CoV-2 RNA concentrations and flow rate data to see if the model's fit was affected.

Lastly, the simple prediction models were supplemented with a functioning SEIR model to showcase potential COVID-19 spread through each respective community. The SEIR model leveraged the following equations:

Equation Three: Change in Susceptible Population(s)

$$\Delta Susceptible = \beta * Susceptible * \left(\frac{Infected}{Population}\right)$$

Equation Four: Susceptible Population(s)

$$Susceptible = Susceptible - \Delta Susceptible$$

Equation Five: Exposed Population(s)

$$Exposed = Exposed + \Delta Susceptible - (\alpha * Exposed)$$

Equation Six: Infected Population(s)

$$Infected = Infected + (\alpha * Exposed) - (\gamma * Infected)$$

Equation Seven: Recovered Population(s)

$$Recovered = Recovered + (\gamma * Infected)$$

Various constant assumptions were made including the number of contacts with a COVID-19-positive individual sufficient for infection, beta, the median incubation time of a COVID-19 infection, alpha, and the recovery rate of an infected individual, gamma. Each value was based on the available existing literature from McMahan et al.[9]

Results

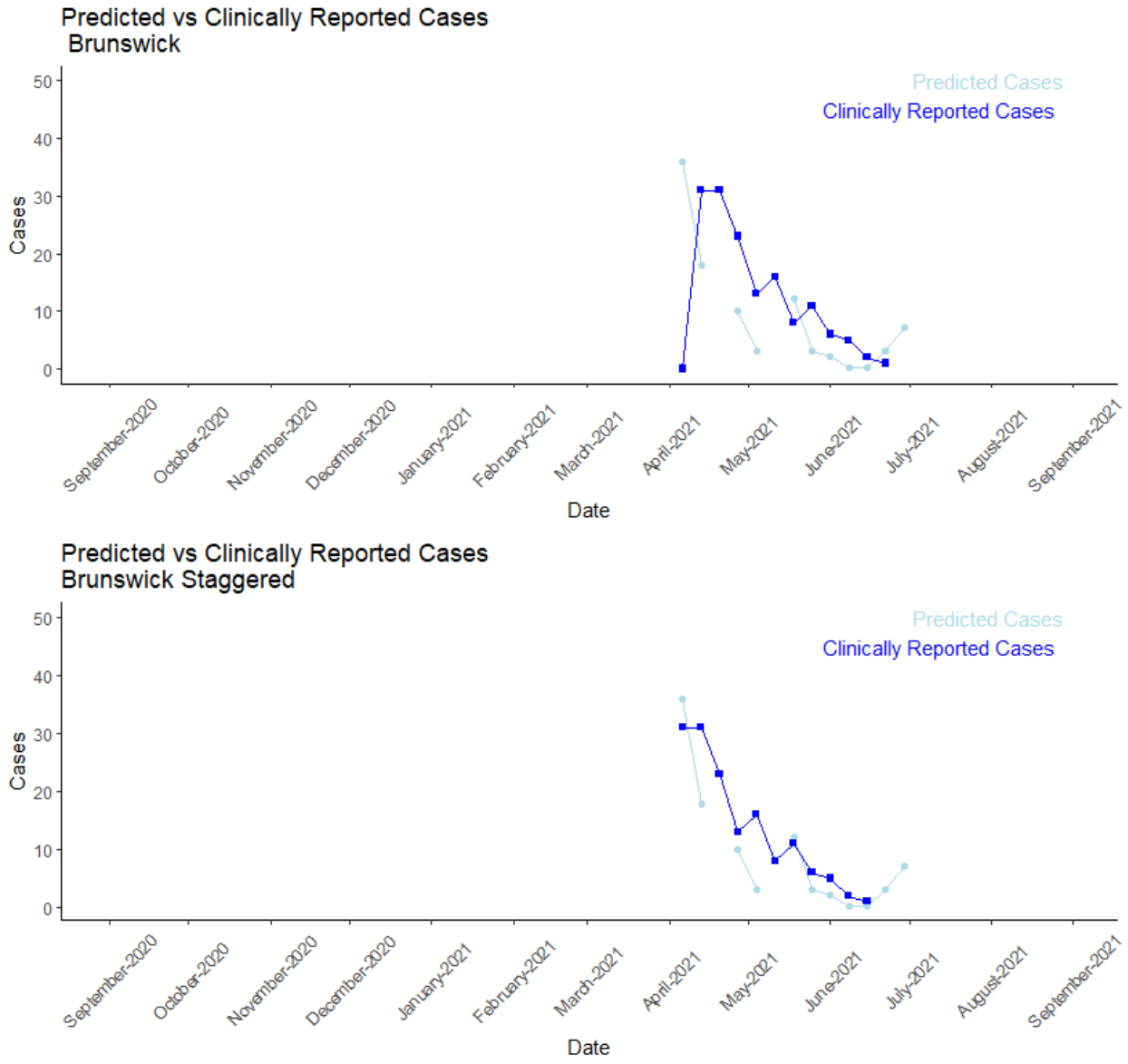
As seen in table 6, Brunswick, East End (Portland) and Westbrook/Gorham all failed to reject the null hypothesis in both 1-week staggered and non-staggered analysis, with p-values well above the 95% confidence interval threshold of 0.05. These three communities, therefore, must accept that there were no statistically significant differences between the predicted cases generated by the simple predictive model and those observed by community clinical diagnostic testing. However, this conclusion was not reflected by Yarmouth. With p-values below the 0.05 alpha, there is significant statistical support to reject the null hypothesis and accept the alternative – there are statistically significant differences between the predicted and clinical

case rates in Yarmouth, with predicted cases being higher than observed cases. A compilation of relevant test statistics can be found in Table 6.

Table 6: The Simple Predictive Model & Wilcoxon Sign Rank Test				
Community	**Non-Staggered**		**Staggered by 1-Week**	
	Test Value	P-value	Test Value	P-value
Brunswick	39	0.2754	36	0.1289
East End (Portland)	618	0.8545	627	0.89902
Westbrook/Gorham	474	0.1156	494	0.2429
Yarmouth	455	0.006141*	406	0.02412*

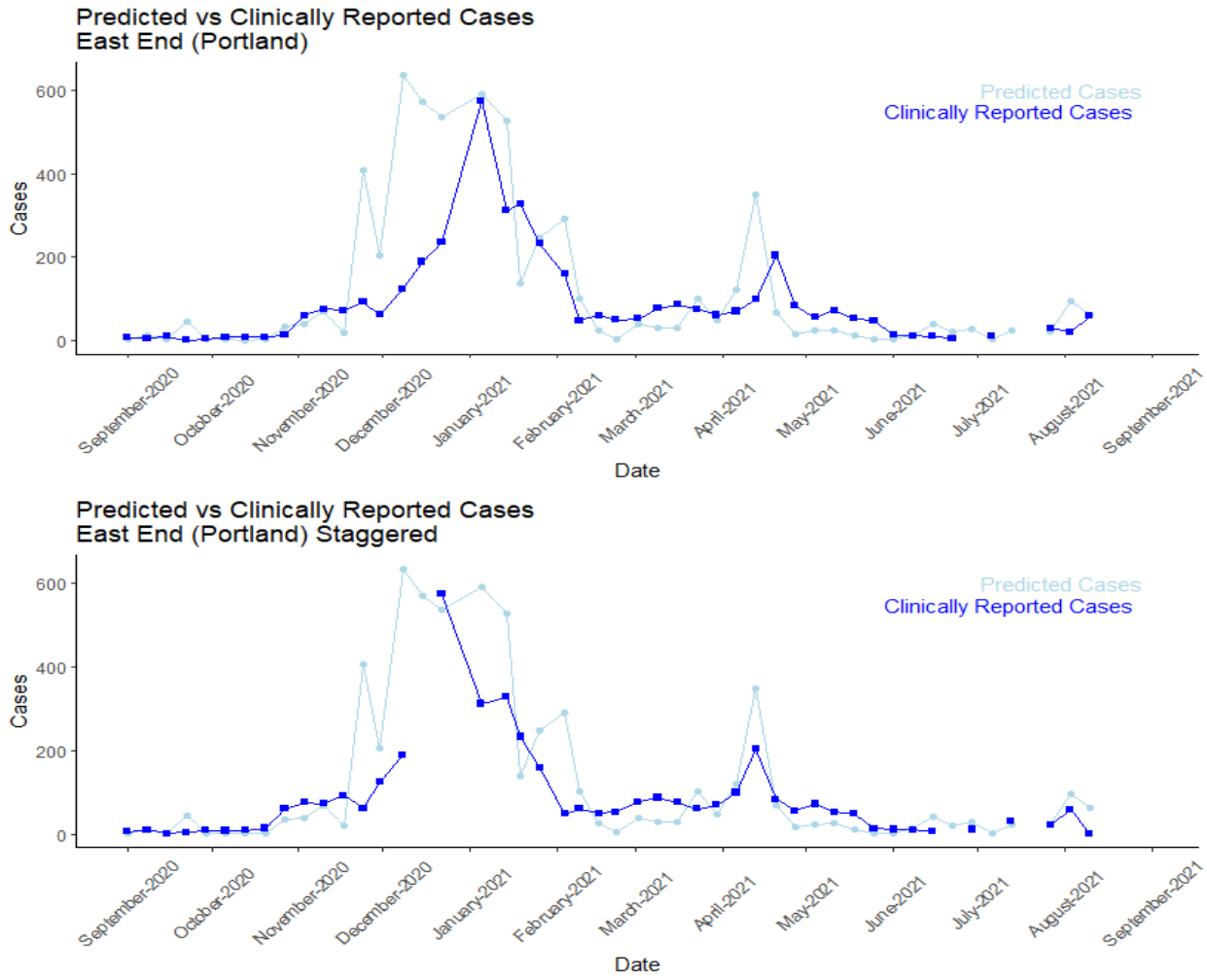
These data can be observed visually via Figure 1, Brunswick, Figure 2, East End (Portland), Figure 3, Westbrook/Gorham, and Figure 4, Yarmouth. Each figure demonstrates differences between predicted case counts and clinically reported cases during a given sampling week.

Figure 1: Brunswick Simple Predictive Model



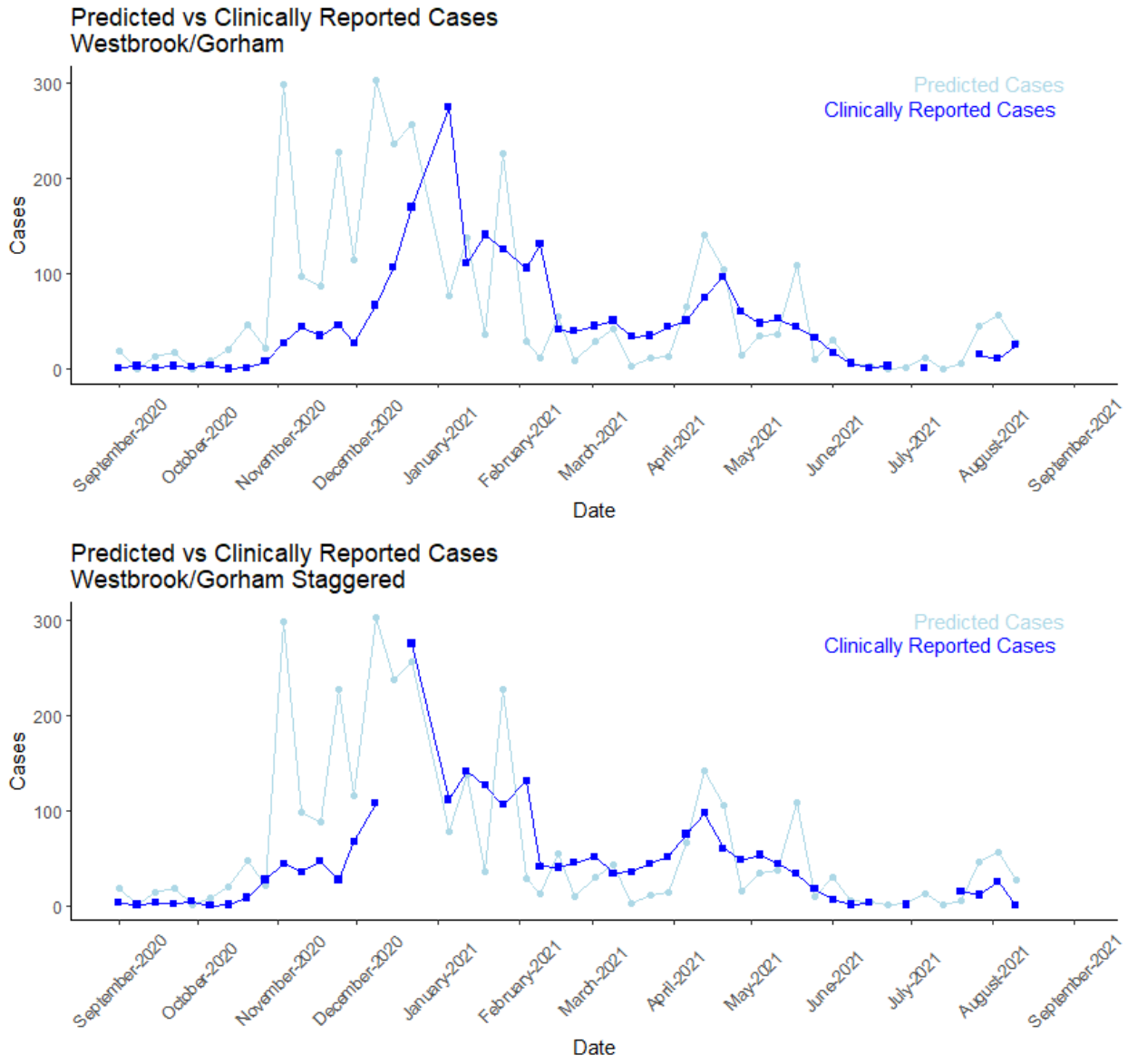
In non-staggered assessments, peaks and valleys in the predicted cases typically precedes similar trends in clinically reported cases. However, the predicted cases became more closely aligned with the clinically reported cases once the 1-week staggered data alignment was considered.

Figure 2: East End (Portland) Simple Predictive Model



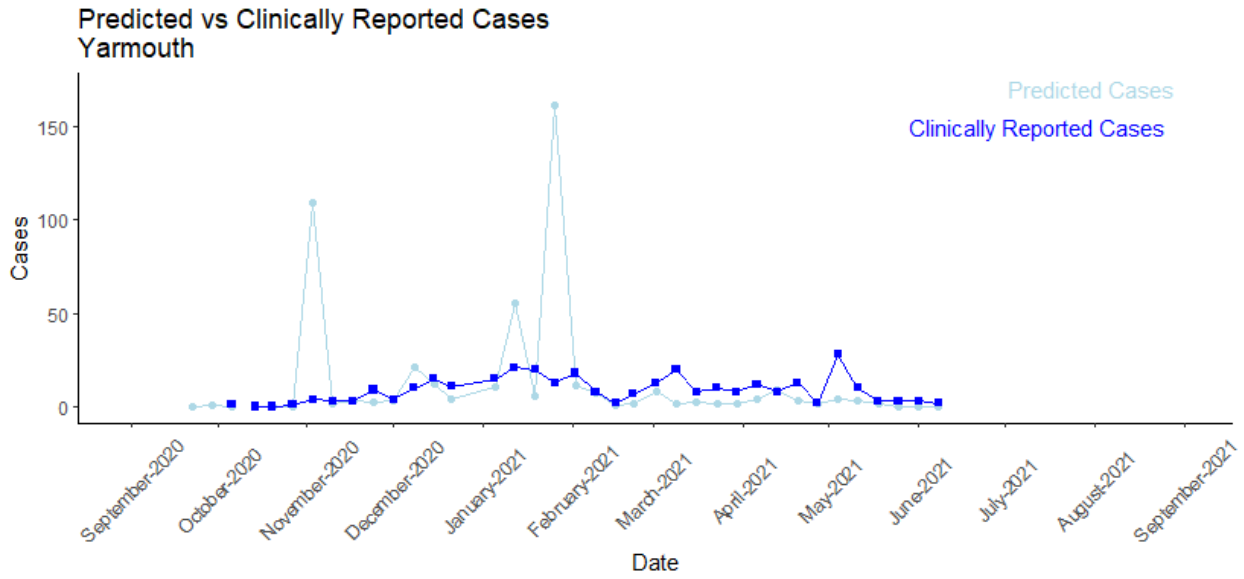
In the non-staggered samples, peaks and valleys in predicted cases preceded similar trends in clinically reported cases. However, like Brunswick, East End (Portland)'s staggered model experienced more closely aligned data between predicted and clinically reported cases.

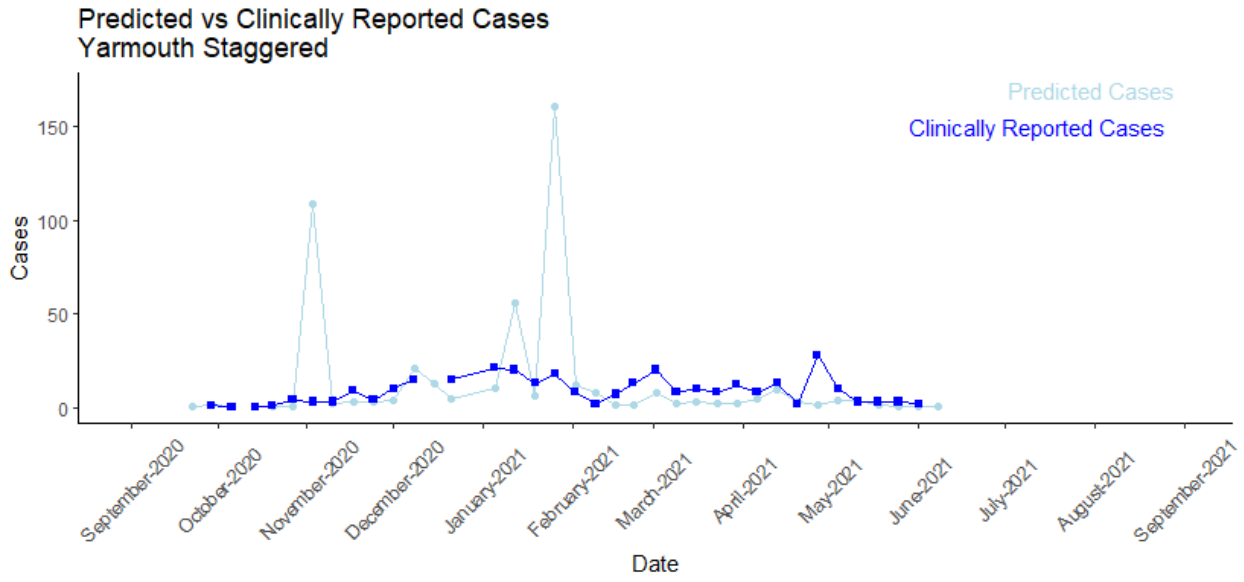
Figure 3: Westbrook/Gorham Simple Predictive Model



In the non-staggered samples, peaks and valleys in predicted cases preceded similar trends in clinically reported cases. Once the 1-week staggered alignment was considered Westbrook/Gorham's predicted and clinically reported case counts were slightly more aligned.

Figure 4: Yarmouth Simple Predictive Model

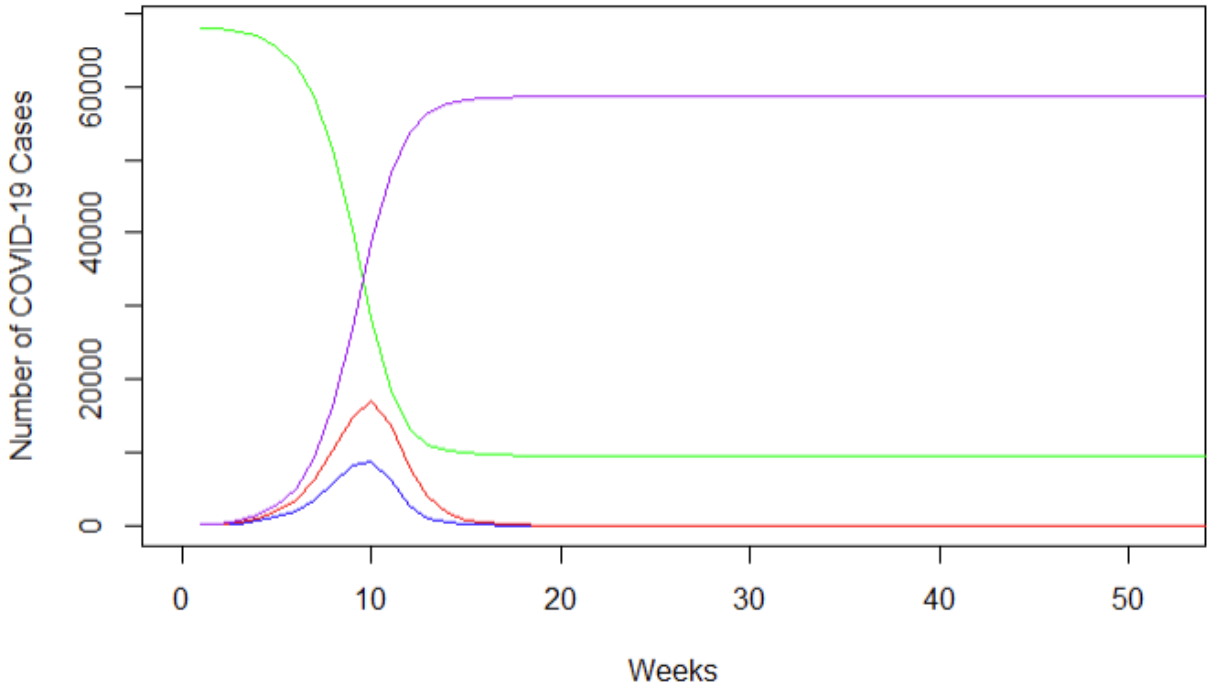




Yarmouth, in both alignment models, experienced a statistically significant difference between predicted case counts and observed clinically reported cases. This phenomenon can be visually observed in the spike in predicted cases in November of 2020 and January of 2021 compared to only slight peaks in clinical cases. These spikes were directly connected with public health warnings sent out by members of the Yarmouth Community Coronavirus Task Force.[35] During this time, cases rose from 7 to 21 in a span of three weeks and wastewater samples yielded accelerating increases in key indicators of COVID-19's presence in the community.[35]

The simple predictive models were paired with customized SEIR simulations that enabled a plug in of initial infected population quantities to better understand the spread of diseases at varied times. Figure 5 demonstrates an example of this model for East End (Portland) assuming an infected population of their mean case rate during data collection, 80 COVID-19 cases.

Figure 5: East End (Portland) SEIR Modeling



In this model, population proportions are tracked over time to understand changes in susceptible (green), exposed (blue), infected (red), and recovered (purple) subgroups. The simulation demonstrates a peak of cases after ~10 weeks with a decrease in spread as recovery from disease states nears the population total. The model does not consider vaccination rates among each community population or the rolling potential of reinfection following the waning of individual immunity over time.[36]

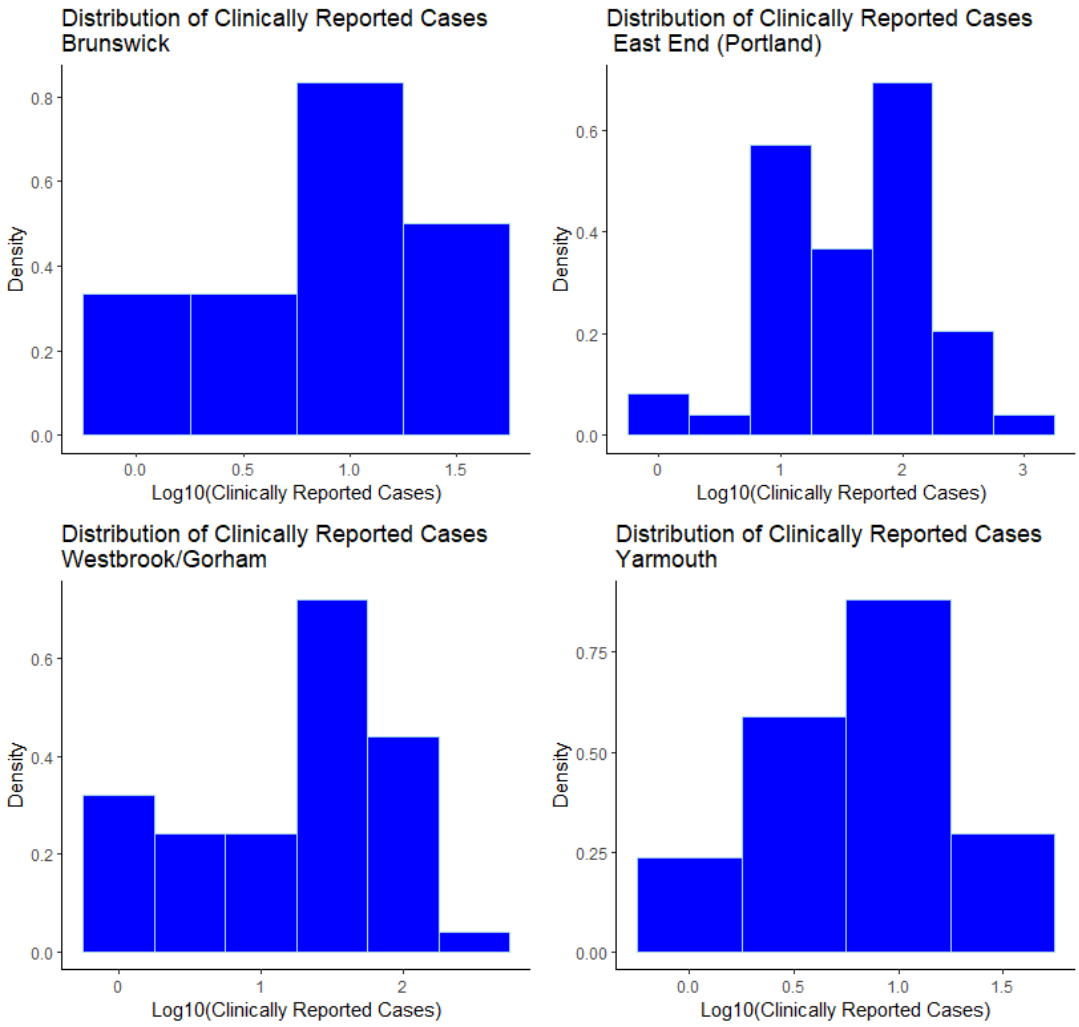
Lastly, the effectivity of linear prediction models was assessed and can be observed in Table 7. All communities showed statistically significant associations between the predicted and the observed case counts except for the unstaggered Brunswick model. When adjusted with a 1-week lag time, Brunswick's predicted cases could account for 86% of the variability observed in the clinically reported cases throughout the study period. East End (Portland) showed medium strength correlations between predictive and clinically reported cases, R^2 values of 0.50 and 0.48 for unstaggered and 1-week staggered analyses respectively. Westbrook/Gorham demonstrated R^2 values of 0.38 and 0.44 for unstaggered and staggered analyses respectively. Lastly, Yarmouth yielded R^2 values of 0.47 and 0.27 for unstaggered and staggered analyses respectively.

Table 7: Linear Regression Model – Log_{10}(Cases) by Log_{10}(Predicted Cases)

Community	Non-Staggered				Staggered by 1-Week			
	Y-Intercept	Slope	Adjusted R2	P-value	Y-Intercept	Slope	Adjusted R2	P-value
Brunswick	0.68	0.15	-0.08	0.591*	0.51	0.67	0.86	1.78E-04
East End (Portland)	0.78	0.54	0.50	6.48E-09	0.76	0.54	0.48	2.62E-08
Westbrook/ Gorham	0.53	0.55	0.38	1.37E-06	0.44	0.6	0.44	1.22E-07
Yarmouth	0.65	0.35	0.47	4.10E-06	0.65	0.29	0.27	1.17E-03

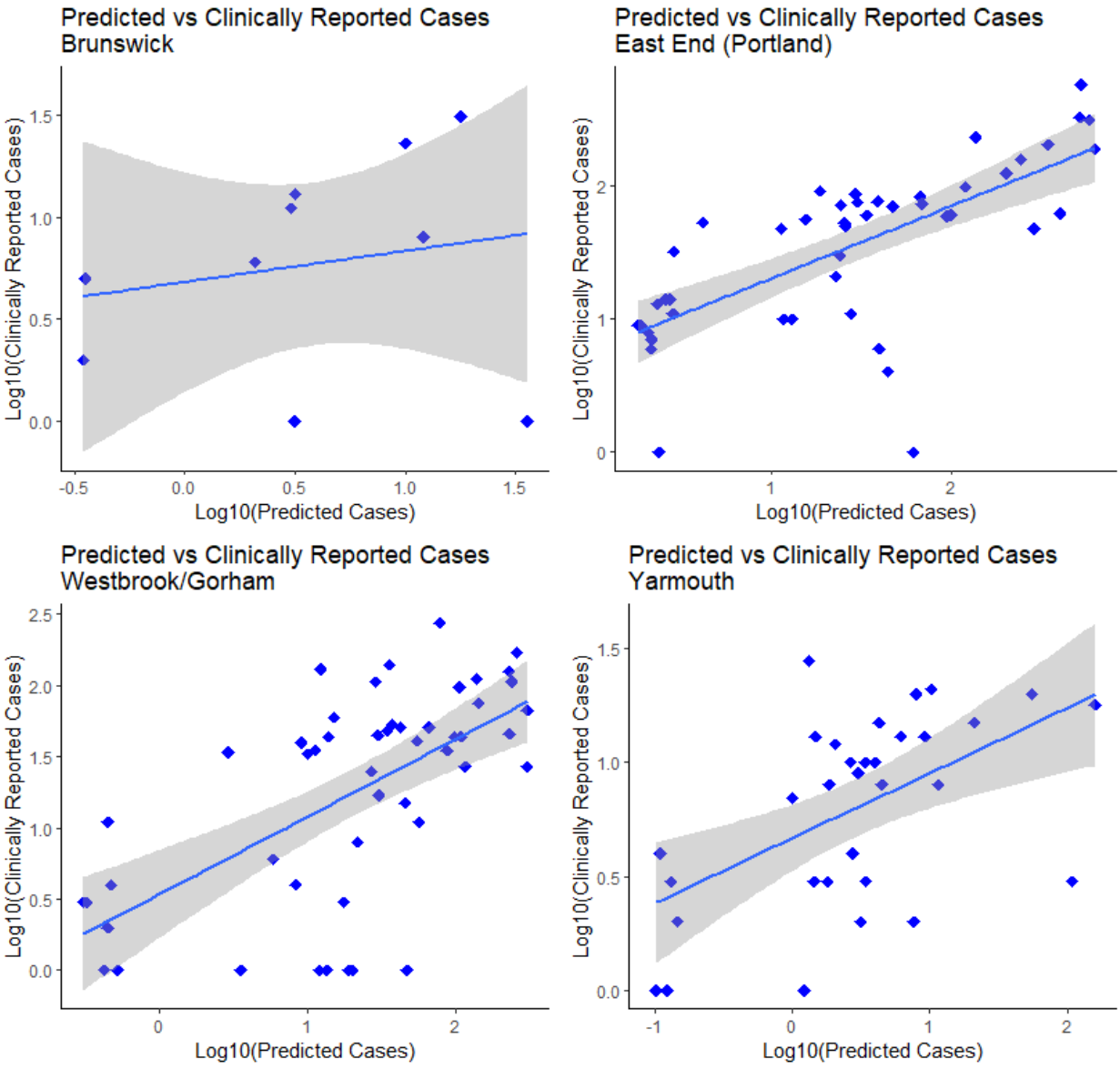
The use-fit of this model can be further understood via the observed data distributions of Log_{10}(Clinical Cases) in Figure 6 and the distribution of Log_{10}(Clinical Cases) by Log_{10}(Predicted Cases) in Figure 7.

Figure 6: Data Distributions of the Linear Models



Following normalization using Log_{10} functions, the clinically reported cases observed reasonably normal distributions – supporting the use of linear regression modeling. The findings in Table 7, regarding the relationship between predicted cases and clinically reported cases, can be visualized in Figure 7.

Figure 7: Linear Regression Modeling



The responsibility for variability in Log_{10}(Clinically Reported Cases) is higher for East End (Portland) and Yarmouth when compared to Westbrook/Gorham and Brunswick. The shaded areas of each graph represent the 95% confidence interval related to where Log_{10}(Clinical Cases) may fall based on Log_{10}(Predicted Cases). Despite 95% confidence, there are still many data points that fall outside of the windows of estimation for each community. However, there is statistically significant relevance to the applicability of linear modeling for predicting COVID-19 case rates across the four assessed Maine communities.

Discussion

WBE's effectiveness requires an extensive cross-functional network of community members, public officials, researchers, public health experts, laboratory personnel and more to achieve beneficial outcomes. WBE supports and optimizes the development of predictive early-warning surveillance that enables communities to respond to their evolving needs based on disease burden.[3-8, 12, 26-27, 28-34, 41-43] Rather than solely relying on traditional clinical testing, communities can practice ongoing cost-effective data collection which offers real time insights into the their populations' health. Additionally, the methods are non-invasive and collect general non-identified community level data – alleviating issues related to sampling bias, limited testing capacity, or untimely reporting of clinical cases.[3, 31, 33]

Its application offers community-level disease detection that includes the ability to capture viral shedding in pre-symptomatic or asymptomatic individuals as well as those who could potentially fail to test positively in clinical diagnostic testing.[1,3,5,7,27,30-31] This functionality could be important for rectifying false negative clinical testing related to COVID-19. The sensitivity, or likelihood of detecting a true positive case, of COVID-19's clinical diagnostic testing varied drastically between its various applications.[37] Confirmed laboratory-repeated testing demonstrated an 85.7% sensitivity, inpatient testing yielded a 95.5% sensitivity and outpatients demonstrated an 89.9% sensitivity.[37] This means that, on average, one out of ten clinically tested positive cases were missed using diagnostic testing.[38] Therefore, the United States, with 103,000,000 confirmed clinical cases, could have theoretically misdiagnosed over 10 million cases in patients who actively sought medical attention.[39] It is important to note, this large miss of possible cases includes only those who sought medical attention and tested

negative – not those who never pursued a test. In parallel, COVID-19 added complexity to clinical diagnosis through its ability to proliferate in a host that exhibited asymptomatic outcomes.[40] A random-effects model leveraged a meta-analysis of 95 studies and found a range of asymptomatic cases between 0.25% in tested populations and 40.50% in case populations.[40] If these asymptomatic individuals test negative, the total U.S. based cases to date is more closely represented by 113-145 million rather than 103 million. This is why predictive supplementary WBE models are important in remedying clinical testing limitations while ensuring increased screening potential and efficiency.

In this study, both the simplified predictive model and a linear regression model proved statistically relevant for predicting community COVID-19 case rates in varied capacity. These findings add to a growing basis of associations between wastewater levels of SARS-CoV-2 genetic materials and reported community cases or incidence.[3-12, 16, 26-34] The simple predictive model put forth by McMahan et al[9] yielded no difference to clinically reported cases with statistical significance for all communities other than Yarmouth in both non-staggered and staggered assessments.

The simple predictive model demonstrated peaks in observed COVID-19 infections after observed rises in predicted case counts in the unstaggered analysis. The 1-week staggered analysis more accurately aligned predicted and observed clinical cases but this likely inhibits the effectivity of a potential 1-week warning regarding possible upcoming clinical case counts. In additional favor for this model, it requires limited technical understanding of predictive modeling or complex equations and only needs continuous WBE as recommended by the CDC.[13, 16]

Yarmouth, the only community that failed to find statistically significant similarities between predicted and observed community cases, experienced COVID-19 outbreaks that resulted in their community taskforce's public alert.[35] This lack of statistical significance can be accredited to many things but certain demographic and socioeconomic considerations have been linked to lowered health-seeking behaviors. Uninsured individuals are less likely to seek necessary care and often experience worse health outcomes than their insured counterparts.[43-44] Additionally, these at-risk populations may choose to avoid care seeking until the point where clinical testing might not be relevant or accurate.[44-45] In parallel, poverty rates have been associated with negative care seeking behaviors.[44] Despite governmental support to increase access to COVID-19 testing and care, poverty may play a negative role in influencing the willingness to seek care when experiencing COVID-19 symptoms.[46] Yarmouth experiences high poverty rates and high rates of uninsured individuals, 10.40% and 6.80% respectively.[14] Further research would be necessary to understand model applicability and alignment with observed infection rates and demographic considerations.

The linear predictive model demonstrated statistical significance in the variability responsibility Log_{10}(COVID-19 Cases) by Log_{10}(Predicted Cases). However, this model also requires normality in data distributions that may not always be present in real life observations. Another constraint is the lack of accuracy in the 95% confidence interval's capture of Log_{10}(COVID-19 Cases) as visualized by Figure 7. Additionally, most models demonstrated medium strength predictive capabilities at best, with the R^2 values nearing 0.50.

The SEIR model is an interesting addition to predictive models in that it can provide simulated projections of sub-population changes over time based on observations captured via

WBE. However, the model is also extremely complicated to write and requires a detailed understanding of differential equations to accomplish. The model developed in this research is extremely basic and therefore has limited applicability for real life use. A more robust model could be created in the future that factors in numerous additional COVID-19 considerations to more accurately describe individual disease states and their respective changes over time. An ideal future state could include a tool that allows community stakeholders to plug in their observed SARS-CoV-2 wastewater RNA, COVID-19 vaccination rates, population size and density, average daily temperature and more.[9, 36]

Conclusion

The effectiveness of the simple predictive and linear models bolsters existing evidence that WBE can play a vital supplementary role in COVID-19 disease surveillance and prediction.[3-8, 12, 26-27, 28-34, 41-43] As WBE continues to serve a more important role in the future of public health surveillance, understanding its potential applications is critical for ensuring the most effective implementation and cost effectiveness of its utilization. The statistical significance of both models demonstrates its capability to serve as an early-warning surveillance rather than a reactive clinical testing response. [3-8, 12, 26-27, 28-34, 41-43] Additionally in its favor, it is a non-invasive method of surveillance and collects data from a generalized non-identified community population which limits potential sampling biases.[3, 31, 33] It can alleviate limited testing capabilities or potentially delayed reporting of clinical diagnostic case counts to public health entities while empowering them to plan ahead for potentially upcoming clinically positive COVID-19 cases.[3, 31, 33] Lastly, with respect to COVID-19, WBE grants the ability to capture

potential infection data for individuals who are pre-symptomatic, asymptomatic, or subject to false negative testing results.[1, 3, 5, 7, 27, 30-31, 37-40]

While clinical or at-home diagnostic testing may remain the standard for identifying positive community cases, implementing WBE programs can supplement this information and provide efficient cost-effective community screening and case prediction in the future. Despite waning COVID-19 rates compared to peak pandemic numbers, variants are still posed to lead to regression in infection prevention and public health.[43] The CDC continues to encourage the monitoring of variants to ensure that the mutations which give rise to new infection opportunities are understood and properly responded to.[47] WBE is becoming an important public health surveillance tool and leveraging it in the future with supplemental predictive models could provide numerous benefits to community level understanding and response related to population health and SARS-CoV-2.

References:

1) Daughton CG. Wastewater surveillance for population-wide Covid-19: The present and future. *Sci Total Environ*. 2020;736:139631. doi:10.1016/j.scitotenv.2020.139631

2) Hu B, Guo H, Zhou P, Shi Z. Characteristics of SARS-CoV-2 and COVID-19. *Nature reviews. Microbiology*. 2021;19(3):141-154. https://www.ncbi.nlm.nih.gov/pubmed/33024307. doi: 10.1038/s41579-020-00459-7.

3) Mainardi PH, Bidoia ED. Challenges and emerging perspectives of an international SARS-CoV-2 epidemiological surveillance in wastewater. *An Acad Bras Cienc*. 2021;93(suppl 4):e20210163. Published 2021 Dec 1. doi:10.1590/0001-3765202120210163

4) Giraud-Billoud M, Cuervo P, Altamirano JC, et al. Monitoring of SARS-CoV-2 RNA in wastewater as an epidemiological surveillance tool in Mendoza, Argentina. *Sci Total Environ*. 2021;796:148887. doi:10.1016/j.scitotenv.2021.148887

5) Aguiar-Oliveira ML, Campos A, R Matos A, et al. Wastewater-Based Epidemiology (WBE) and Viral Detection in Polluted Surface Water: A Valuable Tool for COVID-19 Surveillance-A Brief Review. *Int J Environ Res Public Health*. 2020;17(24):9251. Published 2020 Dec 10. doi:10.3390/ijerph17249251

6) Keshaviah A, Hu XC, Henry M. Developing a Flexible National Wastewater Surveillance System for COVID-19 and Beyond [published correction appears in Environ Health Perspect. 2021 May;129(5):59001]. *Environ Health Perspect*. 2021;129(4):45002. doi:10.1289/EHP8572

7) Kirby AE, Walters MS, Jennings WC, et al. Using Wastewater Surveillance Data to Support the COVID-19 Response - United States, 2020-2021. *MMWR Morb Mortal Wkly Rep*. 2021;70(36):1242-1244. Published 2021 Sep 10. doi:10.15585/mmwr.mm7036a2

8) Jakariya M, Ahmed F, Islam MA, et al. Wastewater-based epidemiological surveillance to monitor the prevalence of SARS-CoV-2 in developing countries with onsite sanitation facilities. *Environ Pollut*. 2022;311:119679. doi:10.1016/j.envpol.2022.119679

9) McMahan CS, Self S, Rennert L, et al. COVID-19 wastewater epidemiology: a model to estimate infected populations. *Lancet Planet Health*. 2021;5(12):e874-e881. doi:10.1016/S2542-5196(21)00230-8

10) Phan T, Brozak S, Pell B, et al. A simple SEIR-V model to estimate COVID-19 prevalence and predict SARS-CoV-2 transmission using wastewater-based surveillance data. *The Science of the total environment*. 2023;857(Pt 1):159326. doi: 10.1016/j.scitotenv.2022.159326.

11) Zhu Y, Oishi W, Maruo C, et al. COVID-19 case prediction via wastewater surveillance in a low-prevalence urban community: A modeling approach. *Journal of water and health*. 2022;20(2):459-470. doi: 10.2166/wh.2022.183.

12) Zhao L, Zou Y, Li Y, et al. Five-week warning of COVID-19 peaks prior to the Omicron surge in Detroit, Michigan using wastewater surveillance. *Sci Total Environ*. 2022;844:157040. doi:10.1016/j.scitotenv.2022.157040

13) Tracking SARS-CoV-2 variants. World Health Organization Web site. https://www.who.int/activities/tracking-SARS-CoV-2-variants. Updated 2022. Accessed Jan 9th, 2023.

14) U.S. Census Bureau QuickFacts. (2021). Population Survey. U.S. Census Bureau. https://www.census.gov/quickfacts/fact/table/portlandcitymaine,gorhamtowncumberlandcountymaine, westbrookcitymaine,yarmouthtowncumberlandcountymaine,brunswicktowncumberlandcountymaine/PST045221 (Accessed on January 31st, 2023)

15) Wu Y, Kang L, Guo Z, Liu J, Liu M, Liang W. Incubation Period of COVID-19 Caused by Unique SARS-CoV-2 Strains: A Systematic Review and Meta-analysis. *JAMA Netw Open.* 2022;5(8):e2228008. doi:10.1001/jamanetworkopen.2022.28008

16) Anneser E, Riseberg E, Brooks YM, Corlin L, Stringer C. Modeling the relationship between SARS-CoV-2 RNA in wastewater or sludge and COVID-19 cases in three new england regions. *Journal of water and health*. 2022;20(5):816-828. https://www.ncbi.nlm.nih.gov/pubmed/35635775. doi: 10.2166/wh.2022.013.

17) RStudio Team (2022). *RStudio: Integrated Development for R.* RStudio, PBC, Boston, MA URL http://www.rstudio.com/. Version 2022.7.0+548

18) Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." _Journal of Open Source Software_, *4*(43), 1686. doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.

19) Waring E, Quinn M, McNamara A, Arino de la Rubia E, Zhu H, Ellis S (2022). _skimr: Compact and Flexible summaries of Data_. R package version 2.1.4, <https://CRAN.R-project.org/package=skimr>.

20) Kassambara A (2020). _ggpubr: 'ggplot2' Based Publication Ready Plots_. R package version 0.4.0, <https://CRAN.R-project.org/package=ggpubr>.

21) Karline Soetaert, Thomas Petzoldt, R. Woodrow Setzer (2010). Solving Differential Equations in R: Package deSolve. Journal of Statistical Software,33(9), 1--25. doi:10.18637/jss.v033.i09

22) Koenker R (2022). _quantreg: Quantile Regression_. R package version 5.94,<https://CRAN.R-project.org/package=quantreg>.

23) R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

24) Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New Yor k. ISBN 0-387-95457-0

25) Hadley Wickham (2007). Reshaping Data with the reshape Package.Journal of Statistical Software, 21(12), 1-20. URL http://www.jstatsoft.org/v21/i12/.

26) Weidhaas J, Aanderud ZT, Roper DK, et al. Correlation of SARS-CoV-2 RNA in wastewater with COVID-19 disease burden in sewersheds. *Sci Total Environ*. 2021;775:145790. doi:10.1016/j.scitotenv.2021.145790

27) Betancourt WQ, Schmitz BW, Innes GK, et al. COVID-19 containment on a college campus via wastewater-based epidemiology, targeted clinical testing and an intervention. *Sci Total Environ*. 2021;779:146408. doi:10.1016/j.scitotenv.2021.146408

28) Kumar M, Joshi M, Patel AK, Joshi CG. Unravelling the early warning capability of wastewater surveillance for COVID-19: A temporal study on SARS-CoV-2 RNA detection and need for the escalation. *Environ Res*. 2021;196:110946. doi:10.1016/j.envres.2021.110946

29) Wurtzer S, Marechal V, Mouchel JM, et al. Evaluation of lockdown effect on SARS-CoV-2 dynamics through viral genome quantification in waste water, Greater Paris, France, 5 March to 23 April 2020. *Euro Surveill*. 2020;25(50):2000776. doi:10.2807/1560-7917.ES.2020.25.50.2000776

30) Ahmed W, Tscharke B, Bertsch PM, et al. SARS-CoV-2 RNA monitoring in wastewater as a potential early warning system for COVID-19 transmission in the community: A temporal case study. *Sci Total Environ*. 2021;761:144216. doi:10.1016/j.scitotenv.2020.144216

31) D'Aoust PM, Graber TE, Mercier E, et al. Catching a resurgence: Increase in SARS-CoV-2 viral RNA identified in wastewater 48 h before COVID-19 clinical tests and 96 h before hospitalizations. *Sci Total Environ*. 2021;770:145319. doi:10.1016/j.scitotenv.2021.145319

32) Lodder W, de Roda Husman AM. SARS-CoV-2 in wastewater: potential health risk, but also data source. *Lancet Gastroenterol Hepatol*. 2020;5(6):533-534. doi:10.1016/S2468-1253(20)30087-X

33) Randazzo W, Truchado P, Cuevas-Ferrando E, Simón P, Allende A, Sánchez G. SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area. *Water Res*. 2020;181:115942. doi:10.1016/j.watres.2020.115942

34) Saguti F, Magnil E, Enache L, et al. Surveillance of wastewater revealed peaks of SARS-CoV-2 preceding those of hospitalized patients with COVID-19. *Water Res*. 2021;189:116620. doi:10.1016/j.watres.2020.116620

35) Hoey D. Yarmouth detects 'rapid rise' of COVID-19 markers in wastewater. Press Herald Web site. https://www.pressherald.com/2021/01/17/yarmouth-detects-rapid-rise-of-covid-19-markers-in-wastewater/. Updated 2021. Accessed Feb 28th, 2023.

36) Levin EG, Lustig Y, Cohen C, et al. Waning Immune Humoral Response to BNT162b2 Covid-19 Vaccine over 6 Months. *N Engl J Med*. 2021;385(24):e84. doi:10.1056/NEJMoa2114583

37) Kortela E, Kirjavainen V, Ahava MJ, et al. Real-life clinical sensitivity of SARS-CoV-2 RT-PCR test in symptomatic patients. *PLoS One*. 2021;16(5):e0251661. Published 2021 May 21. doi:10.1371/journal.pone.0251661

38) West CP, Montori VM, Sampathkumar P. COVID-19 Testing: The Threat of False-Negative Results. *Mayo Clin Proc*. 2020;95(6):1127-1129. doi:10.1016/j.mayocp.2020.04.004

39) United states of america situation. World Health Organization Web site. https://covid19.who.int/region/amro/country/us. Updated 2022. Accessed Mar 3rd, 2023.

40) Ma Q, Liu J, Liu Q, et al. Global Percentage of Asymptomatic SARS-CoV-2 Infections Among the Tested Population and Individuals With Confirmed COVID-19 Diagnosis: A Systematic Review and Meta-analysis. *JAMA Netw Open*. 2021;4(12):e2137257. Published 2021 Dec 1. doi:10.1001/jamanetworkopen.2021.37257

41) McCall C, Wu H, Miyani B, Xagoraraki I. Identification of multiple potential viral diseases in a large urban center using wastewater surveillance. Water Res. 2020 Oct 1;184:116160. doi: 10.1016/j.watres.2020.116160. Epub 2020 Jul 7. PMID: 32738707; PMCID: PMC7342010.

42) Layton BA, Kaya D, Kelly C, et al. Evaluation of a Wastewater-Based Epidemiological Approach to Estimate the Prevalence of SARS-CoV-2 Infections and the Detection of Viral Variants in Disparate Oregon Communities at City and Neighborhood Scales. *Environ Health Perspect*. 2022;130(6):67010. doi:10.1289/EHP10289

43) Lu E, Ai Y, Davis A, et al. Wastewater surveillance of SARS-CoV-2 in dormitories as a part of comprehensive university campus COVID-19 monitoring. *Environ Res*. 2022;212(Pt E):113580. doi:10.1016/j.envres.2022.113580

44) McWilliams JM. Health consequences of uninsurance among adults in the United States: recent evidence and implications. Milbank Q. 2009 Jun;87(2):443-94. doi: 10.1111/j.1468-0009.2009.00564.x. PMID: 19523125; PMCID: PMC2881446.

45) Tolbert J. What issues will uninsured people face with testing and treatment for COVID-19? KFF Web site. https://www.kff.org/coronavirus-covid-19/fact-sheet/what-issues-will-uninsured-people-face-with-testing-and-treatment-for-covid-19/#:~:text=People%20who%20are%20uninsured%20will,to%20cost%20(Figure%202). Updated 2020. Accessed Mar 3rd, 2023.

46) Biden-harris administration requires insurance companies and group health plans to cover the cost of at-home COVID-19 tests, increasing access to free tests. U.S. Department of Health & Human Services Web site. https://www.hhs.gov/about/news/2022/01/10/biden-harris-administration-requires-insurance-companies-group-health-plans-to-cover-cost-at-home-covid-19-tests-increasing-access-free-tests.html. Updated 2022. Accessed Mar 3rd, 2023.

47) Monitoring variants. Center for Disease Control and Prevention Web site. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html. Updated 2022. MAr 3rd, 2023.

Appendix

Table 8 – Missing Data Points					
Date	**BRSV**	**Percent Recovery**	**Cumulative Cases**	**Community Infection Rates**	**Copies Per Day**
1/26/2020	EE, WB	EE, WB	EE, WB	EE, WB	EE, WB
7/28/2020	EE, WB	EE, WB	N/A	N/A	N/A
8/4/2020	EE, WB	EE, WB	N/A	N/A	N/A
8/18/2020	EE, WB	EE, WB	N/A	N/A	N/A
8/25/2020	EE, WB	EE, WB	N/A	N/A	N/A
9/1/2020	EE, WB	EE, WB	N/A	N/A	N/A
9/8/2020	EE, WB	EE, WB	N/A	N/A	N/A
9/15/2020	EE, WB	EE, WB	N/A	N/A	N/A
9/22/2020	EE, WB, YR	EE, WB, YR	YR	YR	N/A
9/29/2020	EE, WB, YR	EE, WB, YR	N/A	YR	N/A
10/6/2020	EE, WB, YR	EE, WB, YR	N/A	N/A	N/A
10/13/2020	EE, WB, YR	EE, WB, YR	N/A	N/A	YR
10/20/2020	EE, WB, YR	EE, WB, YR	N/A	N/A	N/A
10/27/2020	EE, WB, YR	EE, WB, YR	N/A	N/A	N/A
11/3/2020	EE, WB, YR	EE, WB, YR	N/A	N/A	N/A
11/10/2020	EE, WB, YR	EE, WB, YR	N/A	N/A	N/A
11/17/2020	EE, WB, YR	EE, WB, YR	N/A	N/A	N/A
11/24/2020	EE, WB, YR	EE, WB, YR	N/A	N/A	N/A
11/30/2020	EE, WB, YR	EE, WB, YR	N/A	N/A	N/A
12/8/2020	EE, WB, YR	EE, WB, YR	N/A	N/A	N/A
12/15/2020	EE, WB, YR	EE, WB, YR	N/A	N/A	N/A
4/20/2021	N/A	N/A	N/A	N/A	BR
5/11/2021	N/A	N/A	N/A	N/A	BR
6/29/2021	N/A	N/A	EE, WB, BR	EE, WB, BR	N/A
7/13/2021	N/A	N/A	EE, WB	EE, WB	N/A
7/20/2021	EE	EE	EE	EE, WB	EE, WB
8/10/2021	EE, WB	EE, WB	N/A	N/A	N/A

EE – East End (Portland)
YR – Yarmouth
WB/GR – Westbrook/Gorham
BR – Brunswick
N/A – Not Applicable

Despite attempts for remediation through site contact, the above table illustrates specific sampling dates from which key variable data was missing. Data missing for BRSV or percent recovery would result in no capability to calculate theoretical SARS-CoV-2 RNA recovery levels. Missing community infection rates would negate statistical utilization of that site during the date range – as the dependent variable is unknown. Lastly, missing copies per day were due to missing wastewater flow rate data that was necessary to perform these calculations. This also would result in the lost of a test variable and subsequent omittance from our analysis.

Table 9 – Undetected PCR Recovery Samples			
Date	**City(s)**	**Date**	**City(s)**
9/15/2020	EE	2/22/2021	EE
9/22/2020	YR	6/8/2021	BR, YR
9/29/2020	EE, WB/GR	6/15/2021	BR
10/6/2020	EE, YR	6/22/2021	WB/GR
10/20/2020	EE, YR	7/13/2021	WB/GR
10/27/2020	YR	7/20/2021	EE

EE – East End (Portland)
YR – Yarmouth
WB/GR – Westbrook/Gorham
BR – Brunswick

The above data reflects dates from which data points were not present for SARS-CoV-2 RNA PCR recovery. These values were shared as <762 copies/liter or simply 0. As previously mentioned, the data was corrected during cleaning to reflect the CDC's recommendations to leverage half of 762 as a stand in for undetected samples.[21]

Table 10 – Columns Omitted From Data Analysis	
Title	Function
Initial.vol.of.sample..ml..x	Test specific control measure to ensure consistency
concentrate.vol..ul.	Test specific control measure to ensure consistency
DNA.extraction.vol..ul.	Test specific control measure to ensure consistency
elution.vol..ul.	Test specific control measure to ensure consistency
vol.in.qPCR.rxn..ul.	Test specific control measure to ensure consistency
ct.value	rt-PCR value from the submitted wastewater sample
brsv.theoretical	Test specific control value to calculate actual recovery based on experimental design
town	Duplicate data reflected under Name
initial.vol.of.sample..ml..y	Volume of sample received by wastewater treatment plant
Cq.value	Fluorescent value related to the number of cycles required to achieve threshold levels
lower.value	Lower limit of detection in copies/L
upper.value	Upper limit of quantification in copies/L

The above columns were removed from the R-Studio database to allow for ease of data manipulation and concentration on the variables of interest.