Lifting of travel restrictions brings additional noise in COVID-19 surveillance through wastewater-based epidemiology in post-pandemic period

Li, Xuan; Li, Jibin; Liu, Huan; Mínguez-Alarcón, Lidia; van Loosdrecht, Mark C.M.; Wang, Qilin

DOI
10.1016/j.watres.2025.123114

Publication date
2025

Document Version
Final published version

Published in
Water Research

Citation (APA)
Li, X., Li, J., Liu, H., Mínguez-Alarcón, L., van Loosdrecht, M. C. M., & Wang, Q. (2025). Lifting of travel restrictions brings additional noise in COVID-19 surveillance through wastewater-based epidemiology in post-pandemic period. *Water Research*, *274*, Article 123114. https://doi.org/10.1016/j.watres.2025.123114

Water Research 274 (2025) 123114

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Lifting of travel restrictions brings additional noise in COVID-19 surveillance through wastewater-based epidemiology in post-pandemic period

Xuan Li [a,1], Jibin Li [a,1], Huan Liu [a], Lidia Mínguez-Alarcón [b,c], Mark C.M. van Loosdrecht [d], Qilin Wang [a,*]

[a] Centre for Technology in Water and Wastewater, School of Civil and Environmental Engineering, University of Technology Sydney, Ultimo, NSW 2007, Australia
[b] Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[c] Channing Division of Network Medicine, Harvard Medical School & Brigham and Women's Hospital, USA
[d] Department of Biotechnology, Delft University of Technology, Julianalaan 67, Delft 2628, BC, the Netherlands

ARTICLE INFO

Keywords:
Wastewater-based epidemiology
COVID-19
Machine learning
Post-pandemic
Travel restriction

ABSTRACT

The post-pandemic world still faces ongoing COVID-19 infections, although international travel has returned to pre-pandemic conditions. Wastewater-based epidemiology (WBE) is considered an efficient tool for the population-wide surveillance of COVID-19 infections during the pandemic. However, the performance of WBE in post-pandemic era with travel restrictions lifted remains unknown. Utilizing weekly county-level wastewater surveillance data from June 2021-November 2022 for 222 counties in 49 states (covering 104 million people) in the United States of America, we retrospectively evaluated the correlations between SARS-CoV-2 RNA (C_{RNA}) and reported cases, as well as the impacts of international air travel, demographics, socioeconomic aspects, test accessibility, epidemiological, and environmental factors on reported cases under the corresponding C_{RNA}. The lifting of travel restrictions in June 2022, shifted the correlation between C_{RNA} and COVID-19 incidence in the following 7-day and 14-day from 0.70 (IQR: 0.30–0.88) in June 2021-May 2022 (pandemic) to 0.01 (IQR: -0.31–0.36) in June-November 2022 (post-pandemic), and from 0.74 (IQR: 0.31–0.90) to -0.01 (IQR: -0.38–0.45), respectively. Besides, after lifting the travel restrictions, under the same C_{RNA}, the reported case numbers were impacted by many factors, including the variations of international passengers, test accessibility, Omicron prevalence, ratio of population aged between 18 and 65, minority vulnerability, and healthcare system. This highlights the importance of demographics, infection testing, variants and socioeconomic status on the accuracy and implication of WBE to monitor COVID-19 infection status in post-pandemic era. Our findings facilitate the public health authorities to dynamically adjust their WBE-based tools/strategies to the local contexts to achieve optimal community surveillance.

1. Introduction

COVID-19 remains a global health threat, despite being no longer declared a global health emergency since May 2023 by World Health Organization (WHO) (Harris, 2023). Although most countries have returned to pre-outbreak lifestyles, COVID-19 has caused 1431,900 reported cases and 31,230 deaths globally from January to April in 2024 (WHO, 2024). Accurate case monitoring is crucial for governments and public health authorities to track infection status and dynamically adjust the COVID-19 policies on restrictions and social activities (Arvisais-Anhalt et al., 2021). However, in post-pandemic world, COVID-19 testing requirements have become optional for individuals, and public willingness to test gradually declines and shifts from clinical testing to self-testing (Silk et al., 2023; Walensky, 2022). Therefore, more countries intend to rely on wastewater-based epidemiology (WBE) as an alternative in monitoring and predicting community infections for public health sectors (Silk et al., 2023).

The overall concept of WBE relies on the shedding of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RNA from COVID-19 patients (e.g., feces, sputum, and urine) to sewers and its wastewater

---

[*] Corresponding author.
E-mail address: Qilin.Wang@uts.edu.au (Q. Wang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.watres.2025.123114
Received 12 May 2024; Received in revised form 20 December 2024; Accepted 7 January 2025

concentration (C_{RNA}) (Li et al., 2021; Pan et al., 2020). In theory, C_{RNA} can be used to reflect or estimate the COVID-19 incidences (cases) within the community, assuming comparable shedding loads between individuals (Bibby et al., 2021; Polo et al., 2020). However, viral shedding load varied significantly between patients, primarily influenced by individual hygiene practices (such as disposal habits for bodily fluids like sputum and saliva), age demographics, racial factors, and past/current health conditions, etc. (Li et al., 2022; Noor and Islam, 2020). Thus, WBE-based surveillance or early-warning for COVID-19 incidence largely depends on the observed correlation between C_{RNA} and reported cases/incidences in a certain wastewater catchment (Medema et al., 2020; Polo et al., 2020). This correlation serves as a correction/conversion factor for estimating the incidences from a new wastewater sample (Polo et al., 2020). Thus, a strong correlation between C_{RNA} and reported cases/incidences is crucial for the application and accuracy of WBE in estimating/monitoring COVID-19 incidences in the community.

WBE had provided valuable early-warning and surveillance on COVID-19 community infections, during the pandemic when international travel was restricted. Promising correlations between C_{RNA} and reported cases/incidences (C_{RNA}-Incidence correlation) have been widely reported during pandemic times, ranging from 0.82 to 0.9 (Hillary et al., 2021; Jiang et al., 2022; Prado et al., 2021; Sanjuán and Domingo-Calap, 2021). In post-pandemic times, the lifestyle and travel conditions have changed notably compared with the pandemic period. For instance, the monthly number of total air passengers in the United States of America (USA) increased from 12 M before June 2022, to 18 M in May and June 2022 due to the easing of travel restrictions. Several studies have observed reduced C_{RNA}-Incidence correlations in certain wastewater catchments in post-pandemic times (de Freitas Bueno et al., 2022; McManus et al., 2023; Varkila et al., 2023). However, such observations are limited to a few catchments in a short time period (a few weeks), with limited insight into the reasons behind. To date, the temporal variations (before and after lifting travel restrictions) in C_{RNA}-Incidence correlations across a vast geographical area encompassing numerous wastewater catchments with diverse catchment characteristics (e.g., demographics, socioeconomic, environmental, etc.) are still unclear. Also, the impact of these catchment characteristics on the C_{RNA}-incidence correlation remains unexplored.

This study aims to investigate the variations in the C_{RNA}-incidence correlation before and after the lifting of travel restrictions (a critical indicator of the end of the pandemic emergency status). We retrospectively collected county-level weekly C_{RNA} data, along with corresponding COVID-19 incidence in 222 counties (Table S1) across 49 states in USA (Fig. 1a) from June 2021 to November 2022. These counties in total cover 104 M population, and the data collection periods include both the pandemic period (June 2021-May 2022) and post-pandemic period (June-November 2022, after lifting the travel restriction). We also collected data on 37 explanatory factors (Table S2), including demographics (5 factors), socioeconomic (9 factors), epidemiological (15 factors), environmental conditions (3 factors), test accessibility (3 factors), and international air travel data (2 factors) for the examined counties. The impacts of these factors on the reported COVID-19 incidence under the corresponding C_{RNA} were evaluated through random forest models. The results of this study will provide a comprehensive insight into the variations and impacting factors of the C_{RNA}-incidence correlations in the post-pandemic era, facilitating public health authorities and policymakers to dynamically interpret and improve the WBE performance based on the local conditions.

2. Materials and methods

2.1. Wastewater surveillance and COVID-19 case data of USA counties

Considering the progress of the vaccination and experience in the sample analysis and the analytical method consistency of data from different counties/states, the vaccination progress and experience in response to COVID-19, the weekly SARS-CoV-2 concentrations in wastewater (C_{RNA}) at county-level from June 2021 to November 2022 were collected from Biobot database (biobot.io/data) and used in this study. The Biobot Nationwide Wastewater Monitoring Network provides the largest publicly accessible dataset on COVID-19 community surveillance with over 200 participating locations, which covers 30 % of the USA population (Analytics, 2021). The sampling details, analytical protocol, and data processing were specifically described in Duvallet et al. (2022). Overall, county-level weekly C_{RNA} for 222 counties in 49 states in the USA was included in the surveillance data (a total of 7540 data points). The daily clinical confirmed cases of COVID-19 identified by testing were obtained from USA Fact (https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/). The population size of included counties was collected from USA census database (USA Census Bureau).

2.2. Correlation between reported cases and C_{RNA}

Wastewater surveillance was found to be able to estimate the weekly, 15-day-rolling, and future COVID-19 cases of wastewater samples in previous studies (Jiang et al., 2022; Kennedy et al., 2021; Li et al., 2023b). Thus, weekly new, 15-day-rolling and future COVID-19 cases (confirmed in the following 2–14 days) of wastewater samples were further calculated based on the daily clinical confirmed cases in each county (Table S3). Considering the potential temporal changes in the community (i.e., vaccination rollout, variants, etc.), the data points in each county were divided into several subsets over the time period of 5 weeks. This is in line with the application of WBE, where the correlation observed in historical data is used to prove instructions for estimating the infection status from the current sample (Hillary et al., 2021; Sanjuán and Domingo-Calap, 2021; Trottier et al., 2020). The correlations for each subset of data were determined through Spearman's correlation test using R (Version 4.2.2, http://www.R-project.org/).

2.3. Demographic, CCVI, test accessibility, epidemiological, environmental, and international air travel data of the USA counties

A total of 37 factors (5 in demographic, 9 in CCVI and socioeconomic, 15 in epidemiological, 3 in test accessibility, 3 in environmental, and 2 in international air travel) conditions were included in this study (Table S2). The demographic factors included the ratio of population aged below 5, ratio of population aged between 5 and 17, ratio of population aged between 18 and 65 and ratio of population aged over 65.

For COVID-19 management and response, COVID-19 Community Vulnerability Index (CCVI) was established by Surgo Foundation and used by Centers for Disease Control in USA (Tipirneni et al., 2022). CCVI reflects the socioeconomic status and provides a comprehensive assessment of demographic vulnerability at county level via 40 measures from census data, covering 7 factors including: i) socioeconomic status; ii) minority status and language, iii) housing type, transportation, household composition, and disability ('household & transportation' hereafter), iv) epidemiological risk, v) healthcare system, vi) high risk environment, and vii) population density (Smittenaar et al., 2021). The CCVI overall vulnerability and other 7 theme indices were scored ranging from 0 to 1, with a higher score representing higher vulnerability of certain aspect in this county.

Test accessibility factors include test sites per 100k, distance to test sites and positive ratio of testing. Epidemiological factors included the ratios of vaccination recipients in first dose, second dose and booster, the ratio of population infected with different SARS-CoV-2 variants, Shannon and Inverse Simpson diversity index of variant in different periods. The ratio of population infected with different SARS-CoV-2 variants was calculated based on state-level report from the health department (Section S2 in the supplementary information), as the quantification of

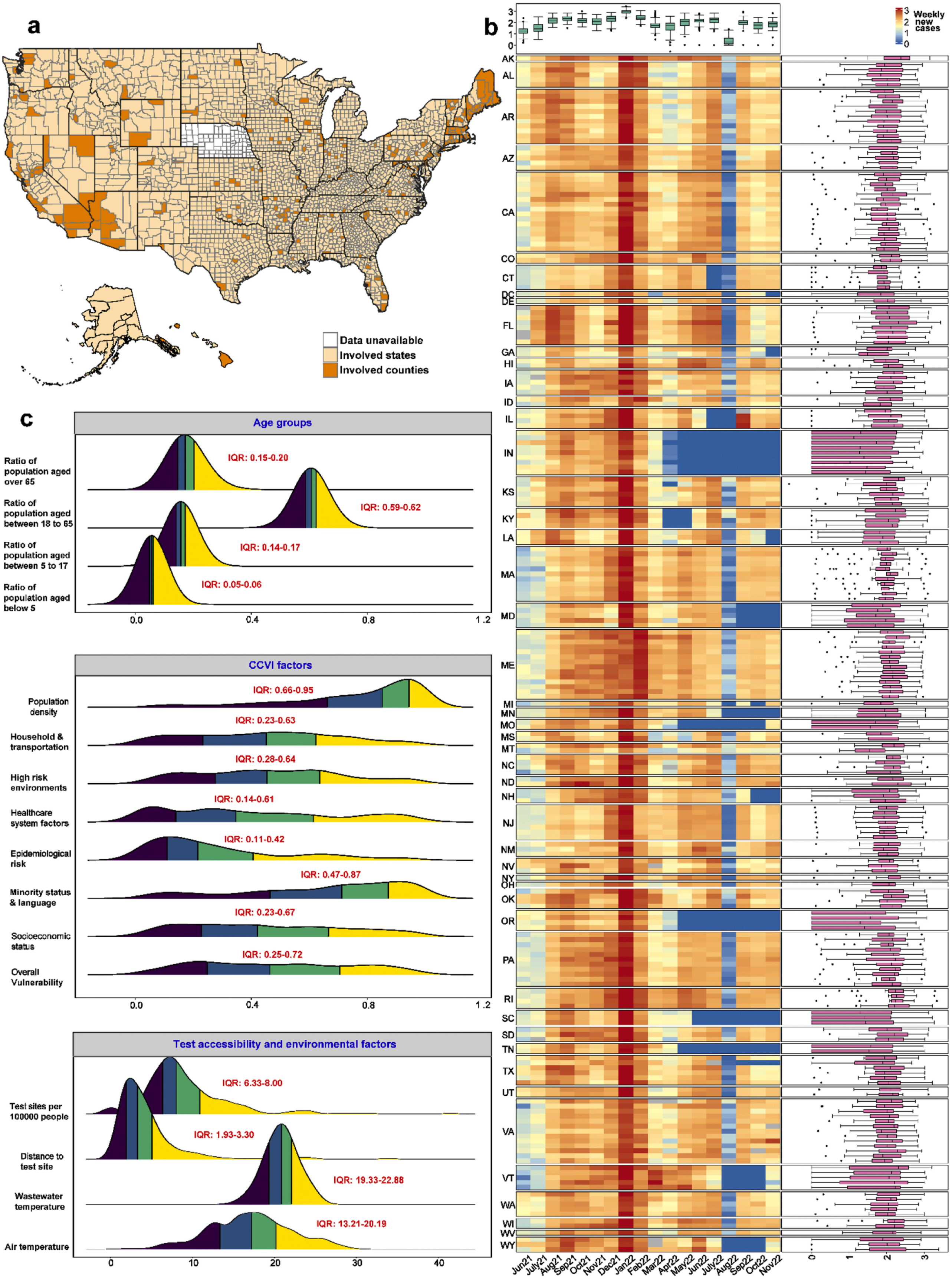


Fig. 1. The comprehensive features of examined 222 USA counties. (a) geographical locations (b) spatial and temporal characteristics of weekly new COVID-19 cases in each county (cases/100k population, shown in the logarithm form in the figure). For better visualization, counties were grouped by state (shown in abbreviations), with the full list provided in Table S1. (c) age groups, CCVI and test accessibility, and environmental factors (a factor without unit means it is dimensionless).

different variant sequence reads in wastewater remains challenging (Karthikeyan et al., 2022). The variant diversity was calculated at state level based on the report from the health department (Section S2 in the supplementary information). Inverse Simpson diversity index (equals to 1/Simpson diversity index) was calculated using R.

Environmental factors include precipitation, air temperature, and wastewater temperature. In order to evaluate the impact of test restrictions cancellation for international air passenger in June 2022, the inbound passenger (i.e., passenger entering USA) and total international passenger (sum of inbound passenger and outbound passenger) were included.

The specific details of data source and data calculation for each factor were shown in Section S1 and Section S2 in the supplementary information. The statistical analysis (maximum, minimum, mean, median, and interquartile range (IQR)) of each factor was implemented using R. The data visualization was also conducted using R.

2.4. Random forest modeling

Random forest model was used to identify the importance of studied factors (a total of 37 input factors in Table S2) and their contributions to clinical confirmed cases (i.e., following 7-day confirmed case and following 14-day confirmed case) through R (version 4.2.0). Random forest is a non-parametric machine learning approach that allows modelling the relationship between target and potential explanatory factors (Breiman, 2001; Ho, 1998). Random forest model does not require a prior hypothesis in terms of the statistical distribution of the data, which makes it feasible to explore more refined connections, particularly the non-linear relationship between target and explanatory factors (Ali et al., 2012; Li et al., 2023a). The included data were randomly divided into three parts, which were the training set (accounting for 70 %), the validation set (accounting for 15 %) and test sets (accounting for 15 %). The training set was used to train the random forest model, then the validation set was used as a conjunction to optimize the model structure, and the test sets were used to assess the prediction capability of model on unseen data. The basic construction of random forest models and the model evaluation strategy were depicted in SI section S3.

2.5. Significance and contribution of explanatory factors

The significance contribution of explanatory factors (input) on following 7-day case (output) and following 14-day case (output) was estimated via rfPermute package in R (version 4.2.2), which was achieved by calculating percentage increase in mean squared error (%IncMSE) with relevant P values via 5-fold cross-validation with 5 replicates (Archer and Archer, 2016). Partial dependence analysis, commonly used alongside machine learning models such as random forests, evaluates the significance of explanatory factors and illustrates the variation trends in the output as input variables (i.e., explanatory factors) change. This approach effectively demonstrates the relationship between a subset of input variables and the predicted output of a model, providing critical insights into how specific factors influence the WBE data (i.e., C_{RNA}) (Cheng et al., 2019). Partial dependence analysis is achieved by marginalizing the model output over the entire possible distribution of input variables, including the value 0 (Cheng et al., 2019; Molnar, 2020). Partial dependence analysis estimates the marginal effect of a single feature by averaging the model's predictions over the distribution of all other features (Molnar, 2020). Thus, partial dependence result reflects relative trends in the model output rather than actual observed values. The partial dependence function for regression, and the differences between permutation and partial dependence were described in detail in SI Section S4. The permutation analysis by rfPermutate package was firstly to find out all significant factors ($P < 0.05$), then the partial dependence analysis was used to illustrate how the significant factor affects the target variable. To quantify the impact on the C_{RNA}-Incidence correlation, partial dependence analyses were conducted for each identified significant factor ($P < 0.05$). The following 7-day and following 14-day cases were used as target variables (output) with C_{RNA} and significant factor as two predictor variables in partial dependence analysis. Due to the variable evaluation differences between permutation and partial dependence analysis, some factors would be identified as significant factor with high %IncMSE in permutation while showed no obvious impact on the variations of target variable. These factors would be excluded in partial dependence discussions. Only the factors simultaneously showing significance ($P < 0.05$) in permutation analysis and showing notable impact on target variable during partial dependence, would be included in the discussions of partial dependence analysis.

3. Results

3.1. COVID-19 status, demographic, socioeconomic, epidemiological, environmental, test accessibility conditions and international air travel of examined counties

The 222 counties (Table S1) across 49 states in USA with population size ranged from 0.01 to 10 M, covering 104 M populations (Fig. 1a). During the study period (June 2021-November 2022), the weekly new cases ranged from 0 to 271,299 cases (IQR: 49–222) with clear regional variations (Fig. 1b). The weekly new cases of examined counties began to increase since June 2021 and peaked in January 2022 with 2 infection waves observed (December 2021-Feburary 2022 and May 2022-July 2022) (Fig. 1b). During the study period, C_{RNA} ranged from 0.3 to 8833.1 copies/mL (IQR: 221.7–910.8) and also peaked in January 2022 (Fig. S1).

The demographic status of examined counties was represented by population ratios of different age groups (Fig. 1c). Population aged between 18 and 65 (years old, omitted hereafter) was the dominant group among examined counties with a population ratio of 0.44–0.74 (IQR: 0.59–0.62). The range of ratio for population aged below 5 was 0.03–0.08 (IQR: 0.05–0.06), between 5 and 17 was 0.11–0.24 (IQR: 0.14–0.18), and over 65 was 0.10–0.40 (IQR: 0.15–0.20).

For better adaptation and transferability to other regions, CCVI was used to reflect the socioeconomic and partial epidemiological status of the examined counties. These counties had an overall vulnerability of 0.47 (IQR: 0.25–0.72), CCVI in socioeconomic status of 0.42 (IQR: 0.23–0.67), minority status and language of 0.71 (IQR: 0.47–0.87), household & transportation of 0.45 (IQR: 0.23–0.63), epidemiological risk of 0.22 (IQR: 0.11–0.42), healthcare system of 0.36 (IQR: 0.14–0.61), high risk environments of 0.46 (IQR: 0.28–0.64) and population density of 0.86 (IQR: 0.66–0.95) (Fig. 1c).

The epidemiological factors include vaccination factors, COVID-19 variants, and variant diversity (Table S2). The ratio of vaccination recipients among the total population increased from 0.58 (IQR: 0.51–0.65) in June 2021 to 0.77 (IQR: 0.62–0.88) in November 2022 for the first dose, from 0.51 (IQR: 0.43–0.56) to 0.67 (IQR: 0.54–0.75) for the second dose, and from 0.34 (IQR: 0.32–0.36) to 0.54 (IQR: 0.45–0.57) for booster (Fig. S2). Clear regional and temporal differences in vaccination recipient ratios were observed among counties (Figs. S3-S5). The major variants of SARS-CoV-2 shifted from Alpha to Delta after June 2021, then to Omicron after January 2022, with 18 subvariants of Delta and Omicron (lineages) being examined (Figs. S6), and clear regional and temporal variations were observed (Figs. S7-S9). The variant diversity was reflected through Shannon index of 0.74 (IQR: 0.55–0.93) and Inverse Simpson index of 1.79 (IQR: 1.35–2.10) in these counties, which changed along with geographic regions and time (Figs. S10-S11). These 2 diversity indices all peaked in June 2021, which was consistent with the period in which most variants coexisted (Fig. S6a).

The test accessibility also showed temporal and regional variations (Fig. 1c, Figs. S12-S13). Positive ratio of testing remained low with medians < 0.10 (IQR: 0.06–0.15) for most of the time, but peaked in January 2022 (0.30, IQR: 0.22–0.34) and June-July 2022 (0.17, IQR: 0.15–0.22). For most counties, the distance to test sites was 3.2 miles (IQR: 1.9–3.3), while the test sites per 100,000 people (100k) was 8 (IQR:6–9) (Fig. 1c and Fig. S13). Examined counties showed obvious seasonal variations for environmental factors (i.e., low temperature in winter and high temperature in summer) in air temperature and wastewater temperature, which medians were 18.1 °C (IQR: 13.2–20.2 °C) and 21.2 °C (IQR: 19.3–22.9 °C), respectively (Fig. 1c, Figs. S14-S15).

As for international air travel, inbound passenger and total international passenger of examined states ranged from 0 to 2.87 M (IQR: 0.08–0.45 M) and 0–5.70 M (IQR: 0.15–0.89 M), respectively. Similar temporal and regional differences were observed in inbound passenger and total international passenger, showing increasing trends over time with a peak in June-August 2022 and majorly arriving in the same states (e.g., CA, FL and TX) (Figs. S16-S17). These variations were likely associated with the recovering international tourism and business activities, particularly after the lifting of test restriction and regional aviation differences (Flaherty et al., 2022).

3.2. Correlations between C_{RNA} and incidence of examined counties during pandemic and post-pandemic periods

The C_{RNA}-incidence correlation coefficients were calculated every 5 weeks for a certain county under various leading times (2–14 d, weekly average, and 15-d-rolling average). Compared with other incidences, we observed higher correlations between C_{RNA} and following 7-day cases (R_{f7}) and following 14-day cases (R_{f14}) among examined counties (Result S1, Fig. S18). This was consistent with the observations in previous studies (Jiang et al., 2022) and attributable to the viral incubation time and time lag between symptom onsite and testing (Badr et al., 2020). Thus, R_{f7} and R_{f14} were selected for the following correlation analysis.

The USA government lifted the restrictions on negative testing results for international passengers before boarding a flight to USA after June 2022 (Bart et al., 2023), meaning that even positive visitors can freely enter USA. The report of USA department of transportation (DOT) showed that monthly average air passenger after June 2022 increased to 18 M which was close to the pre-pandemic levels (20 M) (Statistics, 2023). Thus, the periods before and after June 2022 resembled the pandemic period and post-pandemic period, respectively. Clear temporal and regional differences were observed for R_{f7} and R_{f14} among examined counties during two time periods (Fig. 2). Before June 2022, R_{f7} was 0.70 (IQR 0.30–0.88) and R_{f14} was 0.74 (IQR 0.31–0.90) (Fig. 2a-b), suggesting the good predictability of WBE for COVID-19 cases in following 7-day and 14-day during the pandemic period. After June 2022, the R_{f7} and R_{f14} reduced to 0.01 (IQR: −0.31–0.36) and −0.01 (IQR: −0.38–0.45), respectively. In particular, 61 % and 59 % of examined counties showed negative correlations for R_{f7} (−0.20, IQR: −0.41–−0.01) and R_{f14} (−0.26, IQR: −0.44–−0.13) after June 2022, and 24 counties even showed strong negative correlations of R_{f7} and R_{f14} ($R<-0.5$) (Fig. 2c-d). Such decreases in correlation coefficients in post-pandemic times also showed clear regional differences within these counties. For instance, California had the largest changes in the number of international passengers before and after the lifting of travel restrictions, where the R_{f7} decreased from 0.61 (IQR: 0.31–0.87) before June 2022 to −0.05 (IQR: −0.21–0.29) after June 2022, and R_{f14} from 0.71 (IQR: 0.30–0.88) before June 2022 to −0.01 (IQR: −0.29–0.39) after June 2022. This suggested the potential contributions of international air travel. The variations in R_{f7} and R_{f14} before and after June 2022 were also observed in the same county. For example, in East Baton Rouge Parish in Louisiana, the R_{f7} (0.79) was lower than R_{f14} (0.84) before June 2022 while the R_{f7} (0.53) was higher than R_{f14} (0.31) after June 2022.

3.3. Contribution of explanatory factors for reported cases

Random forest models for describing the reported cases in following 7-day and 14-day (cases/per 100k population) under the corresponding C_{RNA}, and demographic, CCVI, epidemiological, environmental factors, test accessibility, and international air travel (a total of 37 input factors, Table S2) were established for the time periods before (pandemic) and after June 2022 (post-pandemic) separately. The models achieved good fitting performance with R (correlation between reported cases and predicted cases) ranging from 0.93 to 0.98, and root mean square error (RMSE) ranging from 98.6 to 131 (Table S4), suggesting that the explanatory factors reasonably captured the changes of reported cases in these time periods.

Clear shifts in the explanatory factors and their importance (reflected by the increase in mean-squared-error MSE, %) for reported cases under the corresponding C_{RNA} were observed before and after June 2022 (Fig. 3). Before June 2022, positive testing ratio and C_{RNA} were found as the most important factors ($P < 0.01$) for following 7-day cases with a % IncMSE of 23 %, followed by the variant ratio of Omicron and Delta (% IncMSE: 20 %, $P < 0.01$), and other CCVI, demographic, and international air travel factors with %IncMSE ranging from 8.7 to 14.8 % and significance ($P < 0.05$) (Fig. 3a). Similarly, positive testing ratio, C_{RNA}, and variant ratio of Omicron and Delta were found as the most important factors (%IncMSE: 18–20 %) for following 14-day cases, followed by other 7 important factors, including demographic, CCVI, international air travel and test accessibility factors with importance (% IncMSE: 8.8–13.9 %) before June 2022 (Fig. 3c). After June 2022, although positive testing ratio, C_{RNA}, and variant ratio of Omicron and Delta were found as factors with significant contributions, their importance was much lower, with %IncMSE ranging from 6.8 to 14.2 % (Fig. 3b and d). Furthermore, other factors including international air travel, minority and demographic factors also showed significant contributions of 7.7–11.3 % for following 7-day and 14-day cases after June 2022. The detailed contributions of these significant factors are shown and discussed in the following sections.

In addition, 14 and 13 factors were found insignificant ($P > 0.05$) for following 7-day and 14-day cases, including all environmental factors, epidemiological factors (excluding ratios of Omicron and Delta), CCVI factors (excluding healthcare system and minority status), and population ratios aged between 5 and 17 and below 5 (Fig. 3). This indicated that their impacts on reported cases were limited in comparison to significant factors, thus were not investigated in the following sections.

3.3.1. Impact of international passenger on reported cases under the corresponding C_{RNA}

Inbound passenger and total international passenger were all identified as significant factors ($P < 0.05$) before and after June 2022. The inbound passenger exhibited high importance for following 7-day and 14-day cases with %IncMSE of 10.3–11.3 % before June 2002 and 7.1–10.3 % after June 2022 (Fig. 3). Also, total international passenger showed high importance for following 7-day and 14-day cases with % IncMSE of 8.7–10.1 % before June 2002 and 6.9–8.7 % after June 2022 (Fig. 3). The increases of inbound passenger and total international passenger clearly increased the number of cases in the following 7-day and 14-day under the same C_{RNA} (Fig. 4). Such effect was more obvious when the inbound passenger exceeded 0.7 M and total international passenger exceeded 1.7 M during pandemic period (before June 2022) (Fig. 4a-b and e-f), and when the inbound passenger exceeded 1.5 M and total international passenger exceeded 2.8 M in the post-pandemic period (after June 2022) (Fig. 4c-d and g-h). This is likely due to the imported infections from international passengers, who tested positive in a short period after landing, but their viral shedding may not have been captured by the wastewater sampling (7 or 14 days ago). During the pandemic, although negative tests and proof of full vaccination were required (Bart et al., 2023), imported infection from international travel was commonly observed (Ahmed et al., 2022; Farkas

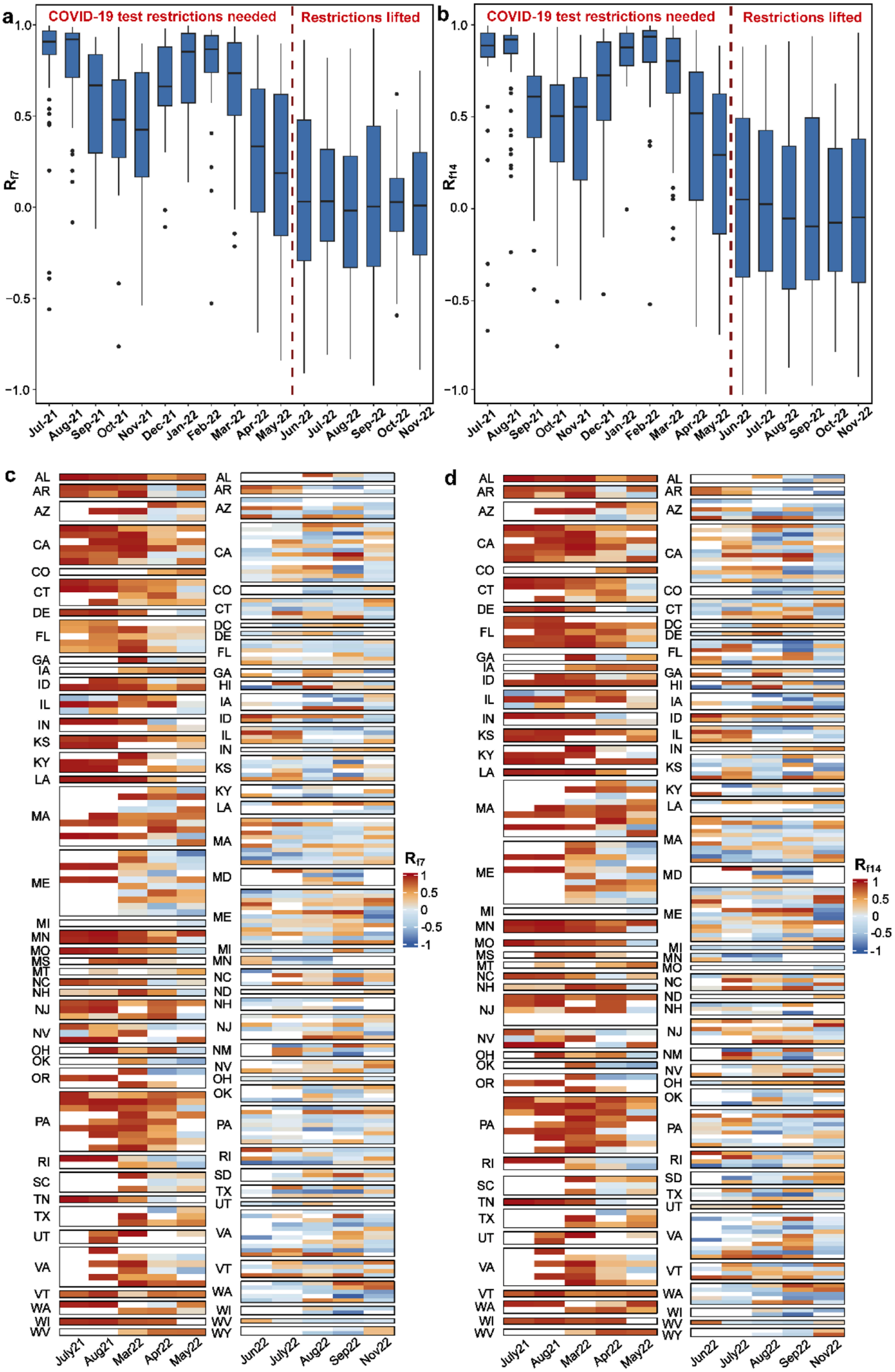


Fig. 2. The temporal variations of the correlation between C_{RNA} and incidence in following 7-day (R_{f7}) and following 14-day (R_{f14}). The temporal variations of (a) overall R_{f7}, (b) overall R_{f14}, (c) county-level R_{f7} and (d) county-level R_{f14} (white-colored cells reflect missing values).

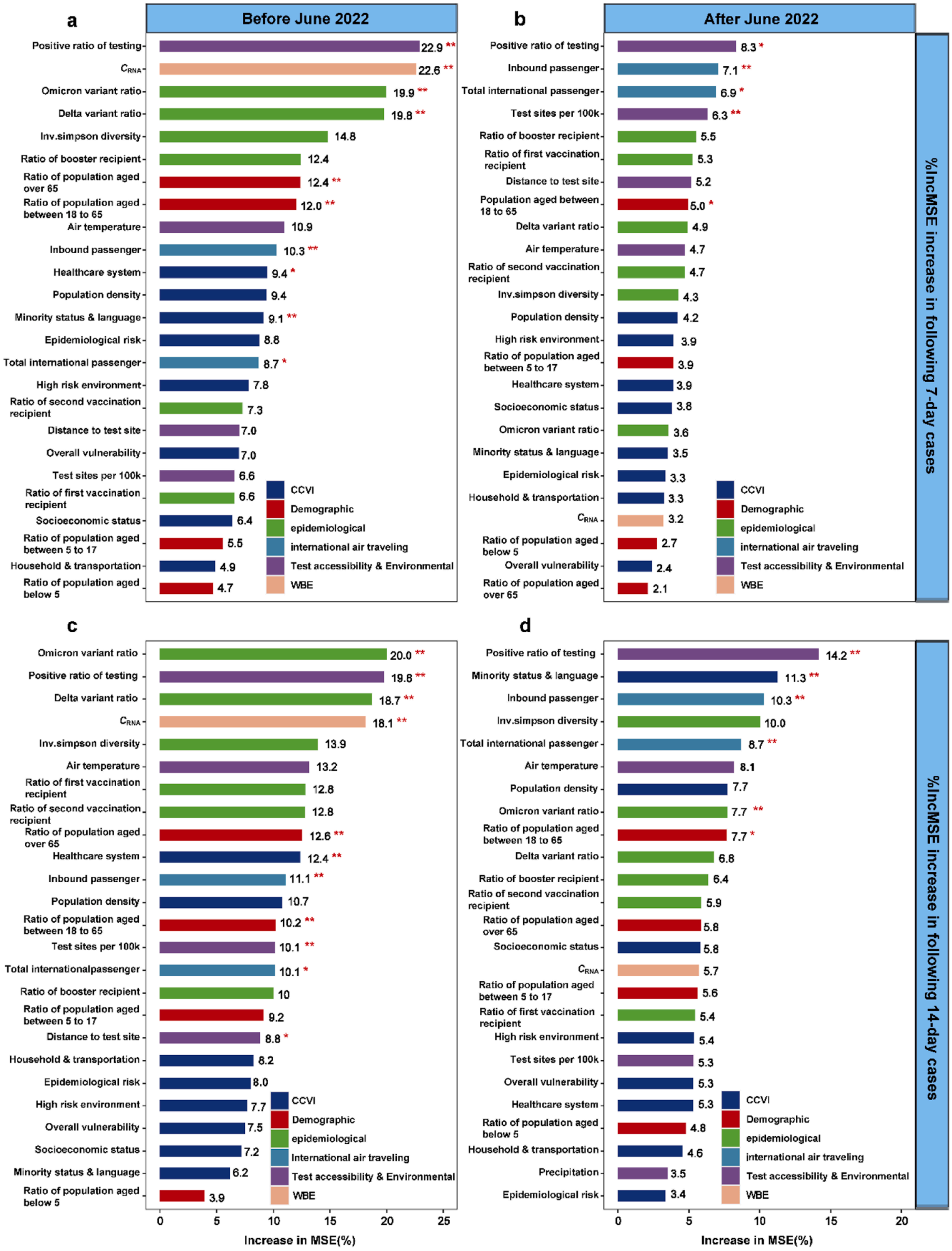
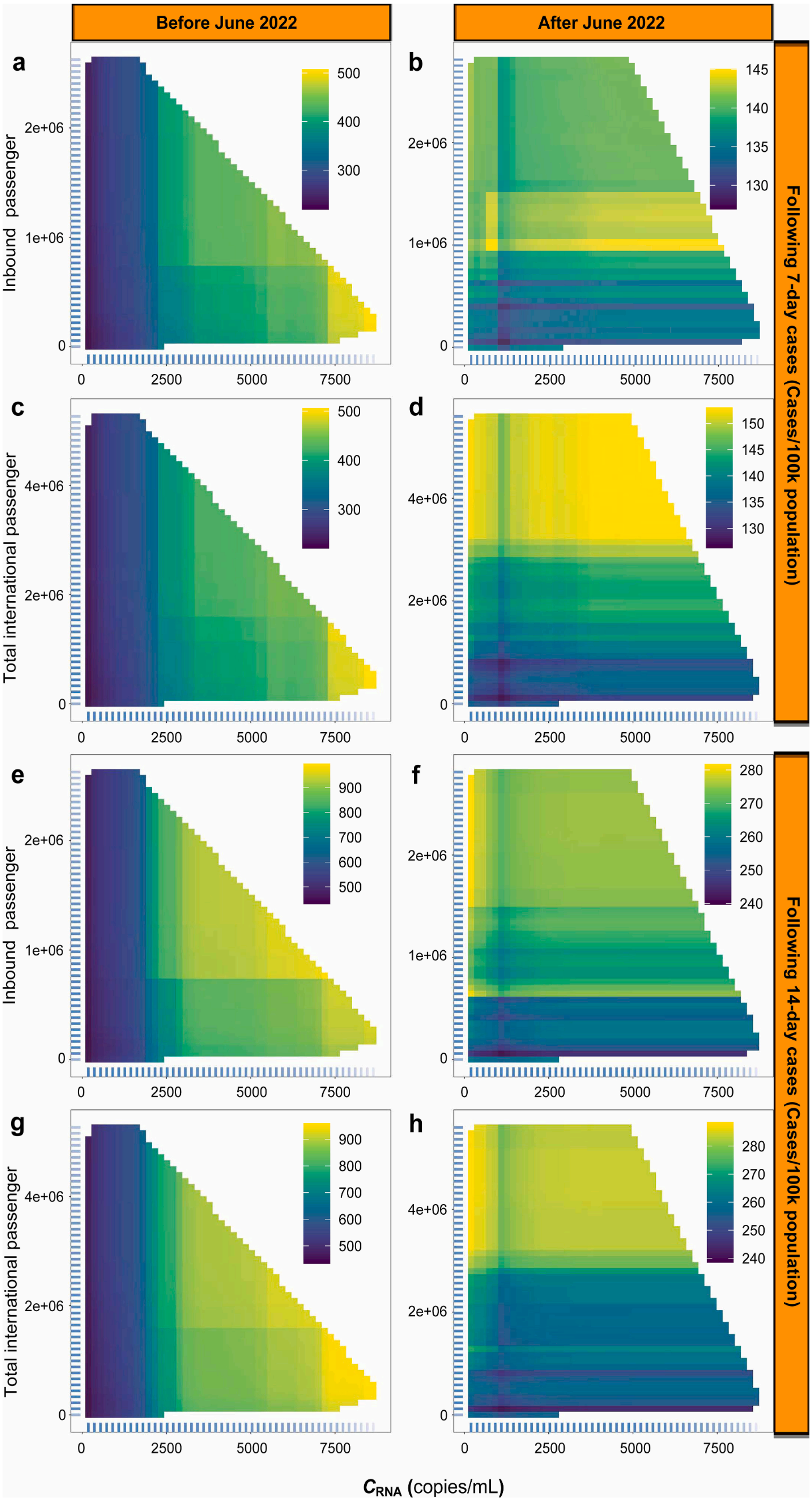


Fig. 3. The factor contribution to the percentage increase in mean squared error (%IncMSE) on following 7-day and 14-day cases before and after June 2022 (a) following 7-day case before June 2022 (b) following 7-day case after June 2022 (c) following 14-day case before June 2022 and (d) following 14-day case after June 2022. Symbol * means 0.01<$P < 0.05$ and ** means $P < 0.01$.

et al., 2023), potentially due to the long incubation period of COVID-19 or false negative results (Arevalo-Rodriguez et al., 2020; Badr et al., 2020). Such imported infections would be more severe in post-pandemic period. The inbound passenger data inherently contains fewer data points at higher ranges (Fig. S16), limiting the model's ability to learn relevant patterns and make accurate predictions (Molnar, 2020). This likely leads to the block of higher new infections observed middle-numbers of inbound passengers in Fig. 4b. Also, inbound passenger showed slightly higher correlations with following 7-day cases ($R = 0.30$, $P < 0.001$) and 14-day cases ($R = 0.32$, $P < 0.001$) compared to total international passengers ($R = 0.22$, $P < 0.001$ and $R = 0.25$, $P < 0.001$, respectively). This is consistent with the higher contribution of



(caption on next page)

Fig. 4. The partial dependence of inbound passenger and total international passenger along with the C_{RNA} on following 7-day and 14-day cases before and after June 2022. Partial dependence of inbound passenger and C_{RNA} on: (a-b) following 7-day cases (cases/per 100k population, the same hereafter) before and after June 2022; Total international passenger and C_{RNA} on: (c-d) following 7-day cases before and after June 2022; Inbound passenger and C_{RNA} on: (e-f) following 14-day cases before and after June 2022; Total international passenger and C_{RNA} on: (g-h) following 14-day cases before and after June 2022. The test restriction of COVID-19 before boarding a flight to the USA was lifted in June 2022. The darker and lighter ticks along the axes indicate more data points and fewer data points in this value range, respectively.

inbound passenger than total international passenger (Fig. 3). This difference is likely because total international passengers include outbound passenger as well. Inbound passengers are more likely to contribute to community transmission compared to outbound passengers (Russell et al., 2021).

3.3.2. Impact of test accessibility factors on reported cases under the corresponding C_{RNA}

For test accessibility factors, positive ratio of testing was identified as the most significant factor ($P < 0.01$) with high %IncMSE for following 7-day cases (22.9 % before June 2022 and 8.3 % after June 2022) and following 14-day cases (19.8 % before June 2022 and 14.2 % after June 2022) (Fig. 3). Under the same C_{RNA}, higher positive ratio of testing increased more following 7-day and following 14-day cases and such increases were more obvious when the positive ratio of testing exceeded 0.25 before and after June 2022 (Fig. S19a-b and e-f). Contrastingly, under the same $C_{\mathrm{RNA},}$ test sites per 100k showed limited change on following 7-day and 14-day cases before June 2022, while it obviously increased following 7-day and 14-day after June 2022 when the test sites per 100k exceeded 30 (Fig. S19c-d and g-h). This indicated that in post-pandemic times, a better test accessibility would help to report/record more COVID-19 infections among the 'true infections' (reflected by C_{RNA}).

3.3.3. Impact of epidemiological factors on reported cases under the corresponding C_{RNA}

Variant ratios of Omicron and Delta were the most significant epidemiological factors ($P < 0.01$) with high %IncMSE ranging from 18.7 to 20.0 % for following 7-day and 14-day cases before June 2022 (Fig. 3a and c). Under the same C_{RNA}, a higher ratio of Delta variant was associated with lower cases in following 7-day and 14-day while higher ratio of Omicron variant increased following 7-day and 14-day cases before June 2022 (Fig. 5). Such changes were more obvious when the ratio of Delta and Omicron exceeded 0.37 and 0.62, respectively (Fig. 5a and e). This is likely due to the lower viral shedding loads of Omicron variants than Delta variants and the progression of the vaccination along with the increasing dominance of Omicron variants (Figs. S3-S6). Contrastingly, only Omicron variant ratio was significant ($P < 0.01$) for following 14-day cases with %IncMSE of 7.7 % after June 2022 (Fig. 3). This is consistent with the observed variant shift, where Omicron variants became dominant with a ratio over 0.99 after June 2022 (Fig. S6). The ratios of Delta variant and Omicron variant inherently contains fewer data points at higher ranges and lower rangers (Fig. S8–9), respectively, limiting the model's ability to learn relevant patterns and make accurate predictions to illustrate the transition from Delta variant to Omicron variant during the pandemic.

3.3.4. Impact of demographics on reported cases under the corresponding C_{RNA}

The ratios of population aged between 18 and 65 and aged over 65 were significant ($P < 0.01$) with %IncMSE ranging from 10.2 to 12.6 % for the reported cases before June 2022 (Fig. 3a and c), while the ratio of population age between 18 and 65 had %IncMSE ranging from 5.0 to 7.7 % after June 2022 ($P < 0.05$) (Fig. 3b and d). Under the same C_{RNA}, the increase of population ratio between 18 and 65 showed limited impact on the reported cases, until population ratio between 18 and 65 was above 0.48, where it notably reduced the reported cases in the following 7 days and 14 days before and after June 2022 (Fig. S20a-b and e-f). Contrastingly, the population aged over 65 showed limited impact on reported cases while increased the reported cases until its ratio exceeded 0.37 before and after June 2022 (Fig. S20c-d and g-h).

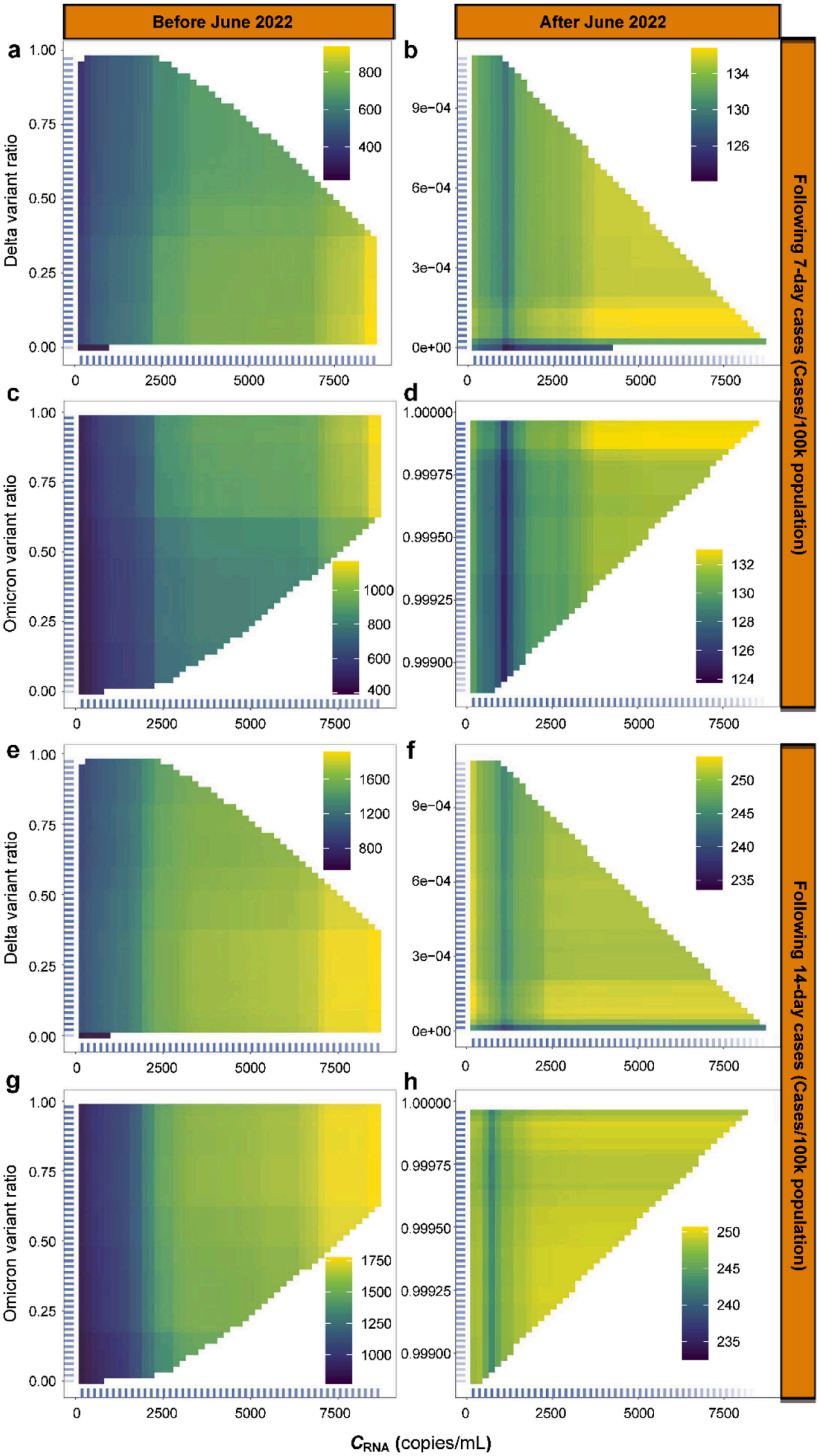
3.3.5. Impact of CCVI factors on reported cases under the corresponding C_{RNA}

For CCVI factors, the vulnerability in healthcare system and minority status and language were significant factors ($P < 0.05$) with %IncMSE ranging from 9.1 to 12.4 % before June 2022 (Fig. 3a and c), while minority status and language showed significance with the second highest %IncMSE of 11.3 % for following 14-day cases after June 2022 (Fig. 3b and d). Under the same C_{RNA}, the following 7-day and 14-day cases slightly increased when the minority status and language exceeded 0.80 before June 2022, while this increase became notable after June 2022 (Fig. 6a-b and e-f), indicating that counties with higher vulnerability of minority status and language would have more reported cases under the same C_{RNA}, especially in post-pandemic times. Contrastingly, under the same C_{RNA}, following 7-day and 14-day cases showed minor variations with the changes in healthcare system vulnerability before June 2022, but they notably decreased after June 2022 when the healthcare system vulnerability exceeded 0.83 (Fig. 6c-d and g-h). Such decreases suggested that counties with more vulnerable healthcare system reported fewer cases than those with better healthcare systems, especially in post-pandemic times.

4. Discussions

To the best of our knowledge, this is the first study to comprehensively explore the impacts of demographic, socioeconomic, epidemiological, environmental, test accessibility, and international air travel factors on the reported cases under the corresponding C_{RNA}. In the post-pandemic era, COVID-19 is no longer classified as a global health emergency, and most travel restrictions have been lifted (Harris, 2023). As public vigilance towards COVID-19 decreases and expanding test accessibility becomes less feasible due to the high economic burden (US $46,640 for only PCR testing cost in a 1000-population community) (Purba et al., 2024), WBE surveillance is expected to play an increasingly important role in monitoring and predicting community infections. Traditionally, WBE-based surveillance relies on the close-to-linear relationship (i.e., strong positive correlations) between the C_{RNA} and reported cases to reflect or predict infections within the community (Polo et al., 2020). However, our study found that this correlation reduced significantly during post-pandemic period in comparison to the pandemic period. The C_{RNA}-Incidence correlation was 0.70 (IQR: 0.30–0.88) and 0.74 (IQR: 0.31–0.90) in pandemic period for cases in the following 7 days and 14 days, respectively. However, this correlation reduced to 0.01 (IQR: −0.31–0.36) and −0.01 (IQR: −0.38–0.45) in post-pandemic period, for cases in the following 7 days and 14 days, respectively (Fig. 2a-b). These findings suggest that the traditional use of the C_{RNA}-Incidence correlation faces challenges in the post-pandemic period, where a direct linear relationship cannot be established between C_{RNA} and reported incidence, indicating the potential influence of other factors.

To explore the reasons behind the dramatic reduction in the C_{RNA}-Incidence correlation during the post-pandemic period, we further examined the contributions of demographics, test accessibility, epidemiological factors, environmental factors, international air travel, and C_{RNA} to reported cases before and after the pandemic. During the



(caption on next page)

Fig. 5. The partial dependence of the variant ratio of Delta and Omicron along with the C_{RNA} on following 7-day and 14-day cases before and after June 2022. Partial dependence results of Delta variant ratio and C_{RNA} on: (a-b) following 7-day (cases/per 100k population, the same hereafter) cases before and after June 2022; Omicron variant ratio and C_{RNA} on: (c-d) following 7-day cases before and after June 2022; Ratio of Delta variant ratio and C_{RNA} on: (e-f) following 14-day cases before and after June 2022; Omicron variant ratio and C_{RNA} and C_{RNA} on: (g-h) following 14-day cases before and after June 2022. The test restriction of COVID-19 before boarding a flight to the USA was lifted in June 2022. The darker and lighter ticks along the axes indicate more data points and fewer data points in this value range, respectively.

pandemic, C_{RNA} was found to be one of the most important factors ($P <$ 0.001) for reported cases in both the following 7 days (%IncMSE: 22.6 %) and 14 days (%IncMSE: 18.1 %) (Fig. 3a and c). However, in the post-pandemic period, C_{RNA} became less important ($P >$ 0.05) for reported cases in both the following 7 days (%IncMSE: 3.2 %) and 14 days (% IncMSE: 5.7 %) (Fig. 3b and d). Meanwhile, factors such as the positive testing ratio, inbound passengers, total international passengers, Omicron variant ratio, minority population, and the ratio of the population aged 18 to 65 remained significant ($P <$ 0.05) in relation to reported cases, contributing 8.7–22.9 % to %IncMSE individually in the post-pandemic period. This aligns with the reduced C_{RNA}-Incidence correlations in the post-pandemic period, suggesting that additional factors should be considered when applying WBE and using C_{RNA}-Incidence correlations for COVID-19 community surveillance.

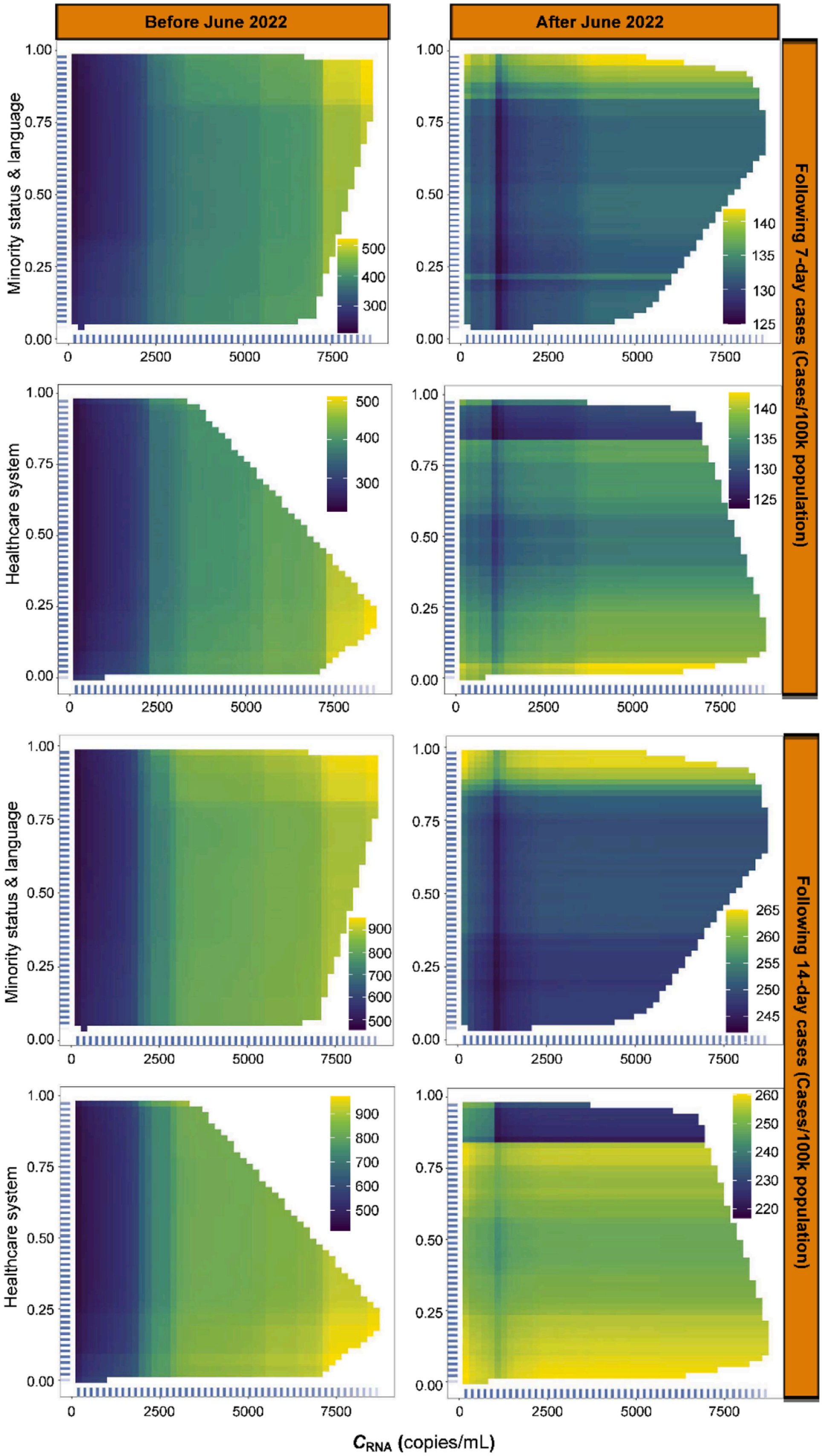
In post-pandemic period, we observed that C_{RNA}-Incidence correlation shifted from 0.70 (IQR: 0.30–0.88) to 0.01 (IQR: −0.31–0.36) for cases in the following 7 days and 0.74 (IQR: 0.31–0.90) to −0.01 (IQR: −0.38–0.45) in the following 14 days, respectively (Fig. 2a-b). Furthermore, compared with pandemic period, C_{RNA} became less important ($P >$ 0.05) to reported cases in both following 7-day (% IncMSE: 3.2 %) and 14-day (%IncMSE: 5.7 %) in post-pandemic period (Fig. 3b and d). Meanwhile, positive ratio of testing, inbound passenger, total international passenger, Omicron variant ratio, minority, and ratio of population aged between 18 and 65 remained as significant factors ($P <$ 0.05) to reported cases with a contribution of 8.7–22.9 % in %IncMSE individually in the post-pandemic period. This indicated that reported cases could not be well reflected by C_{RNA} in post-pandemic period, while socioeconomic, international air travel, epidemiological and demographic factors introduced more noise in the C_{RNA}-Incidence correlation and WBE application.

In post-pandemic era (after June 2022), international travel has returned to its pre-pandemic status. The model analysis further revealed that an increase in international passenger increased future reported cases under the same $C_{RNA,}$ particularly after June 2022 (Fig. 4). This is likely caused by the imported infections from positive international passengers, which led to more reported cases under the same C_{RNA} (captured a few days before their landing). After June 2022, negative test result before boarding a flight to USA is not required for international passengers, meaning that even infected passengers can fly to USA (Bart et al., 2023). The USA DOT report indicated that monthly air passenger numbers increased to 18 M after June 2022, nearing pre-pandemic levels (20 M) (Statistics, 2023). Meanwhile, pre-departure testing decreased by 80 % within three months of lifting test restrictions, and the positive test ratio among inbound passengers increased by 52 % (Bart et al., 2023). The viral shedding of infected international passengers was not captured by the wastewater sampling before landing, eventually leading to higher reported cases than the infections reflected by C_{RNA}.

In post-pandemic period, a higher ratio of Omicron infection led to higher reported cases compared to the Delta infection in the following 7-day and 14-day under the same C_{RNA} (Fig. 5). This is likely related to the lower viral shedding in Omicron-infected patients than in Delta-infected patients and the progression of vaccination along with the increased dominance of Omicron variants. Viral shedding in Delta-infected individuals was 52 % higher than in Omicron-infected individuals in nasal secretions (Martins et al., 2022). This implies that data on circulating variants is important for estimating COVID-19 community infection through WBE. In particular, the frequency of COVID-19 variant changes is expected to increase in the future (Carabelli et al., 2023). However, current variant monitoring primarily relies on clinical reports from patients, which can face delays of weeks due to factors like symptom onset and the availability of analytical facilities (Mercer and Salit, 2021; Robishaw et al., 2021). Recently, a few countries, such as Canada and Ireland, have implemented wastewater surveillance for COVID-19 variants of concern (Hasing et al., 2023; Reynolds et al., 2022). This approach could serve as a valuable supplement for predicting COVID-19 cases in communities through WBE. Also, the increased dominance of Omicron coincided with the progression of vaccination (Fig. S3-S6). Omicron variants have become dominant since January 2022, when the first-dose vaccination, second-dose vaccination and booster vaccination coverage reached 73.3 % (IQR: 60.9–84.3 %), 71.8 % (IQR: 63.3–95.0 %), and 44.8 % (IQR: 40.0–50.3 %), respectively (Fig. S3-S5). The contribution of vaccination coverage on the reported incidences was not significant ($P >$ 0.05, %IncMSE range of 4.7–12.8 %) (Fig. 3), viral load tested in individuals in Switzerland was revealed to be lower in fully vaccinated/booster-vaccinated individuals than in unvaccinated individuals for both Omicron and Delta infections (Puhach et al., 2022). Therefore, it is also important to consider vaccination coverage when utilizing WBE.

Case under-reporting is another major challenge for WBE surveillance in estimating infections in post-pandemic era. Generally, case under-reporting is unavoidable for infectious disease surveillance (Milanesi and De Nicolao, 2023) as the testing largely relies on individual willingness. The reported cases reflect only a portion of the ‘true’ infections (McManus et al., 2023). Theoretically, if a consistent proportion of actual infections is identified by clinical testing over time, reported COVID-19 cases could align closely with the ‘true infections’ and corresponding C_{RNA}. However, our study observed that fewer cases were reported in counties with more vulnerable healthcare systems and fewer test sites under the same C_{RNA} (Fig. 6 and Fig. S19). This is likely attributed to the insufficient test accessibility and lower testing willingness in these counties, consistent with previous findings that testing rates and resident willingness are positively correlated with testing availability (Hendricks et al., 2023). Such observations are also reflected by the impacts of populations aged between 18 and 65 and minority vulnerability in our study. An increase of populations aged between 18 and 65 decreased the reported cases under the same C_{RNA} (Fig. S20). This age group comprises the primary workforce and student population, and a positive test result could disrupt their work or education schedules and even lead to job loss due to mandatory 5-day isolation (Capasso et al., 2022; Walensky, 2021), resulting in a lower willingness to testing. In comparison, increase in the ratio of elderly people among population increased the reported cases under the same C_{RNA} (Fig. S20). Elderly people are more likely to experience severe symptoms due to underlying disease, which increased their likelihood and awareness of seeking COVID-19 testing (Lu et al., 2022). In addition, increases in minority vulnerability increased the reported cases under the same $C_{RNA,}$ particularly in following 14-day. Previous studies revealed that ethnic minorities worked and lived in environments exposed to higher infection risk due to socioeconomic disadvantage, high reliance on public transportation, and crowded households (Goldman et al., 2021; Tai et al., 2021). Thus, they are more likely to experience severe symptoms, which increases their likelihood of seeking COVID-19 testing in following 14-day.

This study provides a retrospective analysis of the C_{RNA}-incidence correlation in WBE applications across 222 counties in the USA,



(caption on next page)

Fig. 6. The partial dependence of the vulnerability of minority status and language and healthcare system along with the C_{RNA} on following 7-day and 14-day cases before and after June 2022. Partial dependence results of minority status and language vulnerability and C_{RNA} on: (a-b) following 7-day cases (cases/per 100k population, the same hereafter) before and after June 2022; Healthcare system vulnerability and C_{RNA} on: (c-d) following 7-day cases before and after June 2022; Minority status and language vulnerability and C_{RNA} on: (e-f) following 14-day cases before and after June 2022; Healthcare system vulnerability and C_{RNA} and C_{RNA} on: (g-h) following 14-day cases before and after June 2022. The test restriction of COVID-19 before boarding a flight to the USA was lifted in June 2022. The darker and lighter ticks along the axes indicate more data points and fewer data points in this value range, respectively.

highlighting the importance for public health sectors to consider factors such as international travel, variant mutations, test accessibility, and the distribution of age and minority groups when utilizing WBE models for COVID-19 surveillance. Potential application of machine-learning or multi-factor linear models, rather than relying solely on C_{RNA}-incidence correlation can help with this. A recent study conducted in the United Kingdom has revealed that multi-factor model showed more accurate COVID-19 case prediction after considering the factors of wastewater sampling, sample processing, and catchment population and facilities (Pellett et al., 2024). Furthermore, correction factors can be introduced when applying WBE for estimating community infections in regions with high population mobility, less test accessibility, or higher distribution of age and minority groups.

There are several limitations in this study. Since the county-level aviation data is unavailable, we use the state-level international passenger data from international airport statistics to represent the holistic international passenger for all studied counties in the same state to investigate the impact of international air travel. However, this sometimes might not be appropriate to some remote counties with limited international passengers (e.g., counties in South Dakota). Thus, future studies are encouraged to use county-level data to directly study the impacts of international passengers when statistical reports are available in the county/state's transportation department. The R and RMSE achieved in our models after June 2022 were both lower than those achieved before June 2022. This suggests the potential contributions of other factors that were not included in the model and require future investigations. Also, we only considered international passenger from air travel due to the data availability, while some international passengers might enter USA by road/water transportation (e.g., road entry in USA/Canada border) (Gurbuz et al., 2023). Future studies are recommended to incorporate these passengers with air passengers to obtain a comprehensive exploration of international passenger impacts on WBE.

5. Conclusions

This study for the first time comprehensively investigated variations of C_{RNA}-Incidence correlation and the impacts of demographic, socioeconomic, epidemiological, environmental, test accessibility, and international air travel factors on the reported cases under the corresponding C_{RNA} in 222 USA counties during the pandemic period (June 2021-May 2022) and post-pandemic period (June-November 2022). This leads to the following key findings:

- The lifting of travel restrictions in June 2022 notably reduced the C_{RNA}-Incidence correlations in following 7-day from 0.70 (IQR 0.30–0.88) to 0.01 (IQR: −0.31–0.36) and in following 14-day from 0.74 (IQR 0.31–0.90) to −0.01 (IQR: −0.38–0.45).
- Socioeconomic, international air travel, epidemiological and demographic factors played major roles in the variations of reported cases in the in following 7-day and 14-day under the corresponding C_{RNA} in the post-pandemic period (after lifting travel restrictions).
- Socioeconomic, international air travel, epidemiological and demographic factors also introduced more noise for COVID-19 incidence monitoring through WBE, due to imported infections, visitor mobility, variant mutation, and case under-reporting.
- Our findings will facilitate the global health departments to dynamically update their WBE models based on local context, improving the feasibility and certainty when using WBE for infection surveillance.

CRediT authorship contribution statement

Xuan Li: Writing – review & editing, Methodology, Data curation, Conceptualization. **Jibin Li:** Writing – original draft, Visualization, Data curation. **Huan Liu:** Data curation. **Lidia Mínguez-Alarcón:** Writing – review & editing. **Mark C.M. van Loosdrecht:** Writing – review & editing. **Qilin Wang:** Writing – review & editing, Supervision, Conceptualization.

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.

Acknowledgement

X.L. thanks the funding support from the Australian Academy of Science through W H Gladstones Population and Environment Fund. J.L. acknowledges the China Scholarship Council. Q.W. acknowledges the Australian Research Council Future Fellowship (FT200100264). The authors also acknowledge the valuable suggestions provided by Dr Li Gao at South East Water during the manuscript preparation.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2025.123114.

Data availability

Data will be made available on request.

References

Ahmed, W., Bivins, A., Simpson, S.L., Bertsch, P.M., Ehret, J., Hosegood, I., Metcalfe, S. S., Smith, W.J., Thomas, K.V., Tynan, J., 2022. Wastewater surveillance demonstrates high predictive value for COVID-19 infection on board repatriation flights to Australia. Environ. Int. 158, 106938.

Ali, J., Khan, R., Ahmad, N., Maqsood, I., 2012. Random forests and decision trees. IJCSI 9 (5), 272.

Analytics, B., 2021. Population Health Analytics Powered By Sewage.

Archer, E. and Archer, M.E. 2016. Package 'rfPermute'. R Project: Indianapolis, IN, USA.

Arevalo-Rodriguez, I., Buitrago-Garcia, D., Simancas-Racines, D., Zambrano-Achig, P., Del Campo, R., Ciapponi, A., Sued, O., Martinez-Garcia, L., Rutjes, A.W., Low, N., 2020. False-negative results of initial RT-PCR assays for COVID-19: a systematic review. PLoS One 15 (12), e0242958.

Arvisais-Anhalt, S., Lehmann, C.U., Park, J.Y., Araj, E., Holcomb, M., Jamieson, A.R., McDonald, S., Medford, R.J., Perl, T.M., Toomay, S.M., 2021. What the coronavirus disease 2019 (COVID-19) pandemic has reinforced: the need for accurate data. Clin. Infect. Dis. 72 (6), 920–923.

Badr, H.S., Du, H., Marshall, M., Dong, E., Squire, M.M., Gardner, L.M., 2020. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. Lancet Infect. Dis. 20 (11), 1247–1254.

Bart, S.M., Smith, T.C., Guagliardo, S.A.J., Walker, A.T., Rome, B.H., Li, S.L., Aichele, T. W., Stein, R., Ernst, E.T., Morfino, R.C., 2023. Effect of predeparture testing on postarrival SARS-CoV-2–positive test results among international travelers—CDC traveler-based genomic surveillance program, Four US Airports, March–September 2022. Morb. Mortal. Wkly. Rep. 72 (8), 206.

Bibby, K., Bivins, A., Wu, Z., North, D., 2021. Making waves: plausible lead time for wastewater based epidemiology as an early warning system for COVID-19. Water Res. 202, 117438.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Capasso, A., Kim, S., Ali, S.H., Jones, A.M., DiClemente, R.J., Tozan, Y., 2022. Employment conditions as barriers to the adoption of COVID-19 mitigation measures: how the COVID-19 pandemic may be deepening health disparities among low-income earners and essential workers in the United States. BMC Public Health 22 (1), 1–13.

Carabelli, A.M., Peacock, T.P., Thorne, L.G., Harvey, W.T., Hughes, J., , 6, C.-G.U.C.d.S.T.I, Peacock, S.J., Barclay, W.S., de Silva, T.I., Towers, G.J., 2023. SARS-CoV-2 variant biology: immune escape, transmission and fitness. Nat. Rev. Microbiol. 21 (3), 162–177.

Cheng, C.-H., Chan, C.-P., Sheu, Y.-J., 2019. A novel purity-based k nearest neighbors imputation method and its application in financial distress prediction. Eng. Appl. Artif. Intell. 81, 283–299.

de Freitas Bueno, R., Claro, I.C.M., Augusto, M.R., Duran, A.F.A., Camillo, L.d.M.B., Cabral, A.D., Sodré, F.F., Brandão, C.C.S., Vizzotto, C.S., Silveira, R., de Melo Mendes, G., Arruda, A.F., de Brito, N.N., Machado, B.A.S., Duarte, G.R.M., de, Lourdes, Aguiar-Oliveira, M., 2022. Wastewater-based epidemiology: a Brazilian SARS-COV-2 surveillance experience. J. Environ. Chem. Eng. 10 (5), 108298.

Duvallet, C., Wu, F., McElroy, K.A., Imakaev, M., Endo, N., Xiao, A., Zhang, J., Floyd-O'Sullivan, R., Powell, M.M., Mendola, S., 2022. Nationwide trends in COVID-19 cases and SARS-CoV-2 RNA wastewater concentrations in the United States. ACS ES&T Water 2 (11), 1899–1909.

Farkas, K., Williams, R., Alex-Sanders, N., Grimsley, J.M., Pântea, I., Wade, M.J., Woodhall, N., Jones, D.L., 2023. Wastewater-based monitoring of SARS-CoV-2 at UK airports and its potential role in international public health surveillance. PLOS Glob. Public Health 3 (1), e0001346.

Flaherty, G.T., Hamer, D.H., Chen, L.H., 2022. Travel in the time of COVID: a review of international travel health in a global pandemic. Curr. Infect. Dis. Rep. 24 (10), 129–145.

Goldman, N., Pebley, A.R., Lee, K., Andrasfay, T., Pratt, B., 2021. Racial and ethnic differentials in COVID-19-related job exposures by occupational standing in the US. PLoS One 16 (9), e0256085.

Gurbuz, O., Aldrete, R.M., Salgado, D., Gurbuz, T.M., 2023. Transportation as a disease vector in COVID-19: border mobility and disease spread. Transp. Res. Rec. 2677 (4), 826–838.

Harris, E., 2023. CDC updates COVID-19 data reporting as US public health emergency ends. JAMA 329 (22), 1907. -1907.

Hasing, M.E., Lee, B.E., Gao, T., Li, Q., Qiu, Y., Ellehoj, E., Graber, T.E., Fuzzen, M., Servos, M., Landgraff, C., 2023. Wastewater surveillance monitoring of SARS-CoV-2 variants of concern and dynamics of transmission and community burden of COVID-19. Emerg. Microbes Infect. 12 (2), 2233638.

Hendricks, B., Price, B.S., Dotson, T., Kimble, W., Davis, S., Khodaverdi, M., Halasz, A., Smith, G.S., Hodder, S., 2023. If you build it, will they come? Is test site availability a root cause of geographic disparities in COVID-19 testing? Public Health 216, 21–26.

Hillary, L.S., Farkas, K., Maher, K.H., Lucaci, A., Thorpe, J., Distaso, M.A., Gaze, W.H., Paterson, S., Burke, T., Connor, T.R., McDonald, J.E., Malham, S.K., Jones, D.L., 2021. Monitoring SARS-CoV-2 in municipal wastewater to evaluate the success of lockdown measures for controlling COVID-19 in the UK. Water Res. 200, 117214.

Ho, T.K., 1998. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20 (8), 832–844.

Jiang, G., Wu, J., Weidhaas, J., Li, X., Chen, Y., Mueller, J., Li, J., Kumar, M., Zhou, X., Arora, S., 2022. Artificial neural network-based estimation of COVID-19 case numbers and effective reproduction rate using wastewater-based epidemiology. Water Res. 218, 118451.

Kennedy, N.A., Lin, S., Goodhand, J.R., Chanchlani, N., Hamilton, B., Bewshea, C., Nice, R., Chee, D., Cummings, J.F., Fraser, A., 2021. Infliximab is associated with attenuated immunogenicity to BNT162b2 and ChAdOx1 nCoV-19 SARS-CoV-2 vaccines in patients with IBD. Gut 70 (10), 1884–1893.

Li, X., Kulandaivelu, J., Guo, Y., Zhang, S., Shi, J., O'Brien, J., Arora, S., Kumar, M., Sherchan, S.P., Honda, R., 2022. SARS-CoV-2 shedding sources in wastewater and implications for wastewater-based epidemiology. J. Hazard. Mater. 432, 128667.

Li, X., Liu, H., Gao, L., Sherchan, S.P., Zhou, T., Khan, S.J., van Loosdrecht, M.C., Wang, Q, 2023a. Wastewater-based epidemiology predicts COVID-19-induced weekly new hospital admissions in over 150 USA counties. Nat. Commun. 14 (1), 4548.

Li, X., Zhang, S., Sherchan, S., Orive, G., Lertxundi, U., Haramoto, E., Honda, R., Kumar, M., Arora, S., Kitajima, M., Jiang, G., 2023b. Correlation between SARS-CoV-2 RNA concentration in wastewater and COVID-19 cases in community: a systematic review and meta-analysis. J. Hazard. Mater. 441, 129848.

Li, X., Zhang, S., Shi, J., Luby, S.P., Jiang, G., 2021. Uncertainties in estimating SARS-CoV-2 prevalence by wastewater-based epidemiology. Chem. Eng. J. 415, 129039.

Lu, G., Zhang, Y., Zhang, H., Ai, J., He, L., Yuan, X., Bao, S., Chen, X., Wang, H., Cai, J., 2022. Geriatric risk and protective factors for serious COVID-19 outcomes among older adults in Shanghai Omicron wave. Emerg. Microbes Infect. 11 (1), 2045–2054.

Martins, M., do Nascimento, G.M., Nooruzzaman, M., Yuan, F., Chen, C., Caserta, L.C., Miller, A.D., Whittaker, G.R., Fang, Y., Diel, D.G., 2022. The Omicron variant BA. 1.1 presents a lower pathogenicity than B. 1 D614G and Delta variants in a feline model of SARS-CoV-2 infection. J. Virol. 96 (17), e00961. -00922.

McManus, O., Christiansen, L.E., Nauta, M., Krogsgaard, L.W., Bahrenscheer, N.S., von Kappelgaard, L., Christiansen, T., Hansen, M., Hansen, N.C., Kähler, J., 2023. Predicting COVID-19 incidence using wastewater surveillance data, Denmark, October 2021–June 2022. Emerging Infect. Dis. 29 (8), 1589.

Medema, G., Heijnen, L., Elsinga, G., Italiaander, R., Brouwer, A., 2020. 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. 7 (7), 511–516.

Mercer, T.R., Salit, M., 2021. Testing at scale during the COVID-19 pandemic. Nat. Rev. Genet. 22 (7), 415–426.

Milanesi, S., De Nicolao, G., 2023. Correction of Italian under-reporting in the first COVID-19 wave via age-specific deconvolution of hospital admissions. PLoS One 18 (12), e0295079.

Molnar, C., 2020. Interpretable Machine Learning. Lulu.com.

Noor, F.M., Islam, M.M., 2020. Prevalence and associated risk factors of mortality among COVID-19 patients: a meta-analysis. J. Community Health 45 (6), 1270–1282.

Pan, Y., Zhang, D., Yang, P., Poon, L.L., Wang, Q, 2020. Viral load of SARS-CoV-2 in clinical samples. Lancet Infect. Dis. 20 (4), 411–412.

Pellett, C., Farkas, K., Williams, R.C., Wade, M.J., Weightman, A.J., Jameson, E., Cross, G., Jones, D.L., 2024. Multi-factor normalisation of viral counts from wastewater improves the detection accuracy of viral disease in the community. Environ. Technol. Innov. 36, 103720.

Polo, D., Quintela-Baluja, M., Corbishley, A., Jones, D.L., Singer, A.C., Graham, D.W., Romalde, J.L., 2020. Making waves: wastewater-based epidemiology for COVID-19 – approaches and challenges for surveillance and prediction. Water Res. 186, 116404.

Prado, T., Fumian, T.M., Mannarino, C.F., Resende, P.C., Motta, F.C., Eppinghaus, A.L.F., Chagas do Vale, V.H., Braz, R.M.S., de Andrade, J.d.S.R., Maranhão, A.G., Miagostovich, M.P., 2021. Wastewater-based epidemiology as a useful tool to track SARS-CoV-2 and support public health policies at municipal level in Brazil. Water Res. 191, 116810.

Puhach, O., Adea, K., Hulo, N., Sattonnet, P., Genecand, C., Iten, A., Jacquérioz, F., Kaiser, L., Vetter, P., Eckerle, I., 2022. Infectious viral load in unvaccinated and vaccinated individuals infected with ancestral, Delta or Omicron SARS-CoV-2. Nat. Med. 28 (7), 1491–1500.

Purba, A.K.R., Rosyid, A.N., Handayani, S., Rachman, B.E., Romdhoni, A.C., Al Farabi, M.J., Wahyuhadi, J., Prananingtias, R., Rahayu, A.N., Alkaff, F.F., 2024. Economic evaluation of COVID-19 screening tests and surveillance strategies in low-income, middle-income, and high-income countries: a systematic review. Med. Sci. Monit. 30, e943863. -943861.

Reynolds, L.J., Gonzalez, G., Sala-Comorera, L., Martin, N.A., Byrne, A., Fennema, S., Holohan, N., Kuntamukkula, S.R., Sarwar, N., Nolan, T.M., 2022. SARS-CoV-2 variant trends in Ireland: wastewater-based epidemiology and clinical surveillance. Sci. Total Environ. 838, 155828.

Robishaw, J.D., Alter, S.M., Solano, J.J., Shih, R.D., DeMets, D.L., Maki, D.G., Hennekens, C.H., 2021. Genomic surveillance to combat COVID-19: challenges and opportunities. Lancet Microbe. 2 (9), e481–e484.

Russell, T.W., Wu, J.T., Clifford, S., Edmunds, W.J., Kucharski, A.J., Jit, M., 2021. Effect of internationally imported cases on internal spread of COVID-19: a mathematical modelling study. Lancet Public Health 6 (1), e12–e20.

Sanjuán, R., Domingo-Calap, P., 2021. Reliability of wastewater analysis for monitoring COVID-19 incidence revealed by a long-term follow-up study. Front. Virol. 1, 776998.

Silk, B.J., Scobie, H.M., Duck, W.M., Palmer, T., Ahmad, F.B., Binder, A.M., Cisewski, J. A., Kroop, S., Soetebier, K., Park, M., 2023. COVID-19 surveillance after expiration of the public health emergency declaration–United States, May 11, 2023. Morb. Mortal. Wkly. Rep. 72 (19), 523.

Smittenaar, P., Stewart, N., Sutermaster, S., Coome, L., Dibner-Dunlap, A., Jain, M., Caplan, Y., Campigotto, C., Sgaier, S.K., 2021. A COVID-19 community vulnerability index to drive precision policy in the US. medRxiv, 21257455, 2021.2005. 2019.

Statistics, U.S.B.o.T., 2023. U.S. Bureau of Transportation Statistics. US International Air Passenger and Freight Statistics Report. https://www.bts.gov/. accessed 1 November 2023.

Tai, D.B.G., Shah, A., Doubeni, C.A., Sia, I.G., Wieland, M.L., 2021. The disproportionate impact of COVID-19 on racial and ethnic minorities in the United States. Clin. Infect. Dis. 72 (4), 703–706.

Tipirneni, R., Schmidt, H., Lantz, P.M., Karmakar, M., 2022. Associations of 4 geographic social vulnerability indices with US COVID-19 incidence and mortality. Am. J. Public Health 112 (11), 1584–1588.

Trottier, J., Darques, R., Ait Mouheb, N., Partiot, E., Bakhache, W., Deffieu, M.S., Gaudin, R., 2020. Post-lockdown detection of SARS-CoV-2 RNA in the wastewater of Montpellier, France. One Health 10, 100157.

Varkila, M.R., Montez-Rath, M.E., Salomon, J.A., Yu, X., Block, G.A., Owens, D.K., Chertow, G.M., Parsonnet, J., Anand, S., 2023. Use of wastewater metrics to track COVID-19 in the US. JAMA Netw. Open 6 (7), e2325591. -e2325591.

Walensky, R.P. 2021. CDC Updates and Shortens Recommended Isolation and Quarantine Period For General population: Media Statement For Immediate release: Monday, December 27, 2021.

Walensky, R.P. 2022. CDC Streamlines COVID-19 Guidance to Help the Public Better Protect Themselves and Understand Their risk: Press Release Embargoed until: Thursday, August 11, 2022, 3: 00 PM ET.

WHO, 2024. WHO COVID-19 Dashboard. World Health Organization, Geneva. Available online. https://covid19.who.int/ (last cited: [06/011/2023]).