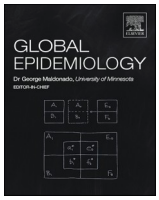
Global Epidemiology 6 (2023) 100124

Contents lists available at ScienceDirect

Global Epidemiology

journal homepage: www.sciencedirect.com/journal/global-epidemiology



Estimating the prevalence of COVID-19 cases through the analysis of SARS-CoV-2 RNA copies derived from wastewater samples from North Dakota

Bong-Jin Choi [a,*], Scott Hoselton [b], Grace N. Njau [c], I.G.C.G. Idamawatta [d], Paul Carson [e], John McEvoy [b]

[a] Department of Statistics and Department of Public Health, North Dakota State University, United States of America
[b] Department of Microbiological Sciences, North Dakota State University, United States of America
[c] North Dakota Department of Health, United States of America
[d] Department of Statistics, North Dakota State University, United States of America
[e] Center for Immunization Research and Education (CIRE), Department of Public Health, North Dakota State University, United States of America

ARTICLE INFO

Keywords:
SARS-CoV-2
COVID-19
Quantile regression
K-nearest neighbor regression
Mean absolute percentage error
SIR model

ABSTRACT

The SARS-CoV-2 virus was first detected in December 2019, which prompted many researchers to investigate how the virus spreads. SARS-CoV-2 is mainly transmitted through respiratory droplets. Symptoms of the SARS-CoV-2 virus appear after an incubation period. Moreover, the asymptomatic infected individuals unknowingly spread the virus. Detecting infected people requires daily tests and contact tracing, which are expensive. The early detection of infectious diseases, including COVID-19, can be achieved with wastewater-based epidemiology, which is timely and cost-effective. In this study, we collected wastewater samples from wastewater treatment plants in several cities in North Dakota and then extracted viral RNA copies. We used log-RNA copies in the model to predict the number of infected cases using Quantile Regression (QR) and K-Nearest Neighbor (KNN) Regression. The model's performance was evaluated by comparing the Mean Absolute Percentage Error (MAPE). The QR model performs well in cities where the population is >10000. In addition, the model predictions were compared with the basic Susceptible-Infected-Recovered (SIR) model which is the golden standard model for infectious diseases.

Introduction

The SARS-CoV-2 virus emerged in December 2019 and has severely affected the world. Scientists and researchers have been working tirelessly to understand and combat the virus as it spreads rapidly around the world. One critical aspect of the virus that has garnered much attention is its transmission. Transmission of SARS CoV-2 is thought to be principally through respiratory droplets, although there is growing evidence that airborne transmission may also play a role [1,2]. Some individuals may remain asymptomatic even after the incubation period, posing significant challenges for contact tracing and frequent testing that can be both time-consuming and costly [3–8]. Therefore, early detection and containment of COVID-19 disease is essential for preventing transmission and limiting its spread.

Wastewater-based epidemiology (WBE) has emerged as a promising approach to detecting the presence of infectious agents including SARS-CoV-2, in a cost-effective manner. Wastewater surveillance involves analyzing sewage samples from communities, which can provide information on the prevalence of diseases within that population [9]. The approach has been used in the past to detect outbreaks of diseases such as polio and hepatitis A, and it is now being applied to COVID-19 [10–13]. The virus can be shedded in the feces, urine, blood, and sputum of infected individuals [14–19] and these shedded viral particles can be detected in wastewater samples. Therefore, wastewater surveillance can capture the presence of the SARS-CoV-2 virus even in asymptomatic individuals, enabling early detection of the disease before it spreads further [20–23].

Moreover, wastewater surveillance is a timely and cost-effective method for detecting infectious agents. Traditional methods of testing, such as individual testing or population-wide testing, can be expensive

* Corresponding author.
E-mail address: bongjin.choi@ndsu.edu (B.-J. Choi).

https://doi.org/10.1016/j.gloepi.2023.100124
Received 13 September 2023; Received in revised form 10 October 2023; Accepted 10 October 2023
Available online 12 October 2023
2590-1133/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

and time-consuming, requiring significant resources. In contrast, wastewater surveillance can provide information on the prevalence of the disease in an entire community at a fraction of the cost of individual testing. The cost of COVID-19 testing can vary based on the type of the test, healthcare provider, and location. Personal tracking tests involve individual testing, where people are tested for COVID-19 using methods such as PCR (polymerase chain reaction) tests, antigen tests, or antibody tests. On average, PCR tests are considered more accurate but may be more expensive than rapid antigen tests. The cost of clinical testing is estimated to be $5.80 per person per year in Latin America and the Caribbean from September 2020 to September 2021 [24].

The cost of wastewater sample analysis can be more cost-effective when considering large populations, as it allows monitoring of a community's infection trends without testing each individual separately. The amount charged can fluctuate relative to the magnitude and regularity of screening, the technology employed, and the research laboratory administering the examination. However, it's considered a more economical approach for tracking COVID-19 in large groups of people. The cost of wastewater testing per person per year is estimated as $0.50 [24]. Thus, researchers and public health entities are increasingly interested in the potential for surveillance of pathogens such as SARS-CoV-2 through routine or episodic testing of wastewater.

Background

SARS CoV-2 has been shown to be reliably shed in the stool of infected patients for up to seven months [25]. A study published by Park et al. [4] investigated gastrointestinal symptoms of COVID-19 patients with mild and asymptomatic symptoms, as well as the shredding of the SARS-CoV-2 virus into their feces. Zhang et al. [26] performed a molecular and serological investigation of COVID-19 patients and found that viral RNA can be detected in anal swabs and blood samples when viral nucleotide has not been detected in oral swabs. Xu et al. [27] clinically investigated the characteristics of pediatric SARS-CoV-2 infection in 10 children and found that there is potential for fecal viral excretion and that testing contaminated water is a viable method for the detection of viral loads. Van Doorn et al. [28] performed a systematic review and meta-analysis of studies based on the fecal-oral transmission of the SARS-CoV-2 virus and found that testing stool samples or anal swabs are more reliable than testing respiratory samples. Additionally, Cheung et al. [29] conducted a systematic review and meta-analysis, and a cohort study and found that stool samples were positive for virus RNA despite negative respiratory samples. Therefore, wastewater surveillance is a good approach to detecting the SARS-CoV-2 virus since it does not rely on the person having symptoms, nor presenting with illness. However, the likelihood of transmission of SARS-CoV-2 virus from sewage-contaminated water or food is very low as reported by Jones et al. [30]. Furthermore, the replicability of SARS-CoV-2 virus is very limited as mentioned by Cerrada-Romero et al. [31].

Furthermore, there are several studies in which wastewater-based epidemiology is used for estimating COVID-19 cases [32–36]. Ahmed et al. [37] collected untreated wastewater samples from one suburban pumping station and two wastewater treatment plant catchments in urban areas in Southeast Queensland and developed a simple equation to estimate positive cases based on the total number of viral RNA copies in wastewater and the number of SARS-CoV-2 RNA copies shed in stool by infected individuals. McMahan et al. [38] implemented a Susceptible-Exposed-Infected-Recovered (SEIR) model to predict the number of infected individuals using the SARS-CoV-2 RNA mass rates in sewage samples. Fazil et al. [39] proposed a stochastic wastewater-based SEIR model, which is similar to the SEIR model proposed by McMahan et al. [38]

Objective

During an outbreak of infectious disease, hospitals need to expand their capacity by adding emergency facilities and increasing medical supplies, staff, and medications. Control measures such as mask mandates, vaccine mandates, social distancing, and business closures are difficult to implement and often controversial. Government and public health authorities must have accurate estimates of disease incidence and prevalence in order to make sound policy decisions. As noted earlier, traditional surveillance methods are severely limited in estimating disease burden when it comes to SARS-CoV-2 transmission due to asymptomatic spread, home testing, and delays in reporting. This study aims to provide further evidence of the utility of wastewater testing for SARS CoV-2 virus to estimate the burden of infection in selected North Dakota populations during key periods of the pandemic, using the K-Nearest Neighbor (KNN) Algorithm [40] and Quantile regression (QR) [41] compared with the Susceptible-Infected-Recovered (SIR) model [42]. By doing so, we can provide an effective method for estimating the number of infected individuals during an outbreak in different populations.

Data collection, viral isolation, and quantification

Wastewater samples were obtained from the intake stream of wastewater treatment plants in various cities in North Dakota. The GLS compact samplers (Teledyne ISCO) were used to collect 24-h composite samples. Subsequently, 250 ml of the composite samples were treated at 60 °C for 30 min and then centrifuged at 4500 $\times g$ for 10 min at room temperature. The supernatant (100 ml) was collected and HCl was added to adjust the samples to pH 4. Samples were vacuum filtered through 0.45 μm MCE membranes (Millipore Sigma). RNeasy PowerMicrobiome kits (Qiagen) were used to extract total RNA from the membranes. Samples were eluted in 50 μl of elution buffer and 5 μl was used for quantitative real-time PCR (Polymerase Chain Reaction). For PCR reactions, we used N1 primers from the 2019-nCoV RUO kit by Integrated DNA Technologies (IDT), along with the Applied Biosystems TaqPath 1-Step RT-qPCR Master Mix. To construct the standard curve, we employed the 2019-nCoV_N Positive control plasmid from IDT.

Statistical methodology

The primary aim of our research is to estimate the number of COVID-19 cases through the analysis of SARS-CoV-2 RNA copies derived from wastewater samples. To achieve this, we will use a combination of statistical modeling and quantitative analysis. For a selected time, we will employ three distinct models to predict the number of infected cases based on natural logarithm-transformed data of viral RNA copies as the independent variable. These models include the KNN regression model, the QR model, and the basic SIR model. The following section summarizes some basic ideas behind each of the above models.

K-nearest neighbor algorithm (KNN)

KNN algorithm is a supervised machine learning algorithm, which can be used to solve classification and regression problems. The KNN model was first developed by Evelyn Fix and Joseph Hodges in 1951 [40]. In this method, the similarities of the features of k-nearest neighbors are used to predict the outcome variable. The Euclidean distance measure, Manhattan distance, and Minkowski distance methods are commonly used to calculate the similarities of nearest neighbors. Suppose $\widehat{y}$ is the predicted target variable value for a given input data point and y_i is the target variable value of the i^{th} nearest neighbor to the input data point. If k is the number of nearest neighbors to consider, then,

$$\widehat{y} = \frac{1}{k}\sum_{i=1}^{k} y_i \tag{1}$$

To perform KNN regression, first, we need to calculate the distance between the input data point and all the data points in the dataset. Then, using the distance metric, select k-nearest neighbors and use their target

variable values to calculate $\hat{y}$ for the input data point. The smaller the k value, the more likely the model will overfit, and the larger the k value, the more likely it will underfit. Therefore, we performed 10-fold cross-validation to select the optimal k using the *tidymodels* package in R. In most of the cities $k = 9, 10,$ or 11 were observed as optimal k. Therefore, we used $k = 10$ for all cities for convenience.

Quantile regression (QR)

Quantile regression [41] is a type of regression analysis that models the relationship between the independent variables and the quantile of the dependent variable. Unlike ordinary least-square regression models, QR models need no assumptions. The QR model for the τ^{th} quantile for i^{th} observation for p variables can be written as

$$Q_\tau(y_i) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \ldots + \beta_p(\tau)x_{ip}, i = 1, 2, \ldots, n \tag{2}$$

The coefficients $\beta_j(\tau)$ of the model can be estimated by solving the following minimization problem [43].

$$\min_{\beta_0(\tau)\ldots\beta_p(\tau)} \sum_{i=1}^{n} \rho_\tau \left(y_i - \beta_0(\tau) - \sum_{j=1}^{p} \beta_j(\tau)x_{ij} \right) \tag{3}$$

Where, $\rho_\tau(r) = \tau\, max(r, 0) + (1 - \tau) max(-r, 0)$ and r is the error of a single data point. This minimization involves the sum of absolute deviations which places equal weight on all observations regardless of their magnitude. In this study, we used a specific case of QR model where $\tau = 0.5$. This corresponds to the median quantile regression estimator, which is robust to outliers in the dependent variable.

Susceptible-infected-recovered (SIR) compartmental model

The SIR model is the first deterministic epidemic model, introduced by Kermack and McKendrick [44]. In this model, the population is divided into three compartments, such as Susceptible (S), Infected (I), and Recovered (I) and we assumed a fixed total population and no new births or deaths. Further, we assumed that all the infected individuals are contagious, and the recovered individuals are immune to the disease. Then, the dynamic of the model compartments can be evaluated using the ordinary differential equations (ODE) as follows.

$$\frac{dS}{dt} = -\beta I \frac{S}{N} \tag{4}$$

$$\frac{dI}{dt} = \beta I \frac{S}{N} - \gamma I \tag{5}$$

$$\frac{dR}{dt} = \gamma I \tag{6}$$

Where,: Total population, $N = S + I + R$

β: Infection Rate - The expected amount of people an infected person infects per day.

γ: The proportion of infected recovering per day.

Our research aimed to estimate the prevalence of COVID-19 in various cities located in North Dakota State. To achieve this goal, we utilized two distinct modeling techniques, namely KNN regression and QR. In both models, we utilized the natural logarithm-transformed RNA copies as the explanatory variable. We then generated predictions using these models and compared their results with those obtained from the basic SIR model. This comparison allowed us to assess the accuracy and reliability of each approach in estimating the number of SARS-CoV-2 infections in the study population.

By comparing the outcomes of these three models, our study provides valuable insights into the efficacy and accuracy of different modeling techniques in predicting the prevalence of COVID-19 disease in various population groups. This information can help inform public health policies and strategies to manage the ongoing pandemic effectively. Our research has the potential to contribute to efforts aimed at mitigating the spread of the virus and controlling its impact on the affected communities.

Results

We analyzed a dataset that included variables such as date of data collection, RNA copies, and clinical SARS-CoV-2 infections for various cities in North Dakota. All the analysis were performed using the R software version 4.1.3. The data was collected between January 1st,2021, to May 5th,2022. The prevalence of SARS-CoV-2 infections was predicted for this time frame using KNN regression and QR models with log-transformed RNA copies as explanatory variables. Further, we fitted an SIR model using clinical cases for the period of January 19th, 2022, to May 5th, 2022. We selected only a certain period to fit the SIR model as SIR model is less effective in predicting multiple waves of the virus. We compared these fitted values from SIR model with the KNN and QR models predictions of January 19th, 2022, to May 5th, 2022. Further, we compared the model performance by location, where Group 1 cities have a population $>$ 10000 and Group 2 cities have a population $\leq$ 10000 [45]. The fitted values from the models were compared with the observed counts of SARS-CoV-2 infections and the results are presented in Fig. 1a and Fig. 1b for two specific cities, where Group 1 City 8 has a population $>$ 10000 and Group 2 City 7 has a population $\leq$ 10000. The model fits for the other cities are provided in the appendix.

We used mean absolute percentage errors (MAPE) of observed and fitted values to evaluate the performance of the models.

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{A_i - F_i}{A_i} \right| \tag{7}$$

Where, n: Number of observations.

A_i : Actual value.

F_i : Fitted value.

Our results showed that the performance of the models varied between cities, with the SIR model outperforming both KNN and QR models in many cities. In addition, we found that the QR model performed better than the KNN model in Group 1 areas, while the KNN model was effective in predicting the prevalence of COVID-19 in some Group 2 areas. We presented a comparison of the MAPE values of the KNN and QR models in various Group 1 and Group 2 areas of North Dakota. The results are summarized in Table 1 and Fig. 2 for Group 1 areas, and Table 2 and Fig. 3 for Group 2 areas. These comparisons provide insight into the relative performance of the KNN and QR models in predicting the prevalence of COVID-19 cases across different regions of North Dakota. (See Table 3.)

The graph presented in Fig. 2 shows that the mean absolute percentage error of the KNN method is consistently higher than that of the QR method for Group 1 areas. Based on this finding, we suggest using the QR model to predict the number of COVID-19 cases in cities with larger populations. This model is likely to provide more accurate predictions compared to the KNN model in such areas, as indicated by its lower MAPE values. By selecting the appropriate modeling technique, we can improve our ability to estimate the prevalence of COVID-19 disease and develop effective strategies to manage its spread in areas with higher populations.

The graph provided in Fig. 3 indicates that the MAPE values obtained from the KNN method are lower for Group 2 areas, particularly when the population size is very small. The graph also illustrates that the QR method performs well for larger populations. Overall, these findings suggest that the choice of modeling technique should be informed by the population size of the study area. In Group 2 areas with smaller populations, the KNN method may provide more accurate predictions. In contrast, for larger populations, the QR method may be a better choice. By carefully selecting the appropriate modeling technique, we can improve the accuracy of our COVID-19 disease predictions and inform

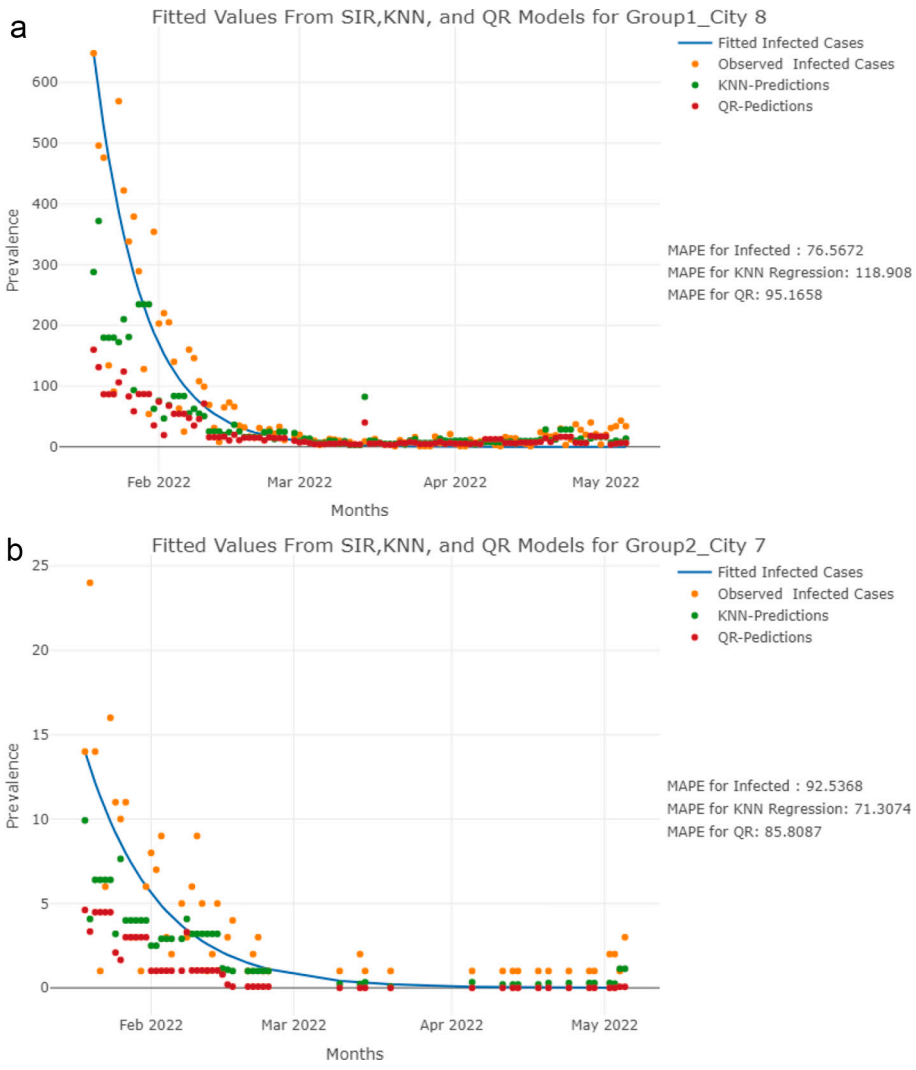


Fig. 1. a. Observed and fitted values comparisons for SIR, KNN, and QR Models for a City in Group 1.
b Observed and fitted values comparisons for SIR, KNN, and QR Models for a City in Group 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Note: The graphs presented in Fig. 1a for a city (population > 10000) and Fig. 1b for a city (population ≤ 10000) depict a comparison between the observed (orange points) and fitted values of *SARS-CoV-2 infections, employing three different models, namely SIR (blue line), KNN (green points), and QR (red points). The number of positive cases was estimated using* log*-RNA copies as the explanatory variable in the KNN and QR models, whereas the observed clinical cases were used to fit the SIR model.*

Table 1
MAPE values of QR and KNN models for Group 1 (population > 10000) cities.

City	MAPE (QR)	MAPE (KNN)
City 1	94.14	127.32
City 2	72.73	93.82
City 3	86.07	109.33
City 4	94.21	107.61
City 5	88.52	105.85
City 6	108.72	140.95
City 7	90.93	116.68
City 8	115.69	139.52

effective public health interventions.

In our study, we evaluated the relative performance of QR and KNN models in predicting COVID-19 prevalence in different North Dakota regions. The Wilcoxon-Sign Rank test was applied to compare the MAPE values of QR and KNN models in Group 1 and Group 2 areas. The Wilcoxon-Sign Rank Test is a non-parametric statistical test, used to determine whether two related samples are drawn from populations with the same distribution. As we used the same data in both methods, MAPE values can be considered as matching pairs.

The Wilcoxon signed rank test results indicate that there is no statistically significant difference ($P\ value = 0.8311$) in the results obtained from the QR and KNN methods when applied to data from Group 2 areas. However, a significant difference exists between the two methods ($P\ value = 0.007813$) when applied to data collected from Group 1 areas.

These findings suggest that the choice of modeling technique should be adapted to the characteristics of the study area. In Group 2 areas, either the QR or KNN method can be used to obtain similar results. However, in Group 1 cities, the QR method may be a more suitable choice for predicting the spread of COVID-19 disease compared to the KNN method. By considering these factors when choosing a modeling approach, we can improve the accuracy of our predictions and inform effective public health interventions.

Discussion

Our analysis was based on data collected from sewage samples and clinical cases observed in several North Dakota cities. The SIR model

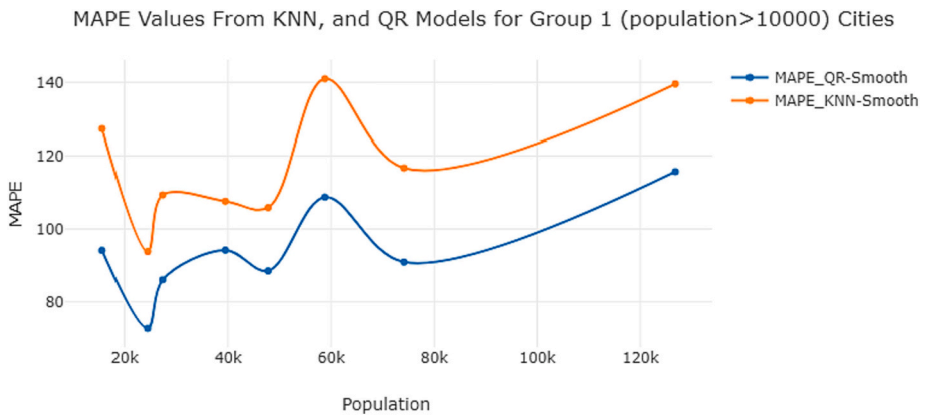


Fig. 2. MAPE values comparisons for KNN, and QR models for Group 1 (population > 10000) cities. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Note: The graph compares the MAPE values of KNN (orange line) and QR (blue line) models for Group 1 areas.

Table 2
MAPE values of QR and KNN models for Group 2 (population ≤ 10000) cities.

City	MAPE (QR)	MAPE (KNN)
City 1	72.34	58.50
City 2	70.27	72.79
City 3	69.84	65.14
City 4	81.80	65.49
City 5	79.81	57.22
City 6	60.79	59.81
City 7	64.21	84.61
City 8	72.33	77.98
City 9	76.44	86.44
City 10	84.39	87.39
City 11	88.87	88.73

based on clinical cases, alone predicted the prevalence of COVID-19 cases better than the KNN and QR methods. Using clinical cases alone for surveillance can be expensive and does not always account for asymptomatic cases. Therefore, using wastewater analysis is a more cost-effective and efficient way to determine the prevalence of COVID-19 cases in a community before the disease spreads.

It is worth noting that the KNN and QR methods also used clinical cases as a dependent variable in their models, which may have contributed to some limitations. The number of COVID-19 cases is based on the entire city, while our water treatment plants may not collect wastewater from the entire city. Capturing the required viral loads can also be difficult due to the collected small volume of wastewater samples. This would lead to an underestimation of the number of cases. Despite possible limitations, such as incomplete coverage of sewage in certain areas and lower viral load in smaller sample volumes, our study provides a good estimate of the prevalence of COVID-19 in whole cities using sewage samples.

Overall, our study highlights the potential of wastewater-based analysis as a valuable tool for disease surveillance in communities. This approach requires further development and optimization to address limitations but holds great promise for better surveillance of COVID-19 disease and other infectious diseases in the future.

CRediT authorship contribution statement

Bong-Jin Choi: Methodology, Conceptualization, Formal analysis, Visualization, Writing - original draft, Writing - review & editing. **Scott Hoselton:** Data curation, Writing - review & editing. **Grace N. Njau:** Writing – review & editing, Project administration. **I.G.C.G. Idamawatta:** Visualization, Writing - review & editing. **Paul Carson:** Writing –

Table 3
Wilcoxon-Sign Rank test results.

Area	Test Statistics	*P*-Value
Group 1	0	0.007813
Group 2	36	0.8311

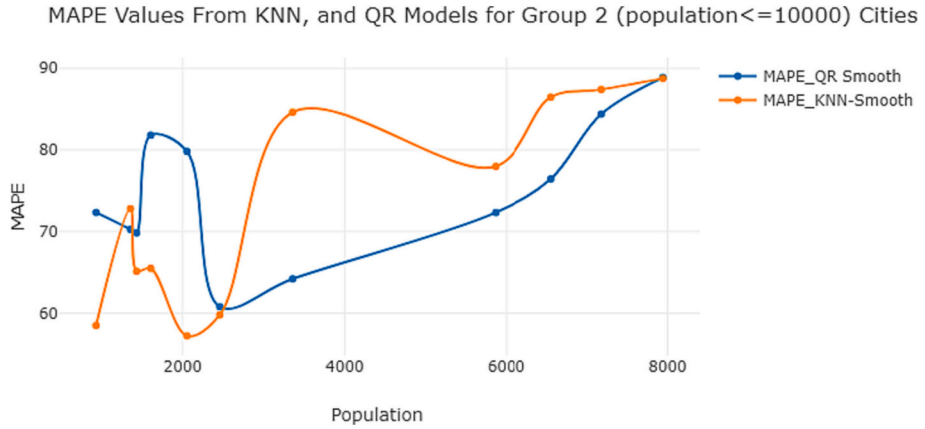


Fig. 3. MAPE values comparisons for KNN, and QR models for Group 2 (population ≤ 10000) cities. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Note: The graph compares the MAPE values of KNN (orange line) and QR (blue line) models for Group 2 areas.

review & editing. **John McEvoy:** Writing – review & editing.

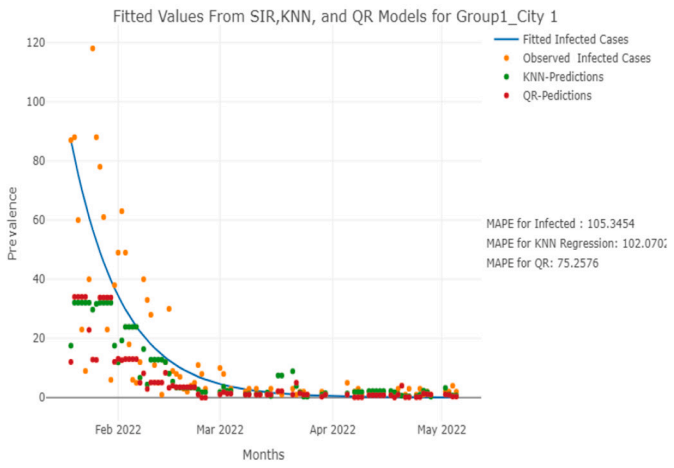
Declaration of Competing Interest

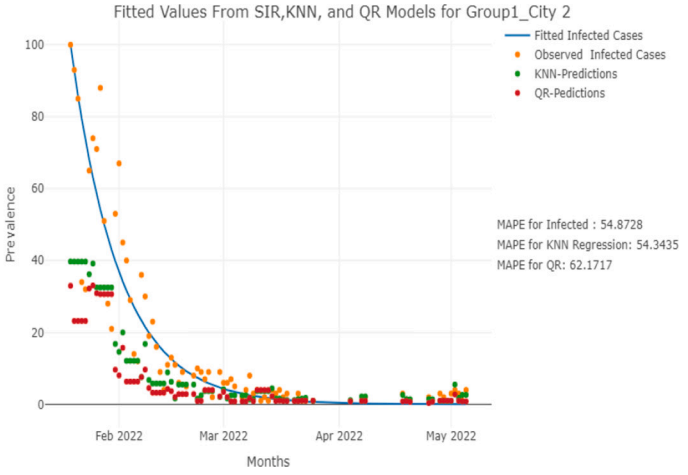
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

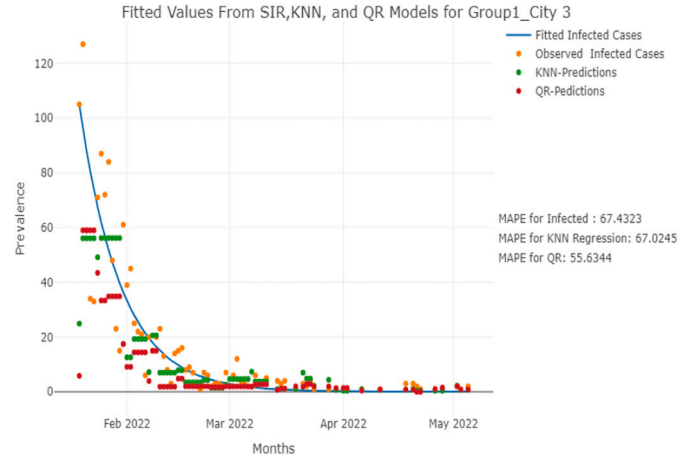
Acknowledgments

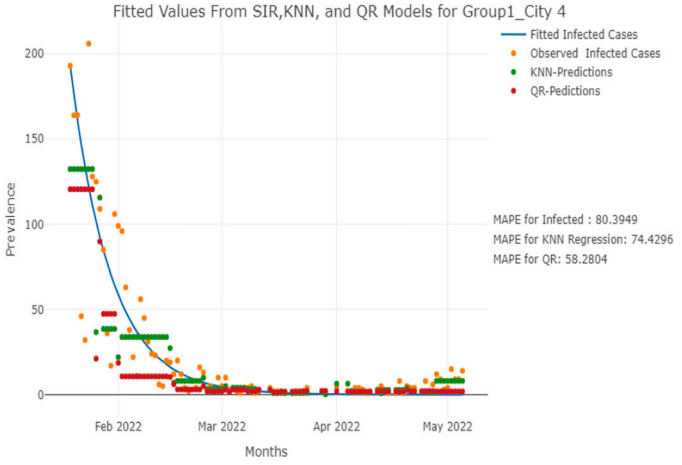
Support for this project was provided in part by the Centers for Disease Control and Prevention: Epidemiology and Laboratory Capacity for Infectious Diseases, National Initiative to Address COVID-19 Health Disparities Among Populations at High-Risk and Underserved, Including Racial and Ethnic Minority Populations and Rural Communities Cooperative Agreements, and Center for Diagnostic and Therapeutic Strategies in Pancreatic Cancer at the North Dakota State University

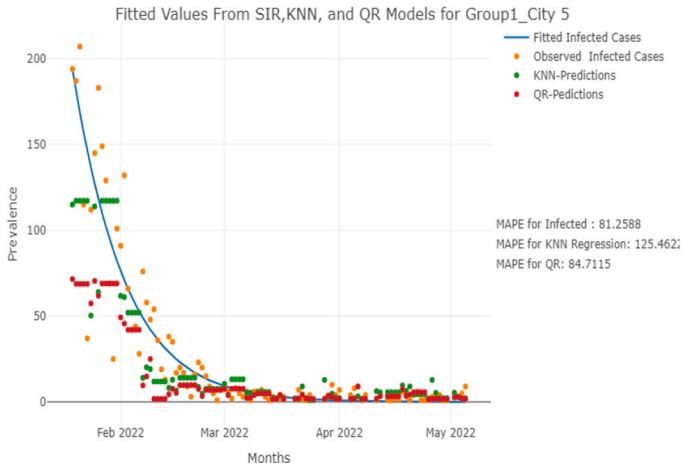
Appendix

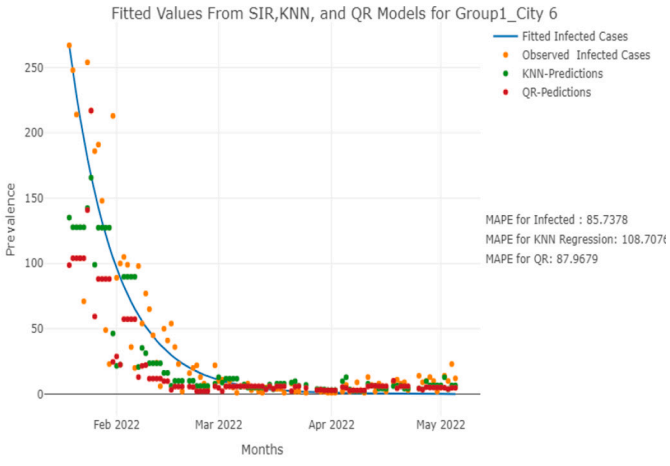


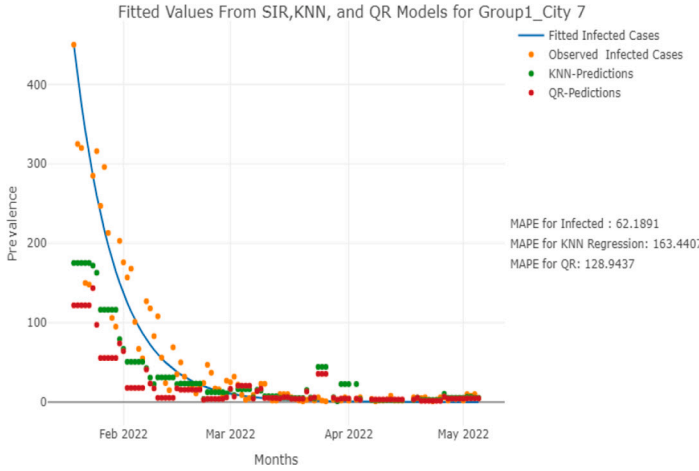


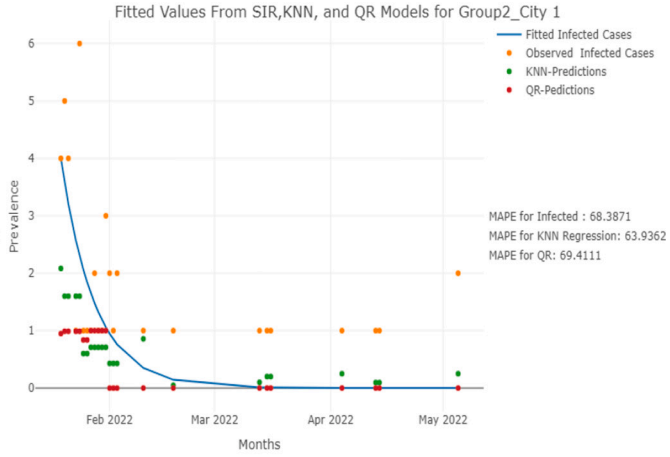


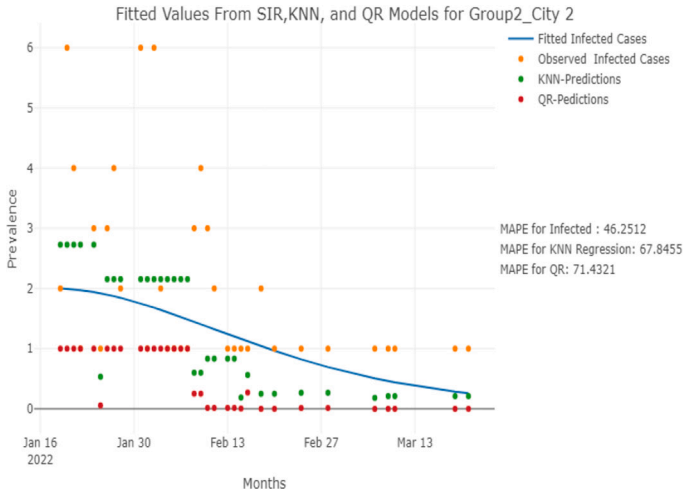


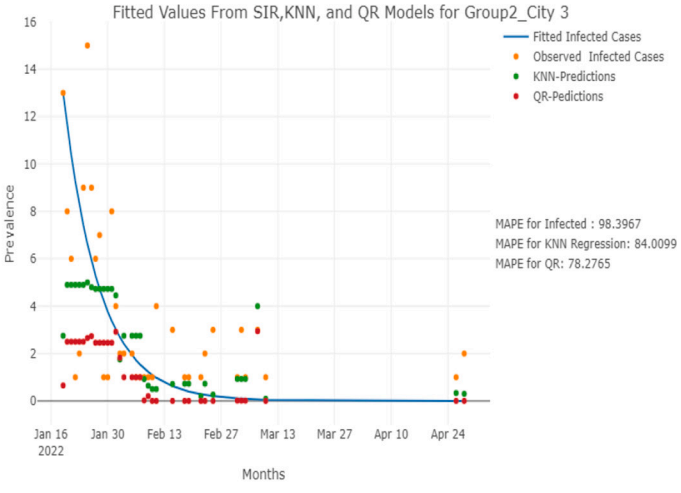

Fitted Values From SIR,KNN, and QR Models for Group1_City 5
Fitted Infected Cases
Observed Infected Cases
KNN-Predictions
QR-Pedictions
MAPE for Infected : 81.2588
MAPE for KNN Regression: 125.462
MAPE for QR: 84.7115
Prevalence
Months
Feb 2022
Mar 2022
Apr 2022
May 2022

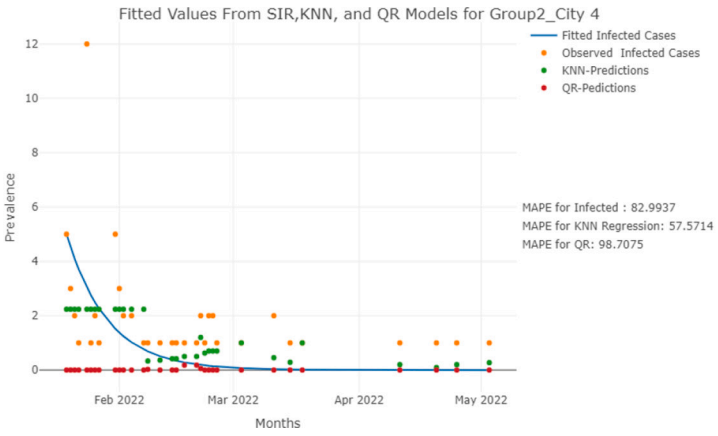

Fitted Values From SIR,KNN, and QR Models for Group1_City 6
Fitted Infected Cases
Observed Infected Cases
KNN-Predictions
QR-Pedictions
MAPE for Infected : 85.7378
MAPE for KNN Regression: 108.707
MAPE for QR: 87.9679
Prevalence
Months
Feb 2022
Mar 2022
Apr 2022
May 2022

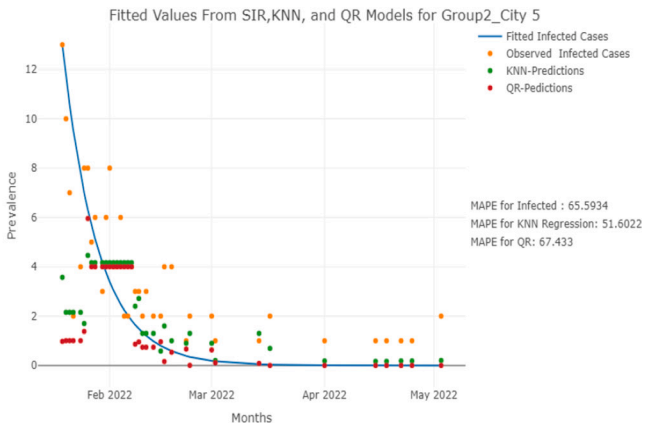

Fitted Values From SIR,KNN, and QR Models for Group1_City 7
Fitted Infected Cases
Observed Infected Cases
KNN-Predictions
QR-Pedictions
MAPE for Infected : 62.1891
MAPE for KNN Regression: 163.440
MAPE for QR: 128.9437
Prevalence
Months
Feb 2022
Mar 2022
Apr 2022
May 2022

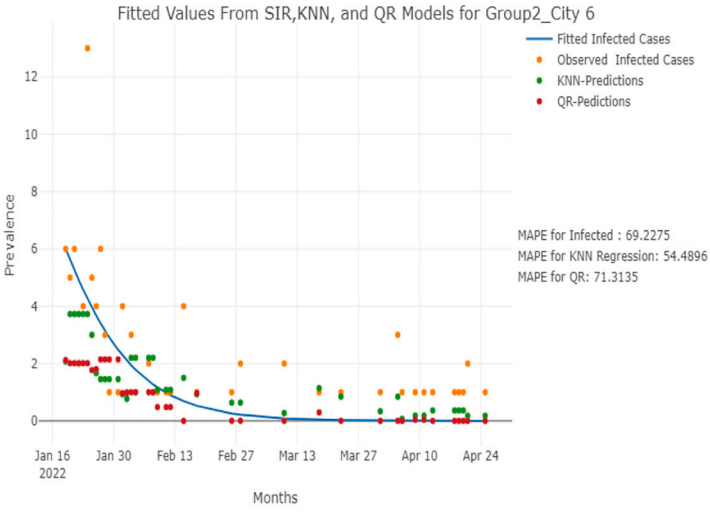

Fitted Values From SIR,KNN, and QR Models for Group2_City 1
Fitted Infected Cases
Observed Infected Cases
KNN-Predictions
QR-Pedictions
MAPE for Infected : 68.3871
MAPE for KNN Regression: 63.9362
MAPE for QR: 69.4111
Prevalence
Months
Feb 2022
Mar 2022
Apr 2022
May 2022

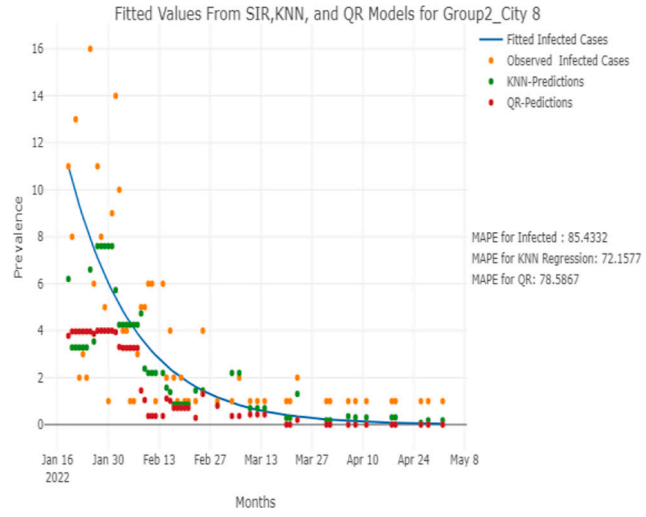

Fitted Values From SIR,KNN, and QR Models for Group2_City 2
Fitted Infected Cases
Observed Infected Cases
KNN-Predictions
QR-Pedictions
MAPE for Infected : 46.2512
MAPE for KNN Regression: 67.8455
MAPE for QR: 71.4321
Prevalence
Months
Jan 16
2022
Jan 30
Feb 13
Feb 27
Mar 13

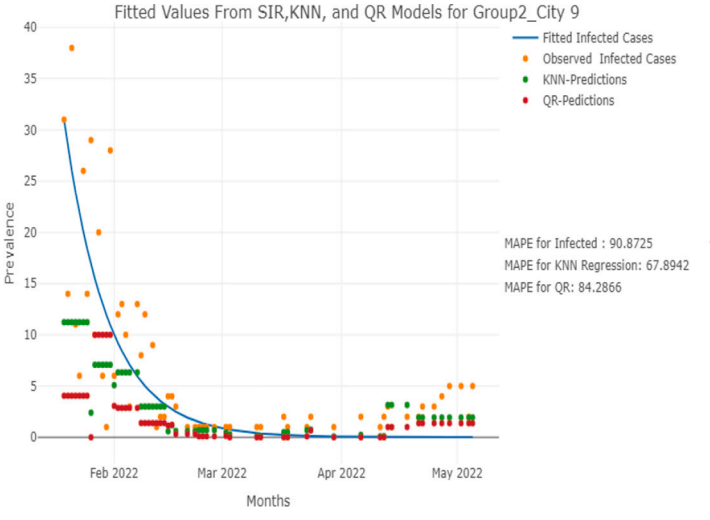

Fitted Values From SIR,KNN, and QR Models for Group2_City 3
Fitted Infected Cases
Observed Infected Cases
KNN-Predictions
QR-Pedictions
MAPE for Infected : 98.3967
MAPE for KNN Regression: 84.0099
MAPE for QR: 78.2765
Prevalence
Months
Jan 16
2022
Jan 30
Feb 13
Feb 27
Mar 13
Mar 27
Apr 10
Apr 24

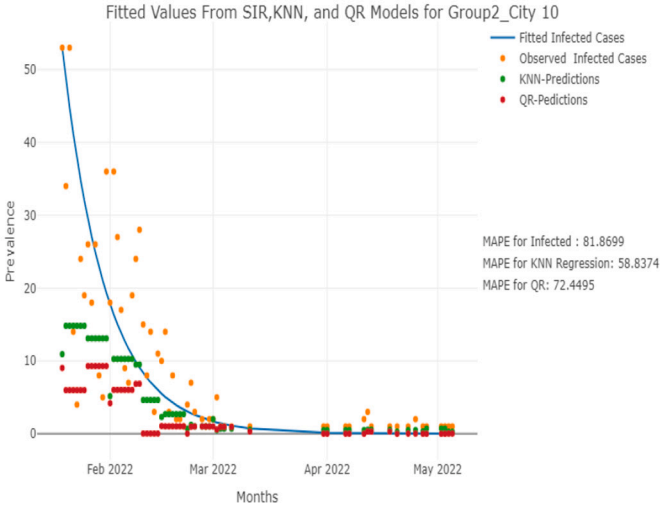

Fitted Values From SIR,KNN, and QR Models for Group2_City 4
Fitted Infected Cases
Observed Infected Cases
KNN-Predictions
QR-Pedictions
MAPE for Infected : 82.9937
MAPE for KNN Regression: 57.5714
MAPE for QR: 98.7075
Prevalence
Months
Feb 2022
Mar 2022
Apr 2022
May 2022

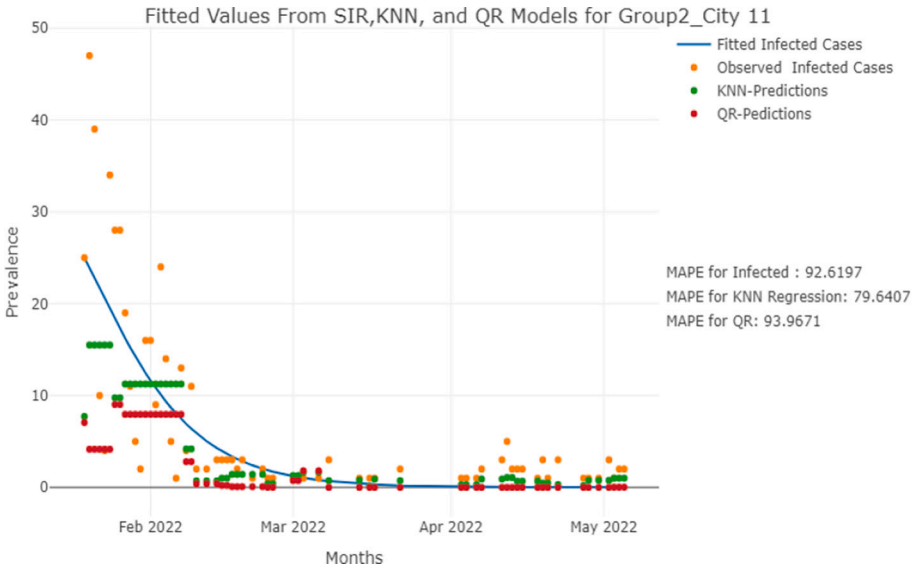

Fitted Values From SIR,KNN, and QR Models for Group2_City 5
Fitted Infected Cases
Observed Infected Cases
KNN-Predictions
QR-Pedictions
MAPE for Infected : 65.5934
MAPE for KNN Regression: 51.6022
MAPE for QR: 67.433
Prevalence
Months
Feb 2022
Mar 2022
Apr 2022
May 2022


Fitted Values From SIR,KNN, and QR Models for Group2_City 6
Fitted Infected Cases
Observed Infected Cases
KNN-Predictions
QR-Pedictions
MAPE for Infected : 69.2275
MAPE for KNN Regression: 54.4896
MAPE for QR: 71.3135
Prevalence
Months
Jan 16 2022
Jan 30
Feb 13
Feb 27
Mar 13
Mar 27
Apr 10
Apr 24


Fitted Values From SIR,KNN, and QR Models for Group2_City 8
Fitted Infected Cases
Observed Infected Cases
KNN-Predictions
QR-Pedictions
MAPE for Infected : 85.4332
MAPE for KNN Regression: 72.1577
MAPE for QR: 78.5867
Prevalence
Months
Jan 16 2022
Jan 30
Feb 13
Feb 27
Mar 13
Mar 27
Apr 10
Apr 24
May 8


Fitted Values From SIR,KNN, and QR Models for Group2_City 9
Fitted Infected Cases
Observed Infected Cases
KNN-Predictions
QR-Pedictions
MAPE for Infected : 90.8725
MAPE for KNN Regression: 67.8942
MAPE for QR: 84.2866
Prevalence
Months
Feb 2022
Mar 2022
Apr 2022
May 2022


Fitted Values From SIR,KNN, and QR Models for Group2_City 10
Fitted Infected Cases
Observed Infected Cases
KNN-Predictions
QR-Pedictions
MAPE for Infected : 81.8699
MAPE for KNN Regression: 58.8374
MAPE for QR: 72.4495
Prevalence
Months
Feb 2022
Mar 2022
Apr 2022
May 2022



References

[1] Jayaweera M, Perera H, Gunawardana B, Manatunge J. Transmission of COVID-19 virus by droplets and aerosols: a critical review on the unresolved dichotomy. Environ Res Sep. 2020;188:109819. https://doi.org/10.1016/j.envres.2020.109819.

[2] Zhao X, Liu S, Yin Y, Zhang T, Chen Q. Airborne transmission of COVID-19 virus in enclosed spaces: an overview of research methods. Indoor Air Jun. 2022;32(6):e13056. https://doi.org/10.1111/ina.13056.

[3] J. Tan *et al.*, "Transmission and clinical characteristics of asymptomatic patients with SARS-CoV-2 infection," Future Virol, p. 10.2217/fvl-2020–0087, doi: https://doi.org/10.2217/fvl-2020-0087.

[4] Bai Y, et al. Presumed asymptomatic carrier transmission of COVID-19. JAMA Apr. 2020;323(14):1406–7. https://doi.org/10.1001/jama.2020.2565.

[5] Rothe C, et al. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. N Engl J Med Mar. 2020;382(10):970–1. https://doi.org/10.1056/NEJMc2001468.

[6] Hu Z, et al. Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. Sci China Life Sci May 2020;63(5):706–11. https://doi.org/10.1007/s11427-020-1661-4.

[7] Nikolai LA, Meyer CG, Kremsner PG, Velavan TP. Asymptomatic SARS coronavirus 2 infection: invisible yet invincible. Int J Infect Dis Nov. 2020;100:112–6. https://doi.org/10.1016/j.ijid.2020.08.076.

[8] Johansson MA, et al. SARS-CoV-2 transmission from people without COVID-19 symptoms. JAMA Netw Open Jan. 2021;4(1):e2035057. https://doi.org/10.1001/jamanetworkopen.2020.35057.

[9] CDC, "National Wastewater Surveillance System," Centers for Disease Control and Prevention. Accessed: Sep. 01, 2022. [Online]. Available: https://www.cdc.gov/healthywater/surveillance/wastewater-surveillance/wastewater-surveillance.html.

[10] Berchenko Y, et al. Estimation of polio infection prevalence from environmental surveillance data. Sci Transl Med Mar. 2017;9(383):eaaf6786. https://doi.org/10.1126/scitranslmed.aaf6786.

[11] Anis E, et al. Insidious reintroduction of wild poliovirus into Israel, 2013. Eurosurveillance Sep. 2013;18(38):20586. https://doi.org/10.2807/1560-7917.ES2013.18.38.20586.

[12] Hellmér M, et al. Detection of pathogenic viruses in sewage provided early warnings of hepatitis a virus and norovirus outbreaks. Appl Environ Microbiol Nov. 2014;80(21):6771–81. https://doi.org/10.1128/AEM.01981-14.

[13] La Rosa G, et al. Surveillance of hepatitis a virus in urban sewages and comparison with cases notified in the course of an outbreak, Italy 2013. BMC Infect Dis Jul. 2014;14(1):419. https://doi.org/10.1186/1471-2334-14-419.

[14] Pan Y, Zhang D, Yang P, Poon LLM, Wang Q. Viral load of SARS-CoV-2 in clinical samples. Lancet Infect Dis Apr. 2020;20(4):411–2. https://doi.org/10.1016/S1473-3099(20)30113-4.

[15] Kim J-M, et al. Detection and isolation of SARS-CoV-2 in serum, urine, and stool specimens of COVID-19 patients from the Republic of Korea. Osong Public Health Res Perspect Jun. 2020;11(3):112–7. https://doi.org/10.24171/j.phrp.2020.11.3.02.

[16] Li X, Zhang S, Shi J, Luby SP, Jiang G. Uncertainties in estimating SARS-CoV-2 prevalence by wastewater-based epidemiology. Chem Eng J Jul. 2021;415:129039. https://doi.org/10.1016/j.cej.2021.129039.

[17] Peng L, et al. SARS-CoV-2 can be detected in urine, blood, anal swabs, and oropharyngeal swabs specimens. J Med Virol 2020;92(9):1676–80. https://doi.org/10.1002/jmv.25936.

[18] Crank K, Chen W, Bivins A, Lowry S, Bibby K. Contribution of SARS-CoV-2 RNA shedding routes to RNA loads in wastewater. Sci Total Environ Feb. 2022;806:150376. https://doi.org/10.1016/j.scitotenv.2021.150376.

[19] Li X, et al. SARS-CoV-2 shedding sources in wastewater and implications for wastewater-based epidemiology. J Hazard Mater Jun. 2022;432:128667. https://doi.org/10.1016/j.jhazmat.2022.128667.

[20] 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 May 2021;196:110946. https://doi.org/10.1016/j.envres.2021.110946.

[21] Kumar M, et al. Lead time of early warning by wastewater surveillance for COVID-19: geographical variations and impacting factors. Chem Eng J Aug. 2022;441:135936. https://doi.org/10.1016/j.cej.2022.135936.

[22] Medema G, Heijnen L, Elsinga G, Italiaander R, Brouwer A. Presence of SARS-Coronavirus-2 RNA in sewage and correlation with reported COVID-19 prevalence in the early stage of the epidemic in the Netherlands. Environ Sci Technol Lett Jul. 2020;7(7):511–6. https://doi.org/10.1021/acs.estlett.0c00357.

[23] 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 Aug. 2020;181:115942. https://doi.org/10.1016/j.watres.2020.115942.

[24] Manuel D, Amadei CA, Campbell JR, Brault J-M, Veillard J. Strengthening public health surveillance through wastewater testing: An essential investment for the COVID-19 pandemic and future health threats. Washington, DC: World Bank; Jan. 2022. Accessed: Aug. 31, 2022. [Online]. Available: https://openknowledge.worldbank.org/handle/10986/36852.

[25] Natarajan A, et al. Gastrointestinal symptoms and fecal shedding of SARS-CoV-2 RNA suggest prolonged gastrointestinal infection. Med Jun. 2022;3(6). https://doi.org/10.1016/j.medj.2022.04.001. pp. 371–387.e9.

[26] Zhang W, et al. Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. Emerg Microbes Infect 2020;9(1):386–9. https://doi.org/10.1080/22221751.2020.1729071.

[27] Xu Y, et al. Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding. Nat Med Apr. 2020;26(4). https://doi.org/10.1038/s41591-020-0817-4. Art no 4.

[28] van Doorn AS, Meijer B, Frampton CMA, Barclay ML, de Boer NKH. Systematic review with meta-analysis: SARS-CoV-2 stool testing and the potential for faecal-oral transmission. Aliment Pharmacol Ther 2020;52(8):1276–88. https://doi.org/10.1111/apt.16036.

[29] Cheung KS, et al. Gastrointestinal manifestations of SARS-CoV-2 infection and virus load in fecal samples from a Hong Kong cohort: systematic review and meta-analysis. Gastroenterology Jul. 2020;159(1):81–95. https://doi.org/10.1053/j.gastro.2020.03.065.

[30] Jones DL, et al. Shedding of SARS-CoV-2 in feces and urine and its potential role in person-to-person transmission and the environment-based spread of COVID-19. Sci Total Environ Dec. 2020;749:141364. https://doi.org/10.1016/j.scitotenv.2020.141364.

[31] Cerrada-Romero C, et al. Excretion and viability of SARS-CoV-2 in feces and its association with the clinical outcome of COVID-19. Sci Rep May 2022;12(1). https://doi.org/10.1038/s41598-022-11439-7. Art. no. 1.

[32] Vallejo JA, et al. Modeling the number of people infected with SARS-COV-2 from wastewater viral load in Northwest Spain. Sci Total Environ Mar. 2022;811:152334. https://doi.org/10.1016/j.scitotenv.2021.152334.
[33] Robotto A, et al. Wastewater-based SARS-CoV-2 environmental monitoring for Piedmont, Italy. Environ Res Jan. 2022;203:111901. https://doi.org/10.1016/j.envres.2021.111901.
[34] Matheri AN, Belaid M, Njenga CK, Ngila JC. Water and wastewater digital surveillance for monitoring and early detection of the COVID-19 hotspot: industry 4.0. Int J Environ Sci Technol Feb 2022. https://doi.org/10.1007/s13762-022-03982-7.
[35] Jiang G, et al. Artificial neural network-based estimation of COVID-19 case numbers and effective reproduction rate using wastewater-based epidemiology. Water Res Jun. 2022;218:118451. https://doi.org/10.1016/j.watres.2022.118451.
[36] La Rosa G, et al. First detection of SARS-CoV-2 in untreated wastewaters in Italy. Sci Total Environ Sep. 2020;736:139652. https://doi.org/10.1016/j.scitotenv.2020.139652.
[37] Ahmed W, et al. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: a proof of concept for the wastewater surveillance of COVID-19 in the community. Sci Total Environ Aug. 2020;728:138764. https://doi.org/10.1016/j.scitotenv.2020.138764.
[38] McMahan CS, et al. COVID-19 wastewater epidemiology: a model to estimate infected populations. Lancet Planet Health Dec. 2021;5(12):e874–81. https://doi.org/10.1016/S2542-5196(21)00230-8.
[39] Fazli M, Sklar S, Porter MD, French BA, Shakeri H. Wastewater-Based Epidemiological Modeling for Continuous Surveillance of COVID-19 Outbreak. medRxiv, p. 2021.10.19.21265221. Oct. 20, 2021. https://doi.org/10.1101/2021.10.19.21265221.
[40] Silverman BW, Jones MC. E. Fix and J.L. Hodges (1951): an important contribution to nonparametric discriminant analysis and density estimation: commentary on Fix and Hodges (1951). Int Stat Rev 1989;57(3):233–8. https://doi.org/10.2307/1403796.
[41] Koenker R, Bassett G. Regression quantiles. Econometrica 1978;46(1):33–50. https://doi.org/10.2307/1913643.
[42] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc Royal Soc London Series A, Contain Papers Math Phys Char 1927;115(772):700–21.
[43] Rodriguez RN. Five Things You Should Know about Quantile Regression. 2017. Accessed: Sep. 12, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Five-Things-You-Should-Know-about-Quantile-Rodriguez/7e824688e34f6010ee203c86643b17ccbd9acf90.
[44] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc Royal Soc London Series A, Contain Papers Math Phys Character 1927;115(772):700–21.
[45] U. C. Bureau, "City and Town Population Totals: 2020-2021," Censusgov Accessed: Apr 28, 2023. [Online]. Available: https://www.census.gov/data/tables/time-series/demo/popest/2020s-total-cities-and-towns.html.