Dynamic SARS-CoV-2 surveillance model combining seroprevalence and wastewater concentrations for post-vaccine disease burden estimates

Check for updates

Rochelle H. Holm [1,10], Grzegorz A. Rempala [2,10], Boseung Choi[2,3,4,10], J. Michael Brick [5],
Alok R. Amraotkar[1], Rachel J. Keith[1], Eric C. Rouchka [6,7], Julia H. Chariker[6,7], Kenneth E. Palmer [8,9],
Ted Smith [1,11] & Aruni Bhatnagar[1,11] ✉

Abstract

Background Despite wide scale assessments, it remains unclear how large-scale severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) vaccination affected the wastewater concentration of the virus or the overall disease burden as measured by hospitalization rates.
Methods We used weekly SARS-CoV-2 wastewater concentration with a stratified random sampling of seroprevalence, and linked vaccination and hospitalization data, from April 2021–August 2021 in Jefferson County, Kentucky (USA). Our susceptible (S), vaccinated (V), variant-specific infected (I_1 and I_2), recovered (R), and seropositive (T) model (SVI_2RT) tracked prevalence longitudinally. This was related to wastewater concentration.
Results Here we show the 64% county vaccination rate translate into about a 61% decrease in SARS-CoV-2 incidence. The estimated effect of SARS-CoV-2 Delta variant emergence is a 24-fold increase of infection counts, which correspond to an over 9-fold increase in wastewater concentration. Hospitalization burden and wastewater concentration have the strongest correlation ($r = 0.95$) at 1 week lag.
Conclusions Our study underscores the importance of continuing environmental surveillance post-vaccine and provides a proof-of-concept for environmental epidemiology monitoring of infectious disease for future pandemic preparedness.

Plain language summary

It is unclear how large-scale COVID-19 vaccination impacts wastewater concentration or overall disease burden. Here, we developed a mathematical surveillance model that allows estimation of overall vaccine impact based on the amount of SARS-CoV-2 in wastewater, seroprevalence and the number of cases admitted to hospitals between April 2021–August 2021 in Jefferson County, Kentucky USA. We found that a 64% vaccination coverage correlated to a 61% decrease in COVID-19 cases. The emergence of the SARS-CoV-2 Delta variant during the time of the surveillance directly correlated with a sharp increase in infection incidence as well as viral counts in wastewater. The hospitalization burden was closely reflected by the viral count found in the wastewater, indicating that post-vaccine environmental surveillance can be an effective method of estimating changing disease prevalence in future pandemics.

[1]Christina Lee Brown Envirome Institute, School of Medicine, University of Louisville, Louisville, KY 40202, USA. [2]Division of Biostatistics, College of Public Health, The Ohio State University, Columbus, OH 43210, USA. [3]Division of Big Data Science, Korea University, Sejong, South Korea. [4]Biomedical Mathematics Group, Institute for Basic Science, Daejeon, South Korea. [5]Westat, Inc., Rockville, MD 20850, USA. [6]Department of Biochemistry and Molecular Genetics, School of Medicine, University of Louisville, Louisville, KY 40202, USA. [7]KY INBRE Bioinformatics Core, University of Louisville, Louisville, KY 40202, USA. [8]Center for Predictive Medicine for Biodefense and Emerging Infectious Diseases, University of Louisville, Louisville, KY 40202, USA. [9]Department of Pharmacology and Toxicology, School of Medicine, University of Louisville, Louisville, KY 40202, USA. [10]These authors contributed equally: Rochelle H. Holm, Grzegorz A. Rempala, Boseung Choi.[11]These authors jointly supervised this work: Ted Smith, Aruni Bhatnagar. ✉e-mail: aruni.bhatnagar@louisville.edu

In the wake of the COVID-19 pandemic, more reliable methods of measuring disease prevalence in communities are urgently needed, particularly methods that do not involve the expensive and cumbersome process of collecting individual level data. Completed development and validation of such methods are likely to be a centerpiece of preparedness for future pandemics. Wastewater concentration, when properly calibrated, can be a surrogate for estimates based on community prevalence of infection[1–3]. Moreover, wastewater-based epidemiology offers the opportunity of estimating community disease prevalence even with asymptomatic disease[2,3]. A handful of previous evaluations of the relationship between severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) wastewater concentration and the COVID-19 vaccine have relied almost exclusively on statistical models calibrated with case counts or other convenience sampling[4–8]. These data run the risk of biased underrepresentation of asymptomatic individuals who may not seek testing, or individuals testing in settings where reporting is low or not required[9]. Other mathematical models are based at a state or national spatiotemporal scale[10–13]. Hence, in this study we consider this question in the context of randomized seroprevalence surveillance, combining mechanistic and statistical frameworks to obtain more robust and realistic estimates of changes in disease prevalence.

We address the question of how changes during the Alpha and Delta variant waves of the pandemic affected wastewater concentrations by looking in detail at a small geographical area which other studies have not done previously. For our analysis, we use repeated cross-sectional community-wide stratified random sampling to measure SARS-CoV-2 nucleocapsid specific antibody-based seroprevalence in Jefferson County, Kentucky (USA), from April 2021 through August 2021 to determine post-vaccine community prevalence at a sub-county scale. We then relate this to a statistical linear model and the available sub-county weekly wastewater surveillance data which yield estimates of the explicit impact of vaccination and seroimmunity on a SARS-CoV-2 wastewater concentration estimate, while controlling for prevalence in different epidemic phases using a population level ecological model. The latter may be easily translated into other important public health indicators such as patterns of hospitalization. The ecological model, SVI_2RT, longitudinally monitors the proportions of individuals in various health stages. These include those who were susceptible (S), vaccinated (V), infected with non-Delta variant (I_1), infected with Delta variant (I_2), recovered (R), or seropositive (T). Here we show the 64% county vaccination rate translate into about a 61% decrease in SARS-CoV-2 incidence. The estimated effect of SARS-CoV-2 Delta variant emergence is a 24-fold increase of infection counts, which correspond to an over 9-fold increase in wastewater concentration. Hospitalization burden and wastewater concentration have the strongest correlation ($r = 0.95$) at 1 week lag.

Methods
Seroprevalence
Community-wide stratified simple random seroprevalence sampling (Supplementary Note 1 and Supplementary Table 1) was conducted in four waves from April to August 2021 in Jefferson County, Kentucky (USA) which is also the consolidated government for the city of Louisville[14]. Seroprevalence sampling was conducted both before and during vaccination, but this analysis only considers the period after COVID-19 vaccines were made widely available to the public ($N = 3303$). In some cases, due to the timing of sampling waves, respondents may have had only the first of a two-dose vaccine series. Serological positivity for nucleocapsid immunoglobulin G was used to identify participants with previous SARS-CoV-2 natural infection; vaccines used in the studied areas relied on SARS-CoV-2 viral spike protein and thus spike protein presence could be attributable to either natural infection or vaccination. Owing to elevated levels of vaccinated respondents in our study (~90%), we only included seroprevalence measured by response to IgG N1 antibodies[14,15]. The nucleocapsid (N1) IgG assay sensitivity was 65% and the specificity was 85%[14]. It was assumed over the study period vaccination induced antibodies did not decay below detection.

Concentration of SARS-CoV-2 and PMMoV in the wastewater
Wastewater samples were collected twice per week from five wastewater treatment plants ($N = 168$; Supplementary Note 2, Supplementary Fig. 1 and Supplementary Table 2) from April to August 2021. From an influent 24-h composite sampler, 125 ml of subsample was collected and analyzed for SARS-CoV-2 (N1) and pepper mild mottle virus (PMMoV). In a few cases due to an equipment malfunction, a grab sample was collected. The geographic area and population serviced by a wastewater treatment plant comprises a sewershed, the zone for which we consider in our model analysis across a range of population sizes, income levels and racial and ethnic diversity. Analysis used polyethylene glycol (PEG) precipitation with quantification in triplicate by reverse transcription polymerase chain reaction (RT-qPCR)[2]. Data for SARS-CoV-2 (N1) and PMMoV are reported as weekly average copies/ml of wastewater with a threshold value for SARS-CoV-2 (N1) assays of 7.5 copies/ml and for PMMoV of 143 copies/ml.

Administrative COVID-19 data
Administrative data on COVID-19 vaccination and infected individuals' hospitalization was provided by the Jefferson County health authority, Louisville Metro Department of Public Health and Wellness (LMPHW), under a Data Transfer Agreement. Vaccination data were geocoded to the urban sewersheds using ArcGIS Pro version 2.8.0 (Redlands, CA). Daily hospitalization data was only available aggregated at a county level.

Analytical model
The hybrid model for estimating the effect of vaccination and variants on longitudinal wastewater concentration was developed by combining a compartmental ecological model with a statistical linear model (Supplementary Note 1–4, Supplementary Information Tables 1–13 and Supplementary Information Figs. 1–9)[16]. The former was used to longitudinally estimate population prevalence from the observed cross-sectional rates of seropositivity. We assumed the overall vaccination pattern as reported by the county, with the overall adult vaccination rate reaching 64%[17] by the end of the study period. The hybrid model was used to relate the ecological model prevalence to the wastewater concentration. The ecological model, SVI_2RT, tracked longitudinally the proportions of individuals who were susceptible (S), vaccinated (V), infected with non-Delta variant (I_1), infected with Delta variant (I_2), recovered (R), or seropositive (T). We note that a version of this model that did not account for vaccination or variants was considered in our earlier work[2].

Upon estimating the parameters in the SVI_2RT model, we compared the model-calculated prevalence estimates for SARS-CoV-2 infections and vaccination levels with the wastewater concentration levels of SARS-CoV-2 (N1) and for that normalized by PMMoV[18]. We also separately calculated two prevalence estimates according to the Alpha and Delta variants. Bayesian linear regression was performed both on the county aggregated data and stratified by sub-county wastewater treatment plant zones (sewersheds). We used the broken stick regression model to separately compare the Alpha and Delta variant effects on the wastewater concentration with regression coefficients directly. To improve the regression model stability, we used weekly average prevalence rates from the SVI_2RT model as the explanatory variable, and weekly aggregated average wastewater concentrations as the single outcome variable. This temporal aggregation also allowed us to use a simple posterior-profile likelihood to estimate the average change point in the broken stick regression model (see, e.g., Schwartz et al.[19] for a similar approach for initial conditions imputation). We assigned non-informative priors to all regression parameters. Specifically, the non-informative Cauchy distribution was assigned to regression coefficients, and the non-informative gamma prior was assigned to the dispersion parameter in error term. The regression model with intercept is used where the intercept may be interpreted as background and calibration noise related to wastewater sampling. We could see temporal differences between the Alpha and Delta variant dominant dates, but this variability in time also considers that samples are weekly aggregated average wastewater concentrations. We did not include these variabilities of intervals in the

model as the magnitudes of the observed wastewater concentration and estimated prevalence in this interval are relatively small, and model changes do not alter the overall model fit.

The strong statistical significance of the regression model relating prevalence and wastewater concentration allowed for indirect estimation of the effect of population vaccination and variants. Under the assumption the relationship between the wastewater concentration and the prevalence is not confounded by the vaccination and variants, we used the original regression equation derived from the collected wastewater and seroprevalence data to estimate the wastewater concentration over time. To estimate the vaccination effect, we compared these concentrations with hypothetical ones obtained when the vaccination term was zeroed out in the SVI_2RT model. In a comparable manner, we estimated the effect of the introduction of the Delta variant. Finally, we performed the longitudinal, regression-based analysis relating the community hospitalization to observed wastewater concentrations. In the three analyses we quantified the effects by calculating the size of the effects relative to the factual (observed) states.

Wastewater samples were prepared for whole genome sequencing[20,21], and the proportion of observed SARS-CoV-2 variants was estimated for each sewershed based on variant dominance. Two variants were present in the study area during the study period: Alpha was dominant from April until July, while Delta was dominant from July until August (Supplementary Note 3, Supplementary Fig. 2 and Supplementary Table 3). To reflect the infections before and after the emergence of the Delta variant, we incorporated into our SVI_2RT model the two different infection compartments (I_1 and I_2) reflecting both the infection competition and temporal heterogeneity caused by the two different variants of the virus.

Ethics
For the seroprevalence and data provided by the LMPHW under a Data Transfer Agreement, the University of Louisville Institutional Review Board approved this as Human Subjects Research (IRB number: 20.0393). Participants in the seroprevalence study provided informed consent. For the wastewater data, the University of Louisville Institutional Review Board classified this as non-human subjects research (reference #: 717950).

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results
Wastewater regression
When examining the relationship between the prevalence estimated from the SVI_2RT model and the observed wastewater levels, the results of the Bayesian regression analysis demonstrate a close trend. This analysis considers both countywide aggregation and the five localized sewershed locations allowing finer geographic resolution. The trend is effectively summarized by the corresponding posterior regression line. To obtain reliable and stable longitudinal wastewater concentration readings, the concentration of SARS-CoV-2 (N1) was normalized by the PMMoV concentration to enhance accuracy and precision and minimize variance in assessing changes in the concentration of the virus over time.

To assess the impact of prevalence on observed wastewater concentration for the Alpha and Delta variants, we employed a regression model known as the broken stick regression to account for variant-specific patterns of virus shedding and infection rates. This model incorporates two regressors: one for the Alpha variant with an estimated prevalence prior to 5 June 2021, and another for the Delta variant with an estimated prevalence after 5 June 2021 (Supplementary Note 2 and 3 and Supplementary Figs. 2 and 3). Consequently, the model encompasses two different regression coefficients corresponding to the respective prevalence. The transition date was determined by the prevailing dominance of the Alpha and Delta variants as inferred from wastewater samples (Supplementary Note 3 and Supplementary Table 3). For the aggregate model, the estimated intercept is -4.222×10^{-4} (CI $= (-9.458 \times 10^{-4}, 7.921 \times 10^{-5})$) and the two

slopes are 0.815 (CI $= (-0.023, 1.717)$) and 0.385 (CI $= (0.318, 0.455)$) (Supplementary Note 4 and Supplementary Table 6). Overall, the regression model fits the data well ($R^2 = 0.90$).

Effect of vaccination on disease incidence and wastewater concentration
We first compared the estimated incidence of the SVI_2RT model under two different vaccination scenarios (factual 64% vaccination rate and counterfactual 0% vaccination rate) while adjusting for the Delta variant emergence (Fig. 1). The peak and the overall temporal dynamics are different under the two scenarios across each location, credible intervals for the incidence with and without vaccination are overlapping and indicate that the scenario curves could have statistically close values during certain times. To quantify these differences more precisely, we computed the location-specific vaccination effects as the ratio of the areas under the two scenario incidence curves. Specifically, we compared the area under the curve (corresponding to a relative cumulative incidence) for the with-vaccination scenario to that of the without-vaccination scenario. The value obtained for the aggregated data was 0.390, with the remaining sewershed specific effects being even stronger at 0.502, 0.393, and 0.479 for MSD1, MSD2, and MSD3–5, respectively. Based on converting these ratios to excess incidence, we estimate that without vaccination, the reported integrated incidence may have increased about 156.2% (CI $= (95.2\%, 175.7\%)$) above the observed level in Jefferson County (Fig. 1a) and about 99.4% (CI $= (94.2\%, 108.5\%)$), 154.5% (CI $= (3.2\%, 154.7\%)$), and 108.8% (CI $= (52.8\%, 109.2\%)$) in different sewershed areas (Fig. 1b–d, Supplementary Note 4 and Supplementary Table 8).

To obtain estimates of the effects of vaccination on the wastewater concentration of the virus, we developed a hybrid inferential model combining the wastewater regression equation with the SVI_2RT estimated prevalence, under two different vaccination scenarios (factual 64% rate and counterfactual 0% rate) (Fig. 2). The use of SVI_2RT (which accounts for the effect of different virulence of the two different SARS-CoV-2 variants) automatically adjusted our analysis for the Delta variant emergence. Because the estimated prevalence from the SVI_2RT model and the normalized wastewater concentration are highly correlated, the hybrid model is seen to fit the data well. As before, to quantify the location-specific vaccination effects, we calculated the location-specific ratios under two curves in an analogous way as when quantifying the vaccination effect on the disease incidence. The ratios of the areas under the two curves, under factual (vaccinated) and counterfactual (unvaccinated) scenarios, were computed. The Jefferson County (Fig. 2a) ratio was equal to 0.314, and the remaining sewershed location ratios (Fig. 2b–d) were equal to, respectively, 0.448, 0.330 and 0.375. The estimate of excess wastewater virus without vaccination is estimated as 218.9% (CI $= (193.5\%, 242.4\%)$), 123.1% (CI $= (105.0\%, 144.0\%)$), 202.8% (CI $= (192.8\%, 203.4\%)$), and 166.9%, (CI $= (146.6\%, 187.1\%)$) respectively (Supplementary Note 4 and Supplementary Table 8).

Effects of virus variant on disease incidence and wastewater concentration
Alpha was the dominant variant at the start of our study period on 30 March 2021 (Supplementary Note 3 and Supplementary Table 3). The Delta variant was first introduced into the two largest urban sewersheds as the dominant variant on 12 July 2021, appearing in the more rural sewersheds in the following 2-week period. More recently we have reported on the re-emergence of Delta in the MSD3 site during the Omicron wave[20], which indicates the persistence of specific variants in wastewater can be variable and are likely influenced by several factors, including the rates of incidence and vaccination.

In our analysis, we assumed a 50% higher infectivity of the SARS-CoV-2 Delta variant in comparison with its Alpha predecessor[22]. In the counterfactual model, where only the Alpha variant was present, the epidemic was seen to dissipate, indicating the effective reproduction number of less than one. This was in contrast with the factual, full SVI_2RT model fit (with both Alpha- and Delta- variants present), where the incidence was seen to rise rapidly. As in the previous section, to quantify the difference between
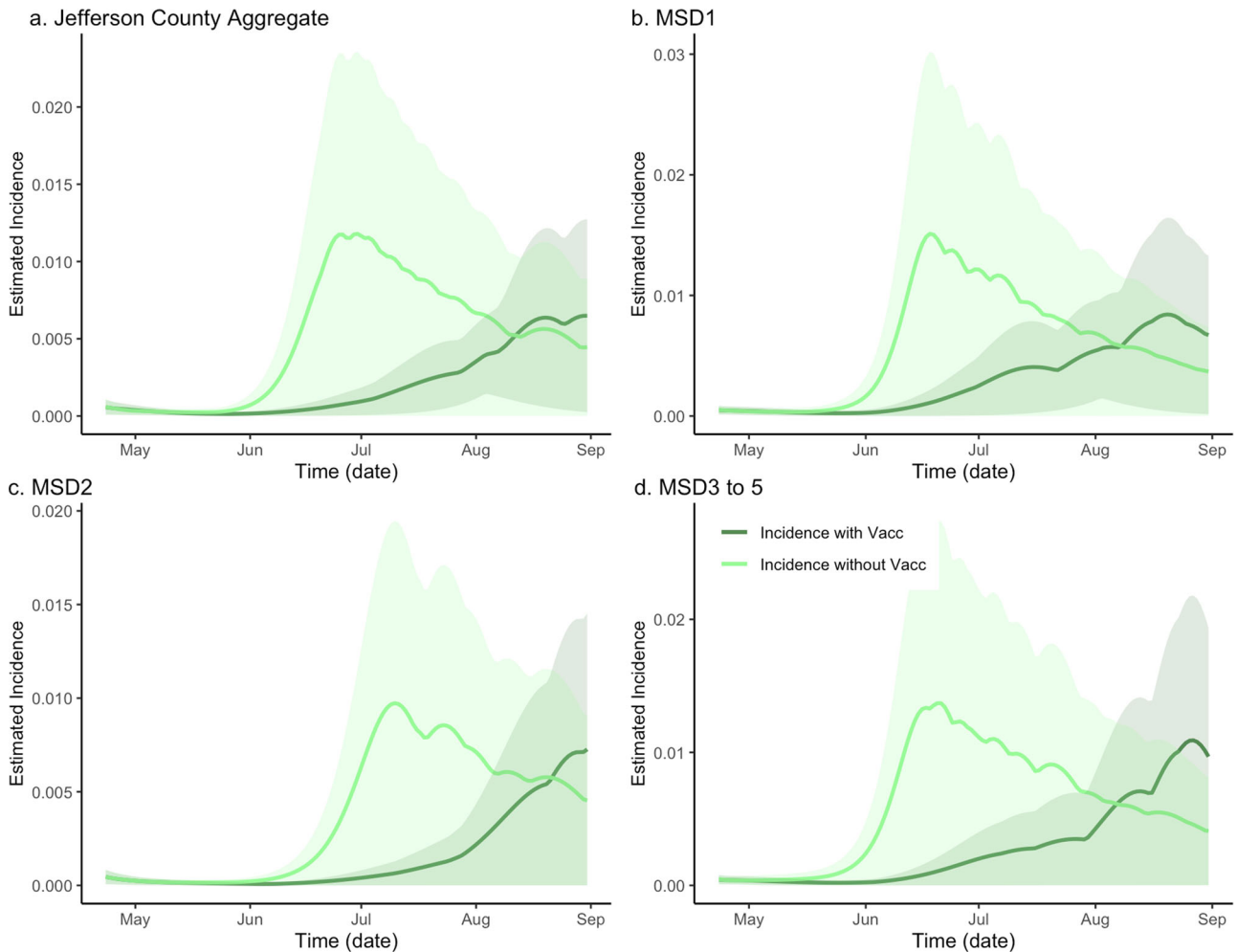
Fig. 1 | The estimated effect of vaccination on incidence in sewersheds of Jefferson County, KY (USA). The dark green line is the factual SVI_2RT model estimated incidence (with vaccination), and the light green line is the corresponding counterfactual estimated incidence with vaccination effect zeroed out. The shaded areas represent 95% credible intervals. The panels compare the vaccination effect in Jefferson County (**a**) as well as stratified by sewershed (**b–d**).

the two curves, which we interpreted as measuring the effect of introducing the Delta variant, we calculated the ratio of areas under the two curves, obtaining the values of 23.524, 31.103, 23.986, and 33.336 for the aggregate, MSD1, MSD2, and MSD3–5 regions respectively (Fig. 3). The estimate of the decrease in total incidence without the variant was found as 95.8% (CI = (95.7%, 95.9%)), 96.8% (CI = (95.5%, 96.8%)), 95.8% (CI = (2.7%, 96.0%)), and 97.0% (CI = (38.6%, 97.1%)), respectively (Supplementary Note 4 and Supplementary Table 8). Note that the two lower bounds of the ratio for the MSD2 and MSD3–5 areas were relatively small. This is because the estimated incidence from both variants had lower CI areas that were close to zero.

To identify the effect of the Delta variant emergence on the observed wastewater concentration, we again applied the hybrid model from the previous section. Genetic variants can have an impact on fecal shedding[23]. In the current analysis, the regression model was applied to predict the longitudinal wastewater concentrations from both factual (both variants present) and counterfactual prevalence data (no Delta variant) (Fig. 4). As with the analysis of the vaccination effects, here we also considered the ratios of areas under the corresponding curves as measures of Delta variant effects in specific locations. Based on the aggregated ratio values of 8.655, and on the location-specific ratio values 5.695, 9.675, and 8.530, the estimate of excess wastewater concentration due to Delta was found as 88.4% (CI = (87.7%, 88.7%)), 82.4% (CI = (81.4%, 84.0%)), 89.7% (CI = (88.5%, 90.8%)), and 88.3% (CI = (87.3%, 89.1%)) respectively (Supplementary Note 4 and Supplementary Table 8). By utilizing the fitted regression coefficients, we

can further examine the impact of the Alpha and Delta variants on wastewater concentrations. To facilitate a comparison, we employed standardized regression coefficients instead of the original scale. Because the range of the Alpha variant prevalence and the number of data points of the Alpha variant are smaller than that of the Delta variant, the slope coefficient of the Alpha variant is larger than that of the Delta variant. For the aggregated model (Fig. 4a), the standardized regression coefficient of the Alpha variant prevalence is 3.464×10^{-4} (CI = (4.460×10^{-7}, 6.946×10^{-4})) and the Delta variant prevalence is 1.992×10^{-3} (CI = (1.627×10^{-3}, 2.344×10^{-3})). Hence, the effect of the Delta variant was found to be 5.8 times greater than that of the Alpha variant. The fitted line of wastewater concentration exhibits a transition point, and the broken stick regression line aligns well with the data (R-square value 0.904). We can also see similar patterns in other sewershed locations (Fig. 4b–d). The standardized regression coefficients for each sewershed area are 5.053×10^{-4} (CI = (1.081×10^{-5}, 9.796×10^{-4})) and 1.880×10^{-3} (CI = (1.387×10^{-3}, 2.339×10^{-3})) for MSD1, 3.586×10^{-4} (CI = (-1.148×10^{-4}, 8.337×10^{-4})) and 3.395×10^{-3} (CI = (2.921×10^{-3}, 3.872×10^{-3})) for MSD2, and 2.518×10^{-4} (CI = (-4.963×10^{-5}, 5.616×10^{-4})) and 1.609×10^{-3} (CI = (1.291×10^{-3}, 1.910×10^{-3})) for MSD3–5 respectively.

Insights into hospitalization rates based on wastewater concentration

Hospitalization estimates under both vaccinated (64% vaccination rate[17]) and unvaccinated (0% vaccination rate) scenarios were obtained by
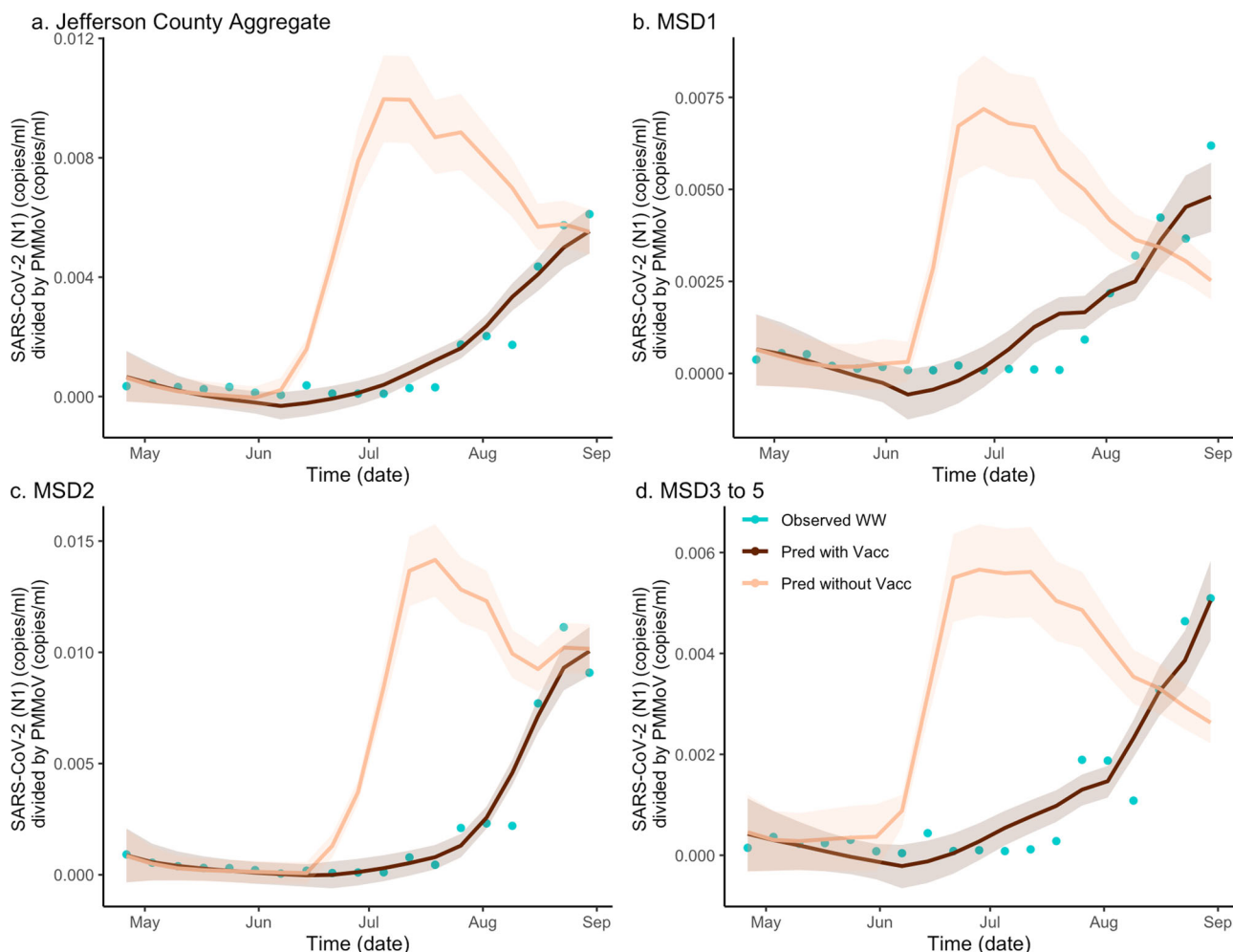
Fig. 2 | The estimated effect of vaccination on SARS-CoV-2 (N1) wastewater concentration normalized by pepper mild mottle virus in sewersheds of Jefferson County, KY (USA). The dark brown line is the regression-based fit to the wastewater concentration data and the light brown line is the prediction of wastewater concentration using synthetic prevalence from SVI_2RT model with vaccination effect zeroed out. The shaded areas represent 95% credible intervals. The blue dots are observed weekly average wastewater concentrations. The panels compare the vaccination effect on wastewater concentration for Jefferson County (**a**) as well as stratified by sewershed (**b**–**d**).

applying a hierarchical regression model where we first regressed wastewater concentration on the SVI_2RT model prevalence and then regressed hospitalization counts on the wastewater concentrations (Fig. 5). As hospitalization is likely to occur sometime after symptom onset, we considered a range of no lag to a 5-week lag period. A 1-week lagged-regression model was the best fit where the length of the lag time was based on the overall model fit criteria. The fitted intercept and slope coefficients were 1.222×10^{-4} (std = 3.345×10^{-5}) and 0.181 (std = 0.0150) for vaccinated and unvaccinated scenarios respectively (R-square of 0.895) (Fig. 5). The maximum number of observed daily average hospitalizations under the vaccination scenario was 110.4 per weekly average (actual 122.0 in daily) at the end of August. However, without vaccination, the maximum predicted number of weekly average hospitalizations increased to 192.1. The ratios between the areas under the prediction curves with and without vaccination were 0.318, indicating a 214% (CI = (192%, 250%)) increase in the number of hospitalizations when no vaccine would be present. In a comparable way, we obtained the hospitalization estimate without the Delta variant. The ratio of the areas under the two graphs (with and without the Delta variant) is 3.037, indicating a 67% (CI = (53.5%, 89.4%)) decrease in the hospitalization rate. Furthermore, we conducted a regression analysis linking the hospitalization rate to wastewater concentration. The resulting slope coefficient was 0.1762 (sd = 0.0119), and an R-square value was 0.9241 (Supplementary Note 4 and Supplementary Table 12). Notably, the

predictions from this simple regression model outperformed those of the hierarchical regression model discussed earlier. This suggests that wastewater concentration can serve as a robust predictor for forecasting hospitalization rates.

Discussion

The results of our large study ($N = 3303$) show the importance of post-vaccine environmental surveillance for the prevalence of the virus in an urban area. A major advantage of this approach is that it decreases bias implicit in publicly-available clinical case data by assessing community prevalence using antibody positivity with four waves of sequential stratified random sampling data. Although our work was localized to Jefferson County, where contemporaneous randomized sampling and wastewater concentration were available, it should be emphasized the model described here may be readily applicable to other locations worldwide. In addition to SARS-CoV-2, the model may be valid for other infectious diseases. Furthermore, despite running our model with data adjusted using both the SARS-CoV-2 (N1) and PMMoV concentrations, we observed that the PMMoV adjustment reduced uncertainty. Indeed, estimation of the effect of vaccination brings the related issue of refined localized model application such as high levels of tourism that may affect community vaccination levels and related observed wastewater concentrations[8]. Here we have presented real world evidence that, in fact, small area wastewater surveillance could be
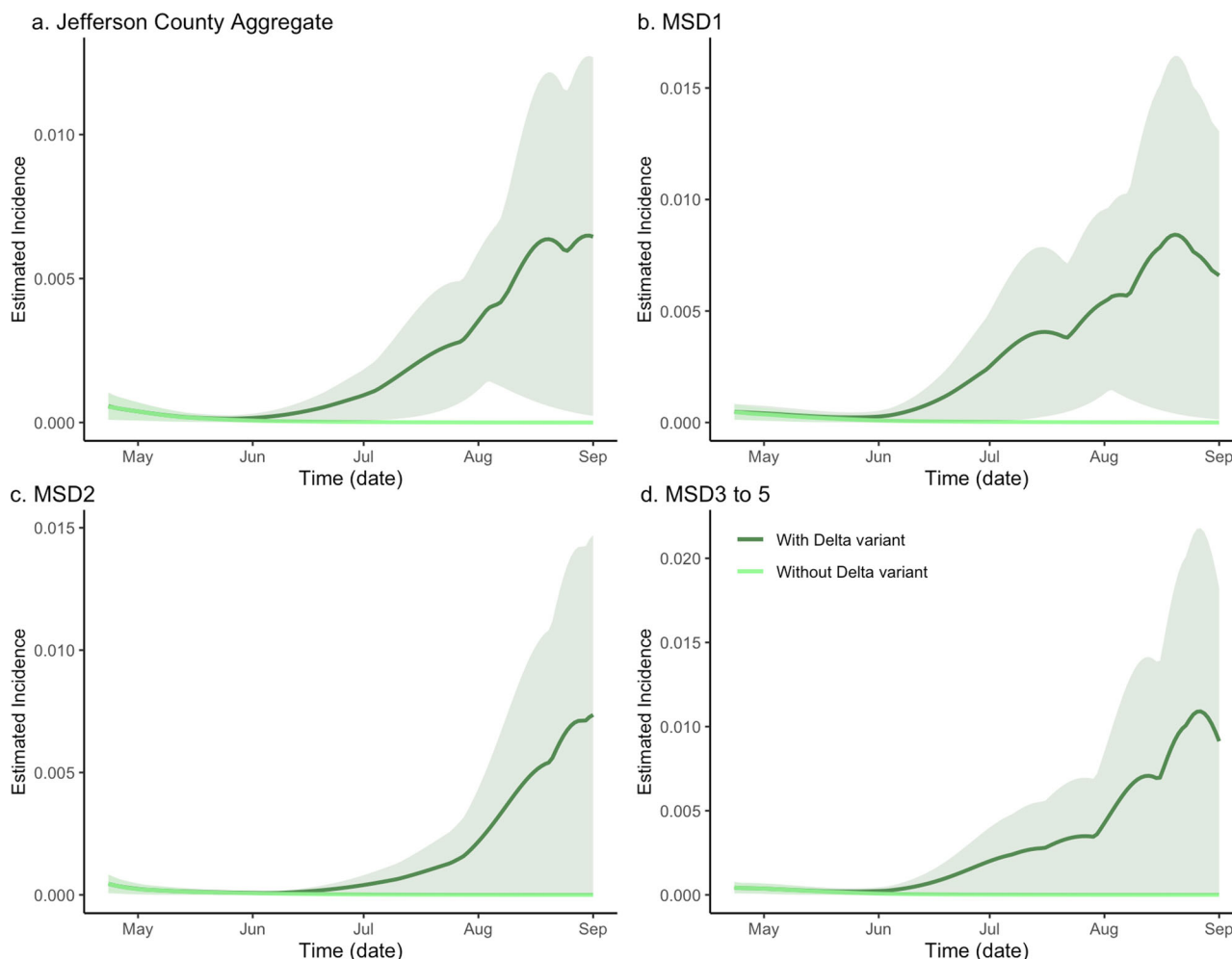
Fig. 3 | The model-based analysis of the Delta variant effect on SARS-CoV-2 incidence rate estimates in sewersheds of Jefferson County, KY (USA). The dark green line is the estimated factual full model incidence (both Alpha and Delta variants present), and the light green line is the counterfactual incidence estimated from the model with no Delta variant. The shaded areas represent 95% credible intervals. The panels compare the incidence rate in Jefferson County (**a**) as well as stratified by sewershed (**b–d**).

used to estimate both—the effects of disease evolution as well as a community intervention, like a vaccination campaign.

It is widely recognized that even though vaccine distribution was more proportional to wealth than need, COVID-19 vaccination was highly effective[24–27]. Therefore, where accessible, the impact of vaccination on community-wide prevalence of infection was readily apparent. However, for other vaccine-preventable disease, there is an urgent need for increased reliance on wastewater as a proxy for community disease impact being built from actual community level data over time, as the estimates by different methods can vary. For instance, when 90% of the student population of a college campus was vaccinated, SARS-CoV-2 in wastewater decreased[4]; but that university campus population generalization was not replicated in our community-wide survey over a longer period. In contrast to our findings, Nourbakhsh et al.[28] found dissimilar trajectories from community clinical and wastewater ratios once vaccination was introduced. We suspect this difference is explained by the bias of relying on clinical data and home testing kits which became more widely available during our studied period than earlier in the pandemic and with no requirement, or in some cases option, for reporting. Whereas the Nourbakhsh et al.[28] study included only publicly reported case data, the randomized selection of community participants in our study population[14] was a comparatively less biased data source for a post-vaccine study period. Our recent work has shown that even though we cannot rule out bias due to self-selection for testing, the

randomized sampling approach provides better estimates of disease prevalence than administratively reported data[14].

Our model is comparable to that used by Jiang et al.[29] in that our analysis also provided estimates of prevalence; however, our estimates are based on a statistically random sample (not a clinical sample) and our regression model has a simple and explicit formula relating prevalence to observed wastewater levels of the virus. Our model also confirms the findings of Hegazy et al.[6] implying the Delta variant emergence strengthened the relationship between wastewater and disease burden. Hence, our analysis provides a further proof-of-concept that our wastewater regression model could be used (after proper calibration) with other similar data to provide surrogate measures of SARS-CoV-2 prevalence in the community without the necessity for individual testing. The regression prediction correlates well with the estimated prevalence with a correlation coefficient of 0.858 (CI = (0.502, 0.975). The hospital burden findings of Wang et al.[30] also compares well to our work; our results showed access to a voluntary community vaccine that reached a coverage level of 64% of the adult population decreased community hospitalizations by ~214%.

Yaniv et al.[5] described the introduction of variant signal in wastewater and noted how vaccination rates and a second booster helped to control the Alpha variant, while an increase in a third booster was found to lead to a decline in Delta. When vaccination levels increase to higher coverage, hospitalizations may decline, even though the levels in wastewater remain
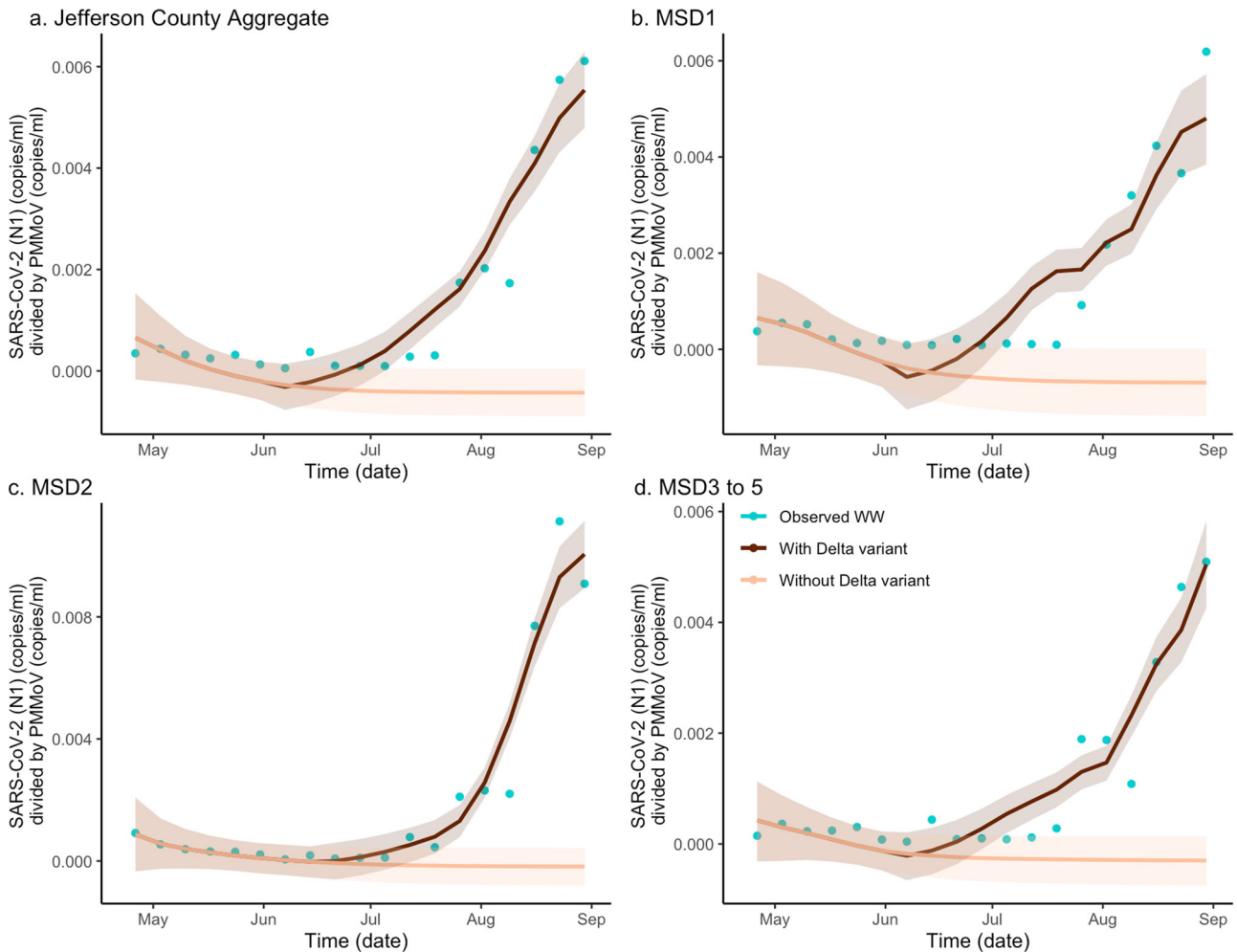
Fig. 4 | The estimated effect of Alpha and Delta variant on SARS-CoV-2 (N1) wastewater concentration normalized by pepper mild mottle virus in sewersheds of Jefferson County, KY (USA). The dark brown line is the regression-based fit to the wastewater concentration with the Alpha and Delta variant and the light brown line is the prediction of wastewater concentration using synthetic prevalence from the SVI_2RT model with the Alpha variant only. The shaded areas represent 95% credible intervals. The blue dots are observed weekly average wastewater concentration. The panels compare the variant effect on wastewater concentration for Jefferson County (**a**) as well as stratified by sewershed (**b–d**).

high[7]. Pandemic preparedness and associated public health response would benefit from new methods less dependent on continuous individual clinical testing.

Our study used five sub-county locations based on the existing wastewater infrastructure allowing observation of a small geographical area but also the aggregation of data for a countywide picture. We found that the antibody positivity varied by the sewershed. The antibody-positive individuals were lowest in sewershed MSD1 and highest in sewershed MSD3–5 (9% for aggregate, 8% for MSD1, 9% for MSD2, and 10% for MSD3–5), indicating that previous infection may have been higher in the less dense portions of the county as compared with the urban core. Nonetheless, there are many other factors differentiating these sewershed areas that could have produced these differences. These include population sizes and demographics, or presence of stormwater or industrial discharge being combined with household sewer water. Regardless, the differences between MSD1 to 5 provide evidence of the benefit of observing results at both an aggregated and a smaller sub-county level.

For replication of our current hybrid SARS-CoV-2 model, wastewater sampling, stratified random sampling of seroprevalence, and linked vaccination data are required; the model is flexible enough to allow additional variant-specific variables. The promise of this model is that with known wastewater levels of the virus, we can predict the effect of vaccination to enable fine-tuned, and milestone-driven, public health response. The results

obtained from our model show unequivocally that the COVID-19 pandemic would have been larger and spread earlier without vaccine access. These findings provide further positive evidence for the significant role of vaccines in public health, a valuable lesson for the pandemic preparedness.

Despite its many strengths, our study has some limitations. The proportion of vaccinated respondents in the seroprevalence study was larger than the greater community (~90% vs. 64%). Vaccine information was self-reported, and we made a simplifying assumption that the magnitude of vaccine leakage effect is negligeable[31] when comparing to other effects. Natural infection of a combined vaccinated and unvaccinated population (and in the absence of another way to verify vaccination) was based on antibody titers of IgG N1, an assay that has 65% sensitivity and 85% specificity[14] a priori, with inevitable under-estimation of infection prevalence. While our serosurvey only captured adults, wastewater assay included minors. COVID-19 infected individuals can, in rare instances, shed fecal SARS-CoV-2 up to 7 months post diagnosis[32]; viral shedding of SARS-CoV-2 can vary in relation to vaccination status and variant[33,34] and thus was not included in our model. One of the major advances of this paper is the presentation of a relatively simple and flexible analytical model capable of using wastewater concentration to evaluate the effect of vaccination and variants on prevalence and hospitalization rates. Our simulations suggest we could use as little as 50% of data to retain the calibration conclusions (Supplementary Note 4 and Supplementary Table 13). This issue is worth
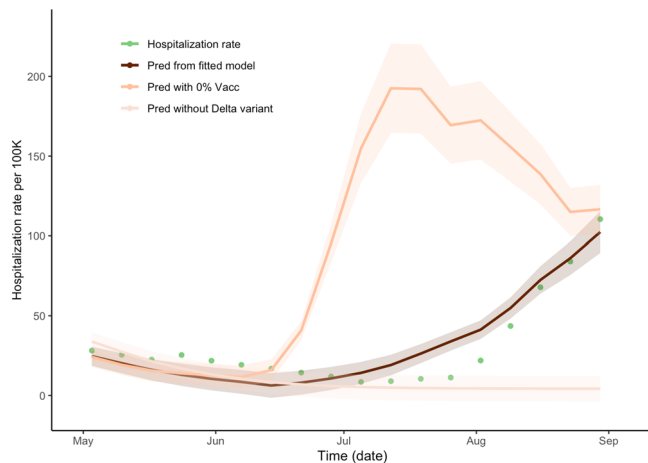
Fig. 5 | Time lag-dependent analysis of the relationship between hospitalization rate and wastewater concentration, Jefferson County, KY (USA). Predictions and 95% confidence intervals of hospitalization rate regressed on week-lagged variables of the weekly average of wastewater concentration according to the changes of the vaccination proportion of the community. The dark line represents the prediction using the observed wastewater concentrations with 64% of community vaccination. The lighter line represents the prediction using the wastewater concentrations obtained from the model under zero community vaccination. The lightest line represents the prediction under the counterfactual modified SVI_2RT model with the Delta-infected model compartment zeroed-out (no Delta variant present). The green dots represent the weekly average of the observed hospitalization rate. The ratios of the areas between the prediction from the fitted model and of no vaccination are 0.318, and in the absence of the Delta variant is 3.037, respectively.

further study outside the present work. Finally, the model we utilized assumed perfect protection for individuals infected with the Alpha variant against the Delta variant, as well as the insignificant seropositivity waning. While these assumptions may not be entirely valid, they appear reasonable[35] and are unlikely to have an impact on our conclusions.

Conclusion

Overall, our work suggests that under certain conditions, it is possible to use wastewater-based epidemiology to assess both immunity acquisition in the community due to natural recovery and vaccination as well as the effect of variant emergence and associated immune evasion to the available vaccines. The effect of vaccination on wastewater concentration as well as on community incidence of SARS-CoV-2 was substantial in Jefferson County. According to our analysis, without vaccination, one would expect about 156% of excess infections over the period of study, which corresponds to a 219% of excess wastewater concentration. The effect of the Delta variant was similarly substantial. We estimated, over the study period in Jefferson County, without Delta the amount of overall infection would decrease on average by 96% which corresponds to 88% decrease in the wastewater SARS-CoV-2 (N1) normalized by PMMoV concentration ratio. The correspondence between wastewater concentration and the number of hospitalizations was found to be strongest with the time lag for about 7 days and correlation = 0.95. Based on the regression model we estimated the effects of vaccination and variants on hospitalization rate. According to the model, without vaccination one would expect about 214% increase and without variants about 67% decrease in hospitalization rate. Using the fitted regression model for hospitalization, the predictions of hospitalization rates are at 50, 100, and 150 per 100 K when SARS-CoV-2 (N1) normalized by PMMoV ratios are 0.0021, 0.0050, and 0.0077, respectively.

Our large, randomized, serosurvey suggests using the mechanistic, population level, vaccination model (SVI_2RT) coupled with longitudinal wastewater sampling reliably estimated the effect of vaccination on the prevalence rate in the community over the period of several months during the second and third wave of COVID-19 pandemic, in the absence of clinical

data. Ours is the first study to look at a specific small area. The model can also be used to estimate the effects of vaccination and variant emergence on the hospitalization rate and on peak hospital beds utilization, estimates critical for adequate preparedness for the next pandemic, should it arise.

References

1. Li, X. et al. 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 (2023).
2. Smith, T. et al. Quantifying the relationship between sub-population wastewater samples and community-wide SARS-CoV-2 seroprevalence. *Sci. Total Environ.* **853**, 158567 (2022).
3. Berchenko, Y. et al. Estimation of polio infection prevalence from environmental surveillance data. *Sci. Transl. Med.* **9**, eaaf6786 (2017).
4. Bivins, A. & Bibby, K. Wastewater surveillance during mass COVID-19 vaccination on a college campus. *Environ. Sci. Tech. Let.* **8**, 792–798 (2021).
5. Yaniv, K., Ozer, E., Lewis, Y. & Kushmaro, A. RT-qPCR assays for SARS-CoV-2 variants of concern in wastewater reveals compromised vaccination-induced immunity. *Water Res.* **207**, 117808 (2021).
6. Hegazy, N. et al. Understanding the dynamic relation between wastewater SARS-CoV-2 signal and clinical metrics throughout the pandemic. *Sci. Total Environ.* **20**, 158458 (2022). 853.
7. Nattino, G. et al. Association between SARS-CoV-2 viral load in wastewater and reported cases, hospitalizations, and vaccinations in Milan, March 2020 to November 2021. *JAMA* **327**, 1922–1924 (2022).
8. Rainey, A. L. et al. Wastewater surveillance for SARS-CoV-2 in a small coastal community: effects of tourism on viral presence and variant identification among low prevalence populations. *Environ. Res.* **208**, 112496 (2022).

9. Bilal, U., Tabb, L. P., Barber, S. & Diez Roux, A. V. Spatial inequities in COVID-19 testing, positivity, confirmed cases, and mortality in 3 US cities: an ecological study. *Ann. Intern. Med.* **174**, 936–944 (2021).

10. Amman, F. et al. Viral variant-resolved wastewater surveillance of SARS-CoV-2 at national scale. *Nat. Biotechnol.* **40**, 1814–1822 (2022).

11. Dyson, L. et al. Possible future waves of SARS-CoV-2 infection generated by variants of concern with a range of characteristics. *Nat. Commun.* **12**, 5730 (2021).

12. Keeling, M. J. et al. Comparison of the 2021 COVID-19 roadmap projections against public health data in England. *Nat. Commun.* **13**, 4924 (2022).

13. KhudaBukhsh, W. R. et al. Projecting COVID-19 cases and hospital burden in Ohio. *J. Theor. Biol.* **561**, 111404 (2023).

14. Keith, R. J. et al. Stratified simple random sampling versus volunteer community-wide sampling for estimates of COVID-19 prevalence. *Am. J. Public Health.* https://doi.org/10.2105/AJPH.2023.307303 (2023).

15. Hamorsky, K. T. et al. Serological assessment of SARS-CoV-2 infection during the first wave of the pandemic in Louisville Kentucky. *Sci. Rep.* **11**, 18285 (2021).

16. Choi, B. & Rempala, G. Statistical method for identifying relationship between wastewater concentration and community-wide prevalence of COVID-19 (V1.0.0). Zenodo. https://zenodo.org/doi/10.5281/zenodo.10685975 (2024).

17. CDC. COVID-19 vaccinations in the United States, County. https://data.cdc.gov/d/8xkx-amqh/visualization (2022).

18. Maal-Bared, R. et al. Does normalization of SARS-CoV-2 concentrations by pepper mild mottle virus improve correlations and lead time between wastewater surveillance and clinical data in Alberta (Canada): comparing twelve SARS-CoV-2 normalization approaches. *Sci. Total Environ.* **856**, 158964 (2023).

19. Schwartz, E. J., Choi, B. & Rempala, G. A. Estimating epidemic parameters: application to H1N1 pandemic data. *Math. Biosci.* **270**, 198–203 (2015).

20. Westcott, C. E. et al. The detection of periodic reemergence events of SARS-CoV-2 Delta strain in communities dominated by Omicron. *Pathogens* **11**, 1249 (2022).

21. Rouchka, E. C. et al. The rapid assessment of aggregated wastewater samples for genomic surveillance of SARS-CoV-2 on a city-wide scale. *Pathogens* **10**, 1271 (2021).

22. Kläser, K. et al. COVID-19 due to the B.1.617.2 (Delta) variant compared to B.1.1.7 (Alpha) variant of SARS-CoV-2: a prospective observational cohort study. *Sci. Rep.* **12**, 10904 (2022).

23. Prasek, S. M. et al. Variant-specific SARS-CoV-2 shedding rates in wastewater. *Sci. Total Environ.* **857**, 159165 (2023).

24. Moore, S. et al. Retrospectively modeling the effects of increased global vaccine sharing on the COVID-19 pandemic. *Nat. Med.* **28**, 2416–2423 (2022).

25. Watson, O. J. et al. Global impact of the first year of COVID-19 vaccination: a mathematical modelling study. *Lancet Infect. Dis.* **22**, 1293–1302 (2022).

26. Nordström, P., Ballin, M. & Nordström, A. Risk of infection, hospitalization, and death up to 9 months after a second dose of COVID-19 vaccine: a retrospective, total population cohort study in Sweden. *Lancet* **399**, 814–823 (2022).

27. Gozzi, N. et al. Estimating the impact of COVID-19 vaccine inequities: a modeling study. *Nat. Commun.* **14**, 3272 (2023).

28. Nourbakhsh, S. et al. A wastewater-based epidemic model for SARS-CoV-2 with application to three Canadian cities. *Epidemics* **39**, 100560 (2022).

29. Jiang, G. et al. Artificial neural network-based estimation of COVID-19 case numbers and effective reproduction rate using wastewater-based epidemiology. *Water Res.* **218**, 118451 (2022).

30. Wang, H. et al. The amount of SARS-CoV-2 RNA in wastewater relates to the development of the pandemic and its burden on the health system. *iScience* **25**, 105000 (2022).

31. McKeigue, P. M. et al. Vaccine efficacy against severe COVID-19 in relation to delta variant (B. 1.617. 2) and time since second dose in patients in Scotland (REACT-SCOT): a case-control study. *Lancet Resp. Med.* **10**, 566–572 (2022).

32. Natarajan, A. et al. Gastrointestinal symptoms and fecal shedding of SARS-CoV-2 RNA suggest prolonged gastrointestinal infection. *Med* **3**, 371–87.e9 (2022).

33. Arts, P. J. et al. Longitudinal and quantitative fecal shedding dynamics of SARS-CoV-2, pepper mild mottle virus, and CrAssphage. *Msphere* e00132-23. https://doi.org/10.1128/msphere.00132-23 (2023).

34. Garcia-Knight, M. et al. Infectious viral shedding of SARS-CoV-2 Delta following vaccination: a longitudinal cohort study. *PLoS Pathog.* **18**, e1010802 (2022).

35. Stein, C. et al. Past SARS-CoV-2 infection protection against re-infection: a systematic review and meta-analysis. *Lancet* **401**, 833–842 (2023).

Acknowledgements

Author contributions

A.B., T.S., K.E.P., and R.J.K conceived and developed the idea for the study. K.E.P. supervised the serology assays. T.S. and R.H.H. supervised the wastewater assays. G.A.R. and B.C. performed the model-based data analyses. R.H.H. and G.A.R. wrote the first draft of the manuscript with all authors contributing to the interpretation of data and critical revision of the article. J.M.B. accessed and verified the data reported in the study. R.H.H., G.A.R., B.C., J.M.B., A.R.A., R.J.K., E.C.R., J.H.C., K.E.P., T.S., and A.B. had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Competing interests

The authors declare the following competing interests: R.J.K. declares participation on a Data Safety Monitoring Board or Advisory Board—Primary Health started March 2022 and runs through current. Primary health has a COVID testing platform for scheduling and reporting results. No other authors declare competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43856-024-00494-y.

Correspondence and requests for materials should be addressed to Aruni Bhatnagar.

Peer review information *Communications Medicine* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints