Research Article

SARS-CoV-2 wastewater concentration and linked longitudinal seroprevalence: a spatial analysis of strain mutation, post-COVID-19 vaccination effect, and hospitalization burden forecasting

Rochelle H Holm*, Grzegorz A Rempala*, Boseung Choi*, J Michael Brick, Alok R Amraotkar, Rachel J Keith, Eric C Rouchka, Julia H Chariker, Kenneth E Palmer, Ted Smith†, Aruni Bhatnagar†

* - Joint first authors
† - Joint senior authors

Christina Lee Brown Envirome Institute, School of Medicine, University of Louisville, Louisville, KY 40202, USA (R H Holm PhD, A R Amraotkar MD, R J Keith PhD, T Smith PhD, A Bhatnagar PhD); Division of Biostatistics, College of Public Health, The Ohio State University, Columbus, OH 43210, USA (G A Rempala DSc); Division of Big Data Science, Korea University, Sejong, South Korea (B Choi PhD); Biomedical Mathematics Group, Institute for Basic Science, Daejeon, South Korea (B Choi PhD); Westat, Inc., Rockville, MD 20850, USA (J M Brick PhD); Department of Biochemistry and Molecular Genetics, University of Louisville, Louisville, KY 40292, USA (E C Rouchka DSc, J HChariker PhD); KY INBRE Bioinformatics Core, University of Louisville, Louisville, KY 40202, USA (E C Rouchka DSc, J H Chariker PhD); Center for Predictive Medicine for Biodefense and Emerging Infectious Diseases, University of Louisville, Louisville, KY 40202, USA (K E Palmer PhD); Department of Pharmacology and Toxicology, School of Medicine, University of Louisville, Louisville, KY 40202, USA (K E Palmer PhD)

Correspondence to: Prof Aruni Bhatnagar, Christina Lee Brown Envirome Institute, School of Medicine, University of Louisville, Louisville, KY 40202, USA; aruni.bhatnagar@louisville.edu; 1.502.852.5724

Summary

Background

Since early in the COVID-19 pandemic, SARS-CoV-2 wastewater concentration has been measured as a surrogate for community prevalence. However, our knowledge remains limited regarding wastewater concentration and effects of the COVID-19 vaccination on 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 spatially linked vaccination and hospitalization data, from April to August 2021. 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 for a spatial analysis of strain mutation, vaccination effect, and overall hospitalization burden.

Findings

We found strong linear association between wastewater concentration and estimated community prevalence (r=0·916). Based on the corresponding regression model, the 64% county vaccination rate translated into about 57% decrease in SARS-CoV-2 incidence. During the study period, the estimated effect of SARS-CoV-2 Delta variant emergence was seen as an over 7-fold increase of infection counts, which corresponded to over 12-fold increase in wastewater concentration. Hospitalization burden and wastewater concentration had the strongest correlation (r=0·963) at 1 week lag time. We estimated the community vaccination campaign resulted in about 63% reduction in the number of daily admissions over the study period. This protective effect was counteracted by the emergence of SARS-CoV-2 Delta strain mutation.

Interpretation

Wastewater samples can be used to estimate the effects of vaccination and hospitalization burden. Our study underscores the importance of continued environmental surveillance post-vaccine and provides a proof of concept for environmental epidemiology monitoring.

Funding

Centers for Disease Control and Prevention (75D30121C10273), Louisville Metro Government, James Graham Brown Foundation, Owsley Brown II Family Foundation, Welch Family, Jewish Heritage Fund for Excellence, the National Institutes of Health, (P20GM103436), the Rockefeller Foundation, and the National Sciences Foundation (DMS-2027001).

Keywords: COVID-19; environmental surveillance; epidemiology; sewer; vaccine; wastewater-based epidemiology

Research in context

Evidence before this study: We searched Web of Science and PubMed for all available articles until August 24, 2022, using the search terms [“seroprevalence” or “antibody”] AND [“wastewater”] AND [“vaccination”]. We examined only English literature. We identified 59 studies. None of these studies considered community level randomized antibody testing paired with vaccination and SARS-CoV-2 wastewater concentration data. Where wastewater and vaccination status have been historically linked is with *Poliomyelitis*, in the known spatial scale of vaccination rates and using wastewater surveillance to confirm presence/absence of community infection. Few studies considered hepatitis A antibodies in workers exposed to sewage to guide vaccination campaigns. We also found some non-human subject research. To our knowledge, there is no real-world setting SARS-CoV-2 study where quantified wastewater concentration is linked to a large longitudinal stratified randomized seroprevalence sampling at a sub-population scale and the population has voluntary access to a vaccination reducing hospitalization burden.

Added value of this study: To our knowledge, this study provides the first analysis of SARS-CoV-2 wastewater concentration as the basis for estimating subpopulation vaccination and virus mutation effects, and hospitalization burden in any country. We used actual seroprevalence data from a large US urban area that was rigorously collected through statistical sampling, to obtain a longitudinal estimate of disease prevalence. We then used a statistical model relating prevalence to wastewater concentration for a spatial analysis of vaccination, virus mutation effects, and for forecasting hospitalization burden. The methodology developed in the current paper has a potential to improve both the effectiveness of monitoring and the predictive accuracy of wastewater-based surveillance systems.

Implications of all the available evidence: Our results show the potential of sustained environmental surveillance post-vaccine in urban areas and on removing bias in population-level estimates of prevalence of SARS-CoV-2 due to over-reliance on reported clinical testing data. The methodology presented here provides further proof wastewater monitoring can be successfully used as a tool for estimating both the community impact of changing disease patterns and various interventions over time. These findings have implications beyond current SARS-CoV-2 pandemic since our proposed approach is quite general and can be applied to other vaccine preventable diseases affecting human health in the absence of clinical testing data.

1. Introduction

There is an increasing realization that the current methods of disease monitoring based on individual testing may be insufficient to effectively combat the new, possibly much more infectious, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants. This leaves public health researchers and policy makers in search for more reliable methods of measuring SARS-CoV-2 prevalence in communities and especially those not involving the (expensive) process of collecting individual level data. Wastewater concentration, when properly calibrated, can be a surrogate for the virus community prevalence analysis.[1,2] Wastewater epidemiology promises an exciting opportunity to estimate community disease prevalence even with asymptomatic, vaccine preventable, disease.[2,3] However, the handful of recent studies considering a relationship between SARS-CoV-2 wastewater concentration and the COVID-19 vaccine have relied almost exclusively on statistical models calibrated with publicly available COVID-19 clinical case data.[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] In this study we consider this question in the context of randomized seroprevalence surveillance, combining mechanistic and statistical frameworks to obtain a more robust and realistic answer.

We used repeated cross-sectional community-wide stratified randomized sampling to measure SARS-CoV-2 nucleocapsid (N) specific IgG antibody-based seroprevalence in Jefferson County, Kentucky (USA), from April through August 2021 to determine post-vaccine community prevalence at a sub-population scale. We then related this to a statistical linear model and the available sub-population weekly wastewater surveillance data which thus yielded an explicit impact of vaccination and seroimmunity on SARS-CoV-2 wastewater concentration estimate, while controlling for prevalence in different epidemic phases. The latter may be easily translated into other important public health indicators such as the patterns of hospitalization.

2. Methods

2.1 Seroprevalence

Community-wide stratified randomized seroprevalence sampling 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. Seroprevalence sampling was both before and during vaccination, but this analysis only considers the period after COVID-19 vaccines were made widely available to the public (N=3436). An address-based sampling frame was used to build four geographic zones. Invitations (~30,000 per wave) were mailed to sampled households and one random adult was selected to join the study. Participants completed an online consent form and survey and scheduled an in-person appointment for testing at a mobile site. In some cases, due to the timing of sampling waves, respondents may have had only the first of a two-dose vaccine series. Owing to elevated levels of vaccinated respondents in our study (~90%), seroprevalence was measured by response to IgG N antibodies.[10] It was assumed over the study period vaccination induced antibodies do not decay below detection. The nucleocapsid (N) IgG test sensitivity was 65% and the specificity was 85%. The seroprevalence sampling by geographical zones are described in more detail in the Supplemental Material section S1.

2.2 Wastewater SARS-CoV-2 concentration

Wastewater samples were collected twice per week from five wastewater treatment plants (N=520; Supplemental Material section S2) from April to August 2021. From an influent 24-hour composite sampler, 125 ml of subsample was collected and analyzed for SARS-CoV-2. 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.[2] Analysis used polyethylene glycol (PEG) precipitation with quantification in triplicate by reverse transcription polymerase chain reaction (RT-qPCR).[11] Data for SARS-CoV-2 (N1) are reported as weekly average copies/ml of wastewater with a threshold value for N1 assays of 7·5 copies/ml.

2.3 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 sewersheds using ArcGIS Pro version 2·8·0 (Redlands, CA). Daily hospitalization data was only available aggregated at a county level.

2.4 Analytical model

The hybrid model for estimating the effect of vaccination and strain mutation on longitudinal wastewater concentration was developed by combining a compartmental ecological model with a statistical linear model. 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%[12] by the end of the study period. The hybrid model was used to relate the ecological model prevalence to the wastewater data. The ecological model, SVI_2RT, tracked longitudinally the proportions of individuals who were susceptible (S), vaccinated (V), infected with non-Delta (I_1), and Delta variant (I_2), recovered (R), or seropositive (T). The model is described in more detail in the Supplemental Material section S3. We note that a version of this model that did not account for vaccination or mutation was considered in our earlier work.[2]

2.4.1 Regression model for wastewater concentration of SARS-CoV-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) normalized by pepper mild mottle virus (PMMoV).[13] Bayesian linear regression was performed both on the county aggregated data and stratified by sub-community wastewater treatment plant zones (sewersheds). To improve the regression model stability, we used weekly average prevalence counts from the SVI_2RT model as the explanatory variable, and weekly aggregated average wastewater concentrations as the single outcome variable. 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 error term.

2.4.2 Estimating vaccination, strain mutation and hospitalization effects

The strong statistical significance of the regression model relating prevalence and wastewater concentration allowed for indirect estimation of the effect of population vaccination and virus mutation. Under the assumption the relationship between the wastewater concentration and the prevalence is not confounded by the vaccination and mutation process, we used the original regression equation derived from the collected wastewater and seroprevalence data to estimate the wastewater concentrations 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 all three analysis we quantified the effects by calculating the size of the effects relative to the factual (observed) states.

2.4.3 Competing risk model with two different virus strains

Wastewater samples were prepared for whole genome sequencing[11, 14], and the proportion of observed SARS-CoV-2 variants was estimated for each sewershed based on the frequency of mutations specific to each variant.

Two variants were present in the study area during the study period: Alpha was dominant April to July, while Delta was dominant July and August.[11, 14] To reflect the infections before and after the emergence of the Delta variant, we incorporated into our SVI_2RT model two different infection compartments (I_1 and I_2) reflecting the infection competition between two different strains of the virus.

2.5 *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). For the wastewater data, the University of Louisville Institutional Review Board classified this as non-human subjects research (reference #: 717950).

3. Results

3.1 Wastewater regression

The results of Bayesian regression analysis relating the prevalence estimated from the SVI_2RT model, and the observed wastewater levels, both in aggregation and by sewershed area, show a significant trend that is well summarized by the corresponding posterior regression line. Normalized SARS-CoV-2 by PMMoV provided more reliable and stable longitudinal concentration readings than using SARS-CoV-2 N1 concentration alone. For the aggregate model, the estimated intercept is $5{\cdot}563 \times 10^{-4}$ (CI =($-9{\cdot}903 \times 10^{-4}$, $-1{\cdot}250 \times 10^{-4}$)) and the slope is 0·453 (CI=(0·374, 0·529)). Overall, we see the regression model fits well with R^2=0·909. See Supplemental Material Figure S3.2 and S3.3.

3.2 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 (observed 64% vaccination rate and counter-factual 0% vaccination rate) while adjusting for the Delta variant emergence (Figure 1). The peak and the overall temporal dynamics are different under the two scenarios across each location. To better quantify these differences, we calculated the location-specific vaccination effects as the ratios of the areas under two scenario curves (with-vaccination area over without-vaccination area). The value obtained for the aggregated data was 0·429 (CI= (0·405, 1)), with the remaining sewershed specific effects being even stronger at (Figure 1; panels B–D) 0·532 (CI= (0·515, 1)), 0·367 (CI= (0·366, 0·785)), and 0·555 (CI= 0·555, 1)), respectively. Based on converting these ratios to excess incidence, we conclude that without vaccination, we would expect to see the incidence increase of about 133% above the observed level in Jefferson County (panel A) and about 88%, 172%, and 80% in respective sewershed areas (Figure 1; panels B–D, see also S3).

To obtain estimates of the vaccination effects on the wastewater concentrations, we developed a hybrid inferential model combining the wastewater regression (see Sec 3.1) equation with the SVI_2RT estimated prevalence, under two different vaccination scenarios (factual 64% rate and counter-factual 0% rate) (Figure 2). Note that the usage of SVI_2RT (which accounts for the effect of different virulence of the two different SARS-CoV-2 strains) automatically adjusted our analysis for the Delta variant emergence. As the estimated prevalence from the SVI_2RT model and the normalized wastewater concentration are highly correlated (see Sec 3.1), the hybrid model is seen to fit 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 (Figure 2; panel A) ratio was equal to 0·358 (CI= (0·333, 0·381)), and the remaining sewershed area ratios (Figure 2; panels B–D) were equal to, respectively, 0·457 (CI= (0·388, 0·537)), 0·276 (CI= (0·260, 0·296)), and 0·426 (CI= 0·407, 0·446)). The estimate of excess wastewater virus without vaccination is estimated as 179%, 119%, 262%, and 135%, respectively (Supplemental material section S3).

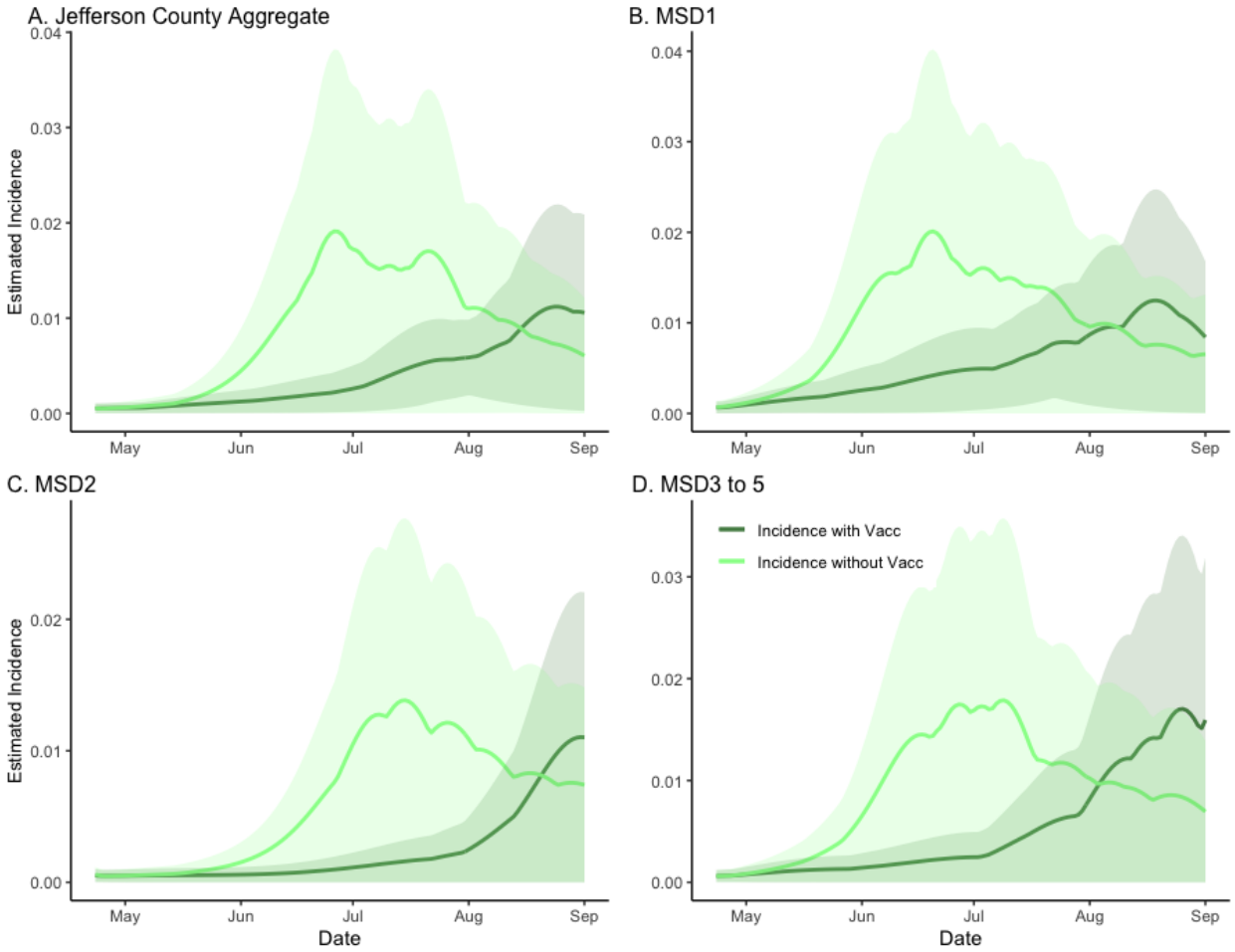

Figure 1. **The estimated effect of vaccination on incidence in sewersheds of Jefferson County, KY (USA).** The dark green line is the factual SVI_2RT model estimated incidence (with vaccination), and the light green line is the corresponding counter-factual estimated incidence with vaccination effect zeroed out. The shaded areas represent 95% credible intervals. The panels compare the vaccination effect in Jefferson County (Panel A) as well as stratified by sewershed (Panels B–D).

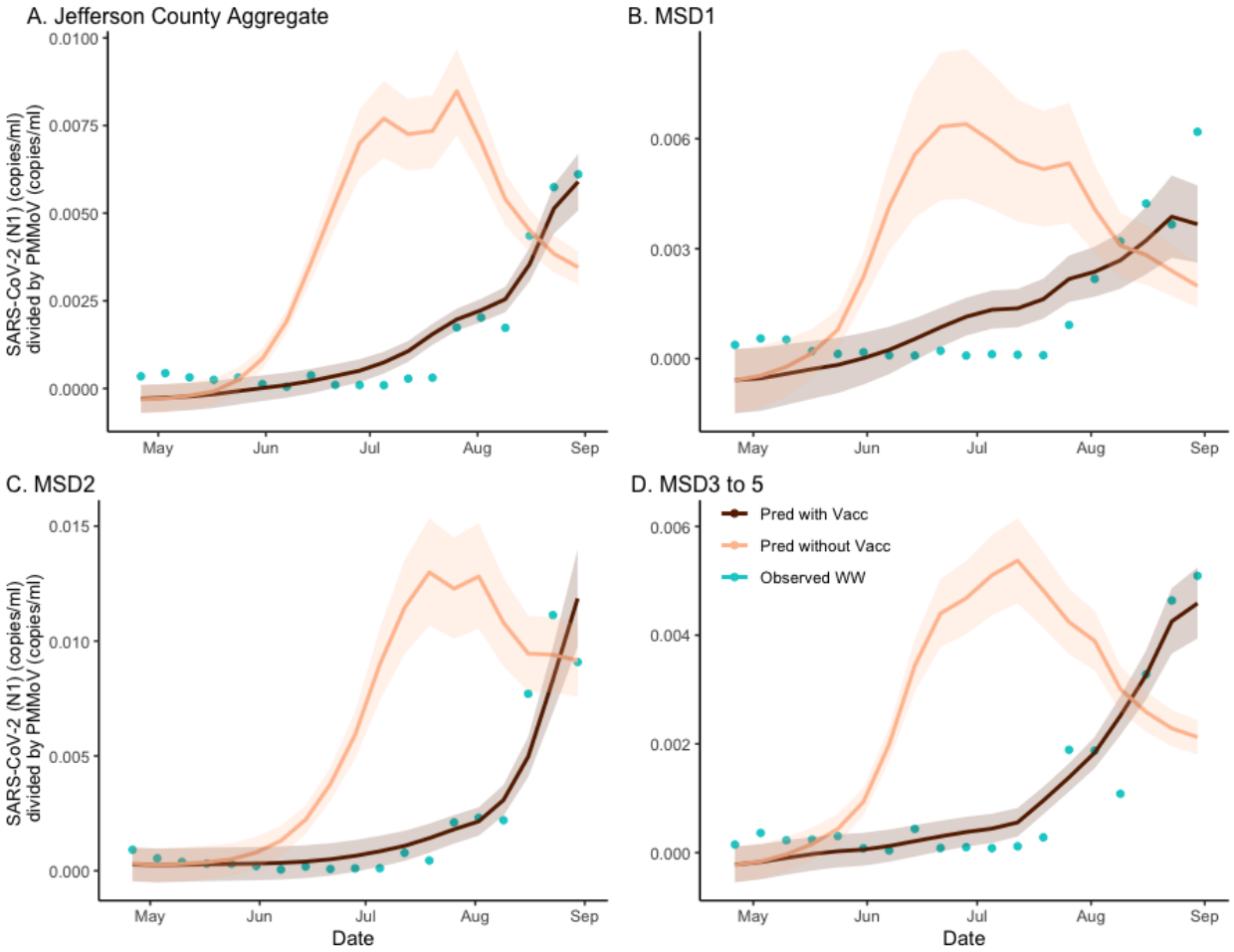

Figure 2. **The estimated effect of vaccination on SARS-CoV-2 wastewater concentration normalized by pepper mild mottle virus in sewersheds of Jefferson County, KY (USA).** The deep brown line is the regression-based fit to the wastewater concentration data and the light brown line is the prediction of wastewater concentration using synthetic prevalence from 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 (Panel A) as well as stratified by sewershed (Panels B–D).

3.3 Effects of virus mutation on disease incidence and wastewater concentration

The time periods during which the Alpha and Delta variants were dominant in each sewershed are are shown in Table S4.1. The onset of Alpha as the dominant variant simultaneously occurred at each site on 3/30/21 and lasted a variable number of weeks before dying out and eventually becoming replaced by Delta. However, Delta had a more gradual introduction into the sewersheds, beginning as the dominant variant on 7/12/21 in two of the sites and not showing up as the dominant variant in one of the sites until two weeks later. Our gradual linear switch from Alpha to Delta is similar to observations in other wastewater data, including a study of 94 sites within Austria.[22] Interestingly, Delta's dominance as the major variant ended simultaneously on 8/30/21 in all five sites, prior to the later emergence of Omicron in the sewersheds in December 2021. More recently, we have reported on the re-emergence of Delta in the MSD03 site specifically during the Omicron wave,[14] which indicates the persistence of specific variants in wastewater can be variable in nature, and are likely influenced by a number of factors, including incidence and vaccination rates.

In the analysis, we assumed a 20% higher infectivity of the SARS-CoV-2 Delta variant as compared to its Alpha predecessor.[15] In the counterfactual model (light curve), where only the Alpha variant was present, the epidemic was seen to dissipate, indicating the basic reproduction number less than one. This was in stark contrast with the factual, full SVI_2RT model fit (with both Alpha- and Delta- variants present), where the incidence (dark curve) was seen to rise rapidly (Figure 3). As in the previous section, to quantify the difference between the two curves, which we interpret as measuring the effect of introducing the Delta mutation, we calculated the ratio of areas under the two curves in each panel, obtaining the values of 7·32 (CI = (7·05, 20·13)), 4·40 (CI = (4·33, 7·64)), 8·58 (CI = (1, 8·60)), and 6·15 (CI = (1, 6·16)) for the aggregate, MSD1, MSD2, and MSD3–5 regions respectively (corresponding to panels A–D). The estimate of the decrease in total incidence without mutation is found as 86%, 77%, 88%, and 84%, respectively.

To identify the effect of the Delta variant emergence on the observed wastewater concentration, we again applied the hybrid model discussed in the previous section. 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). The results are depicted in the panels of Figure 4 both for the aggregated and sewershed-specific analysis. 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 location-specific ratio values of 12·569 (CI = (11·487, 13·914)), 6·235 (CI = (5·290, 7·891)), 14·932 (13·351, 16·898), and 8·413 (CI = (7·654, 9·351)), corresponding to aggregated and sewershed-specific curves, the estimate of excess wastewater virus due to Delta mutation is founded as 92%, 84%, 93%, and 88% respectively. Further analysis is provided in Supplemental Material Table S3.3.

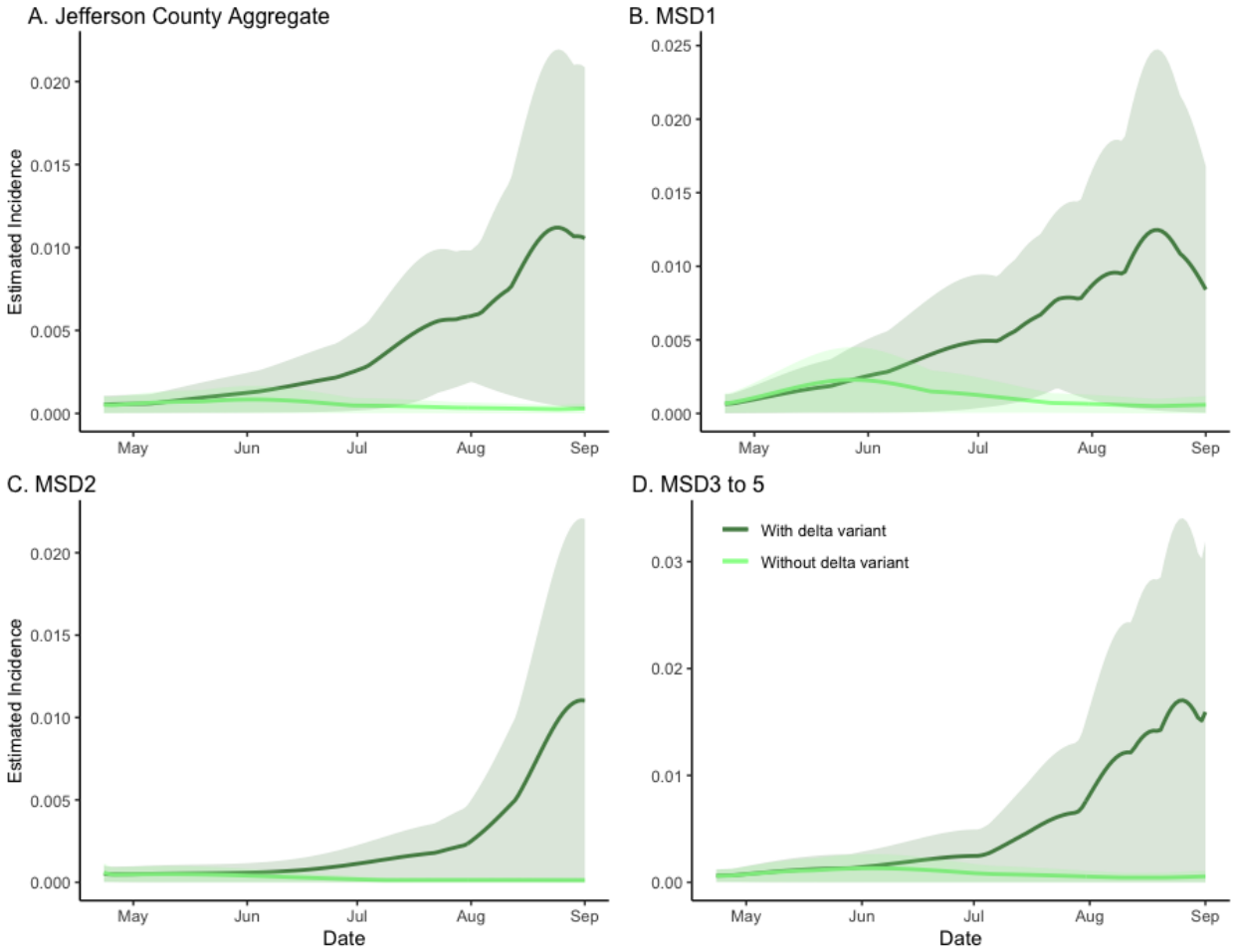

Figure 3. **The model-based analysis of the Delta-variant effect on SARS-CoV-2 incidence rate estimates in sewersheds of Jefferson County, KY (USA).** The dark green line is the estimated factual full model incidence (both Alpha and Delta variants present), and the light green line is the counterfactual incidence estimated from the model with no Delta variant. The shaded areas represent 95% credible intervals. The panels compare the vaccination effect in Jefferson County (Panel A) as well as stratified by sewershed (Panels B–D).

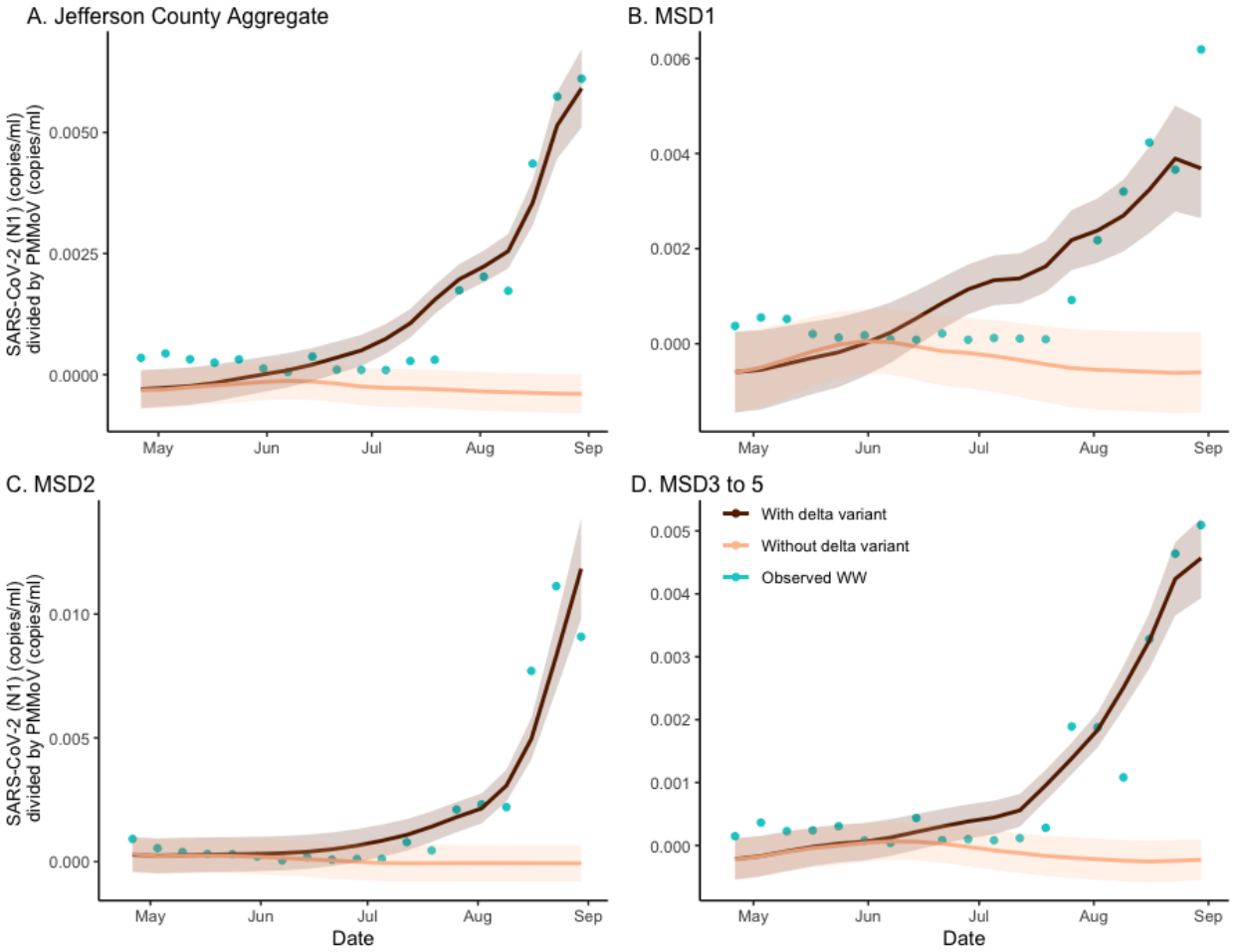

Figure 4. **The estimated effect of Delta variant on SARS-CoV-2 wastewater concentration normalized by pepper mild mottle virus in sewersheds of Jefferson County, KY (USA).** The deep brown line is the regression-based fit to the wastewater concentration and the light brown line is the prediction of wastewater concentration using synthetic prevalence from the model with the Delta variant effect zeroed out. The shaded areas represent 95% credible intervals. The blue dots are observed weekly average wastewater concentration. The panels compare the mutation effect on wastewater concentration for Jefferson County (Panel A) as well as stratified by sewershed (Panels B–D).

3.4 Forecasting hospitalization rates based on wastewater concentrations

Hospitalization estimates under both vaccinated (64% vaccination rate) and unvaccinated (0% vaccination rate) scenarios were obtained by 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 (Figure 5). As hospitalization is likely to occur sometime after symptom onset, we used the 1-week lagged-regression model where the length of the lag was based on the overall model fit criteria. The fitted intercept and slope coefficients were $1{\cdot}284 \times 10^{-4}$ (std=$2{\cdot}279 \times 10^{-5}$) and 0·176 (std=0·0119) for vaccinated and unvaccinated scenarios respectively, with the R-square of 0·928. The maximal number of the observed daily average hospitalizations under 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 150·3. The ratios between the areas under the prediction curves with and without vaccination were 0·368 (CI = (0·413, 0·458)), indicating a 170% 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 mutation. The ratio of the areas under the two graphs (with and without the Delta variant mutation) is 2·632 (CI = (2·382, 5·573)), indicating a 62% decreasing in the hospitalization rate.

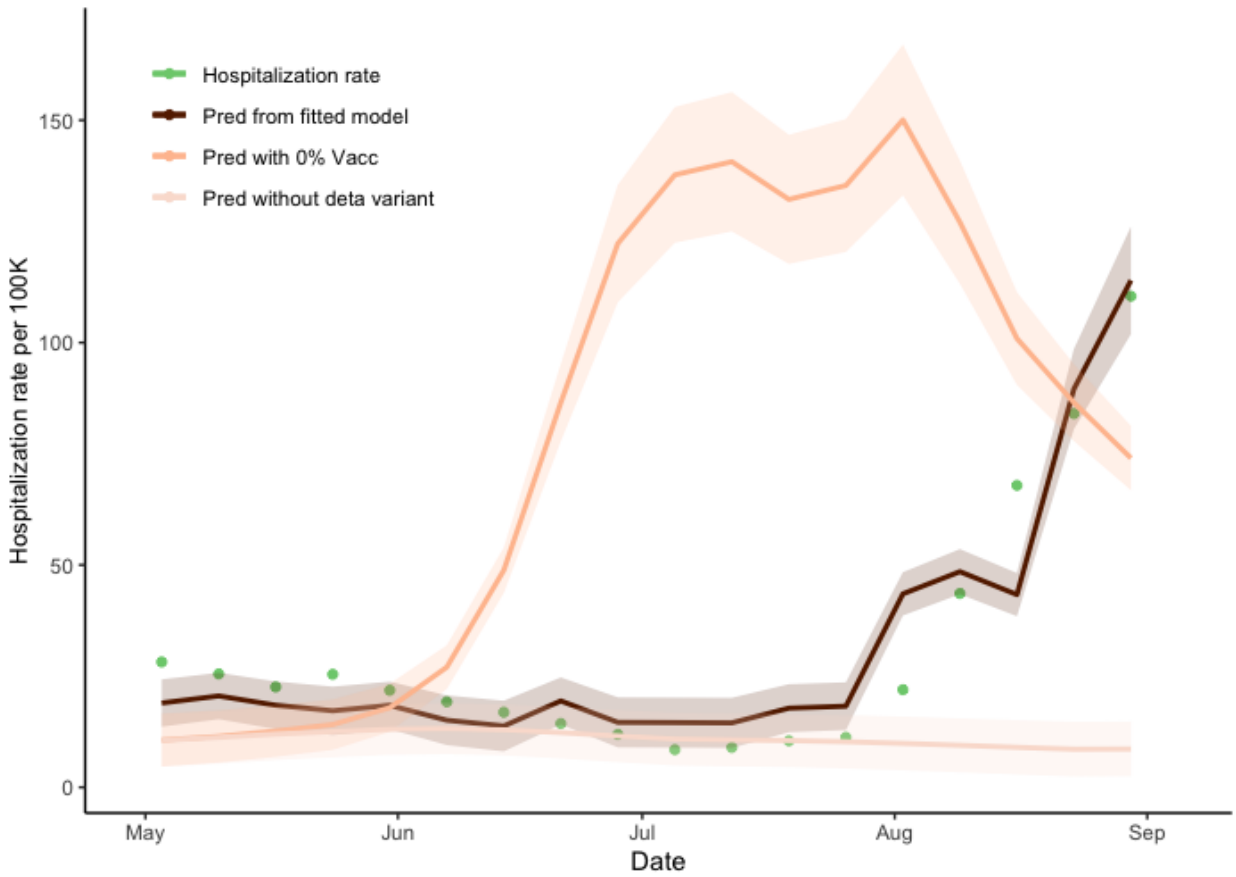

Figure 5. **Time lag-dependent analysis of the relationship between hospitalization rate and wastewater concentration, Jefferson County, KY (USA).** Predictions and 95% confidence intervals of hospitalization rate regressed on week-lagged variables of the weekly average of wastewater concentration according 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 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 0·368, and in the absence of the Delta variant is 2·63, respectively.

Discussion

The results of our large study (N=3436) have shown the importance of environmental surveillance post-vaccine in urban and sub-urban areas; removing bias of publicly available clinical case data by using antibody positivity with four waves of sequential community-wide stratified randomized sampling data. Despite our focus on the Jefferson County example, it should be emphasized the model described here is readily applicable to other locations worldwide. Although our model was run with both the N1 analysis and adjusted N1 analysis, we learned the model provided reduced uncertainty with adjustment. Indeed, the vaccination effect estimates bring 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, wastewater surveillance may be also used to estimate both the effects of a community intervention, like the vaccination campaign and the effects of disease evolution, and emergence of a new viral strain mutation.

The COVID-19 vaccine has been highly effective.[16, 17] There is a need for increased reliance on wastewater as a proxy for community disease impact being built from actual community level data over time for other vaccine preventable diseases affecting human health. When 90% of the student population of a college campus was vaccinated, SARS-CoV-2 in wastewater decreased;[4] but that building level generalization was not replicated in our community-wide survey over a longer period. In contrast to our findings, Nourbakhsh et al.[18] found dissimilar trajectories from clinical and wastewater data 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 for the studied period when compared to earlier in the pandemic and with no requirement, or in some cases option, for reporting. Whereas the Nourbakhsh et al.[18] study included only publicly reported case data, the randomized selection of community participants in our study population were a comparatively less biased data source for this post-vaccine period.

Our model compares to the work of Jiang et al.[19] in that our analysis also provided estimates of prevalence, however our estimates are based on a statistical random sample (not a clinical sample) and our regression model has a simple and explicit formula relating prevalence to observed wastewater concentrations. Our model further confirms the findings in Hegazy et al.[6] implying the Delta variant emergence has strengthened the relationship between wastewater and hospitalization rates. 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 0·916 (CI = (0·764, 0·976). The hospital burden findings of Wang et al.[20] also compared 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 approximately 170%.

Yaniv et. al[5] described the emergence of peaks of positivity rates, showing they corresponded to introduction of new variants. In addition, they noted how vaccination rates and a second booster helped to control Alpha, while an increase in a third booster was found to lead to a decline in Delta. When vaccination levels increase to higher coverage, overall reported incidence may

decline, even though the levels in wastewater remain high.[7] This leads to the hypothesis that circulation among vaccinated individuals creates a level of selective pressure making variants with transmission and vaccine escape more advantageous. In our wastewater data, we find in periods of decline for a specific variant we see more diversity in the overall mutations, indicating both a selection against the variant in decline (typically due to vaccinations and boosters), and a similar advantageous selection for emerging variants.

Our study used five sub-community scales based on the existing wastewater infrastructure allowing observation of regional trends but also the aggregation of data for a countywide picture. We found the antibody positivity varied by the sewershed. The antibody-positive individuals were lowest in sewershed MSD1 and highest in sewershed MSD3–5 (10·0% for aggregate, 8·6% for MSD1, 9·2% for MSD2, and 12·0% for MSD3–5), indicating previous infection may have been higher in the rural portions of the county compared to the urban core. There are many 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. These differences between MSD1 to 5 provide evidence of the benefit of observing results at a sub-county level, rather than only considering MSD1 representing the urban core and the largest sewershed zone.

The trajectory of the pandemic and public health response would benefit from new methods less dependent on continuous individual clinical testing. For replication of this SARS-CoV-2 model, wastewater sampling, stratified random sampling of seroprevalence, and spatially linked vaccination data are required; the model is flexible enough to allow additional variant-specific variables. The promise of this model is if we just have wastewater concentration, we can we predict the effect of vaccination and allow fine-tuned, and milestone driven, public health response. Our model also provides a further positive response for public health offices of the significant role of the vaccine, our competition model shows the epidemic would have been bigger and earlier without vaccine access.

4. Limitations

Our study has several limitations. The proportion of vaccinated respondents was larger than the greater community (~90% vs. 64%) reflecting that even a probability-based sample relies on volunteer participants and vaccinated individuals may also have different health-research behaviors. Vaccine information was self-reported. 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 N, an assay that has 65% sensitivity and 85% specificity, with inevitable under-estimation of infection prevalence. While our serosurvey only captured adults, wastewater testing included minors. COVID-19 infected individuals can in rare instances shed fecal SARS-CoV-2 RNA up to 7 months post diagnosis;[21] viral shedding of each SARS-CoV-2 variant in relation to days after vaccination, or whether a primary or booster series had been taken, is not well defined as the vaccination series guidance extends during the pandemic and thus was not included in our model.

5. Conclusion

Our work indicates it is possible in certain conditions to use wastewater-based epidemiology to assess both the immunity acquisition in the community due to natural recovery and vaccination as well as the effect of new variants emergence and associated immune evasion to the currently available COVID-19 vaccines. The effects 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 133% of excess infections over the period of study, which corresponds to a 179% of excess wastewater concentration. The effect of Delta mutation was similarly substantial. We estimated, over the study period in Jefferson County, without Delta mutation the amount of overall infection would decrease on average by 86% which corresponds to a 92% decrease in wastewater N1 normalized by PMMoV concentration. 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 = 0·963. Based on the regression model we estimated the effects of vaccination and mutation on hospitalization rate. According to the model, without vaccination one would expect about 171% increase and without mutation about 62% decrease in hospitalization rate. Using the fitted regression model for hospitalization, the predictions of hospitalization rates are at 50, 100, and 150 per 100K when the normalized wastewater concentrations are 0·0021, 0·0050, and 0·0078 N1 normalized by PMMoV, respectively. Our large, randomized, serosurvey suggests using the mechanistic, population level, vaccination model (SVI_2RT) coupled with longitudinal wastewater sampling 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. The model can also be used to estimate the effects of vaccination and new variants emergence on the hospitalization rate and on peak hospital beds utilization.

Declaration of interests:
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.

Acknowledgements:
We thank the Louisville/Jefferson County Metropolitan Sewer District for their valuable collaboration in wastewater sample collection.

Contributors:
AB, TS, KEP, and RK conceived and developed the idea for the study. KEP supervised the serology assays; ABS and KK performed the serology assays, performed quality control assessments, and uploaded the data to the redcap database. TS and RH supervised the wastewater analyses. GAR and BC performed the model-based data analyses. RHH and GAR wrote the first draft of the manuscript with all authors contributing to the interpretation of data and critical revision of the article. MB accessed and verified the data reported in the study. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Data sharing:
The seroprevalence, wastewater concentration, and hospitalization information data used in the study can be accessed from the website https://github.com/cbskust/DSA_Seroprevalence. The computer code that implemented our model-based analysis will be made available immediately after publication.

References

1. Li X, Zhang S, Sherchan S, 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* 2023; **441**: 129848. doi:10.1016/j.jhazmat.2022.129848
2. Smith T, Holm RH, Keith RJ, et al. Quantifying the relationship between sub-population wastewater samples and community-wide SARS-CoV-2 seroprevalence. *Sci Total Environ* 2022; **853**: 158567. doi:10.1016/j.scitotenv.2022.158567
3. Berchenko Y, Manor Y, Freedman LS, et al. Estimation of polio infection prevalence from environmental surveillance data. *Sci Transl Med* 2017; **9**(383): eaaf6786. doi:10.1126/scitranslmed.aaf6786
4. Bivins A, Bibby K. Wastewater surveillance during mass COVID-19 vaccination on a college campus. *Environ Sci Tech Let* 2021; **8**(9): 792–8. doi:10.1021/acs.estlett.1c00519
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* 2021; **207**: 117808. doi:10.1016/j.watres.2021.117808
6. [preprint] Hegazy N, Cowan A, D'Aoust PM, et al. Understanding the dynamic relation between wastewater SARS-CoV-2 signal and clinical metrics throughout the pandemic. *medRxiv* 2022. doi:10.1101/2022.07.06.22277318
7. Nattino G, Castiglioni S, Cereda D, 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* 2022; **327**(19): 1922–4. doi:10.1001/jama.2022.4908
8. Rainey AL, Loeb JC, Robinson SE, 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* 2022; **208**: 112496. doi:10.1016/j.envres.2021.112496
9. Bilal U, Tabb LP, Barber S, Diez Roux AV. Spatial inequities in COVID-19 testing, positivity, confirmed cases, and mortality in 3 US cities: an ecological study. *Ann Intern Med* 2021; **174:** 936–44. doi:10.7326/M20-3936
10. Hamorsky KT, Bushau-Sprinkle AM, Kitterman K, et al. Serological assessment of SARS-CoV-2 infection during the first wave of the pandemic in Louisville Kentucky. *Sci Rep-UK* 2021; **11**: 18285. doi:10.1038/s41598-021-97423-z
11. Rouchka EC, Chariker JH, Saurabh K, et al. The rapid assessment of aggregated wastewater samples for genomic surveillance of SARS-CoV-2 on a city-wide scale. *Pathogens* 2021; **10**: 1271. doi:10.3390/pathogens10101271
12. CDC. COVID-19 Vaccinations in the United States, County. 2022. https://data.cdc.gov/d/8xkx-amqh/visualization (accessed January 5, 2023).
13. Maal-Bared R, Qiu Y, Li Q, 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* 2023; **856**: 158964. doi:10.1016/j.scitotenv.2022.158964
14. Westcott CE, Sokoloski KJ, Rouchka EC, et al. The detection of periodic reemergence events of SARS-CoV-2 Delta strain in communities dominated by Omicron. *Pathogens* 2022; **11**(11): 1249. doi:10.3390/pathogens11111249

15 Yang, W, Shaman, J. Development of a model-inference system for estimating epidemiological characteristics of SARS-CoV-2 variants of concern. *Nat Commun* 2021; **12**.1: 1–9. doi:10.1038/s41467-021-25913-9

16 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* 2022; **399**(10327): 814–23. doi:10.1016/S0140-6736(22)00089-7

17 Watson OJ, Barnsley G, Toor J, Hogan AB, Winskill P, Ghani AC. Global impact of the first year of COVID-19 vaccination: A mathematical modelling study. *Lancet Infect Dis* 2022; **22**(9): 1293–302. doi:10.1016/S1473-3099(22)00320-6

18 Nourbakhsh S, Fazil A, Li M, et al. A wastewater-based epidemic model for SARS-CoV-2 with application to three Canadian cities. *Epidemics* 2022; **39**: 100560. doi:10.1016/j.epidem.2022.100560

19 Jiang G, Wu J, Weidhaas J, et al. Artificial neural network-based estimation of COVID-19 case numbers and effective reproduction rate using wastewater-based epidemiology. *Water Res* 2022; **218**: 118451. doi:10.1016/j.watres.2022.118451

20 Wang H, Churqui MP, Tunovic T, 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*. 2022; **25**(9): 105000. doi:10.1016/j.isci.2022.105000.

21 Natarajan A, Zlitni S, Brooks EF, et al. Gastrointestinal symptoms and fecal shedding of SARS-CoV-2 RNA suggest prolonged gastrointestinal infection. *Med* 2022; **3**(6): 371–87.e9. doi:10.1016/j.medj.2022.04.001.

Amman F, Markt R, Endler L, et al. Viral variant-resolved wastewater surveillance of SARS-CoV-2 at national scale. *Nat Biotechnol* 2022: 1–9. doi:10.1038/s41587-022-01387-y

Supplementary Material

Table of Contents

S1. SARS-CoV-2 seroprevalence by wave and sewershed, Jefferson County, KY (USA)..... 23

S2. Studied wastewater treatment plant zones (sewersheds), Jefferson County, KY (USA). 24

S3. Population vaccination model (SVI_2RT)..... 26

S4. Variant detection..... 40

S1. SARS-CoV-2 seroprevalence by wave and sewershed, Jefferson County, KY (USA).

Table S1. **SARS-CoV-2 seroprevalence by wave and sewershed, Jefferson County, KY (USA).**

	Number of unvaccinated participants	Number of vaccinated participants	Number of participants positive for SARS-CoV-2 nucleocapsid (N) specific IgG antibodies	Estimated posterior average seroprevalence per 10^5 people (95% credible interval	Estimated posterior average prevalence per 10^5people (95% credible interval)
Overall					
MSD1	91	1450	132	9153 (3772, 14533)	533 (38, 1028)
MSD2	132	794	81	5427 (1848, 9006)	336 (0, 672)
MSD3-5	23	228	17	8410 (2016, 14804)	588 (0, 1177)
Total	315	2957	296	22989 (13898, 32081)	1457 (618, 2296)
Wave A					
MSD1	26	363	24	2249 (1686, 2811)	62 (1, 122)
MSD2	29	206	25	2277 (1751, 2803)	49 (0, 97)
MSD3-5	4	39	5	2284 (1682, 2887)	53 (1, 106)
Total	68	697	68	6810 (5832, 7788)	163 (70, 257)
Wave B					
MSD1	20	369	17	2927 (2042, 3813)	197 (7, 386)
MSD2	40	192	17	2611 (1832, 3391)	74 (0, 149)
MSD3-5	7	53	2	2952 (1816, 4089)	271 (0, 542)
Total	89	730	42	8491 (6853, 10129)	542(203, 881)
Wave C					
MSD1	16	308	22	6261 (2537, 9984)	618 (43, 1193)
MSD2	29	179	11	3696 (1860, 5531)	197 (0, 394)
MSD3-5	3	60	2	7565 (2048, 13083)	456 (0, 911)
Total	60	657	46	17521 (10617, 24426)	1270 (511, 2030)
Wave D					
MSD1	29	410	69	19572 (8999, 30145)	810 (4, 1617)
MSD2	34	217	28	12408 (1872, 22943)	834(0, 1668)
MSD3-5	9	76	8	14746 (2188, 27304)	1440 (0, 2880)
Total	98	873	140	46726 (27220, 66232)	3084 (1235, 4934)

S2. Studied wastewater treatment plant zones (sewersheds), Jefferson County, KY (USA).

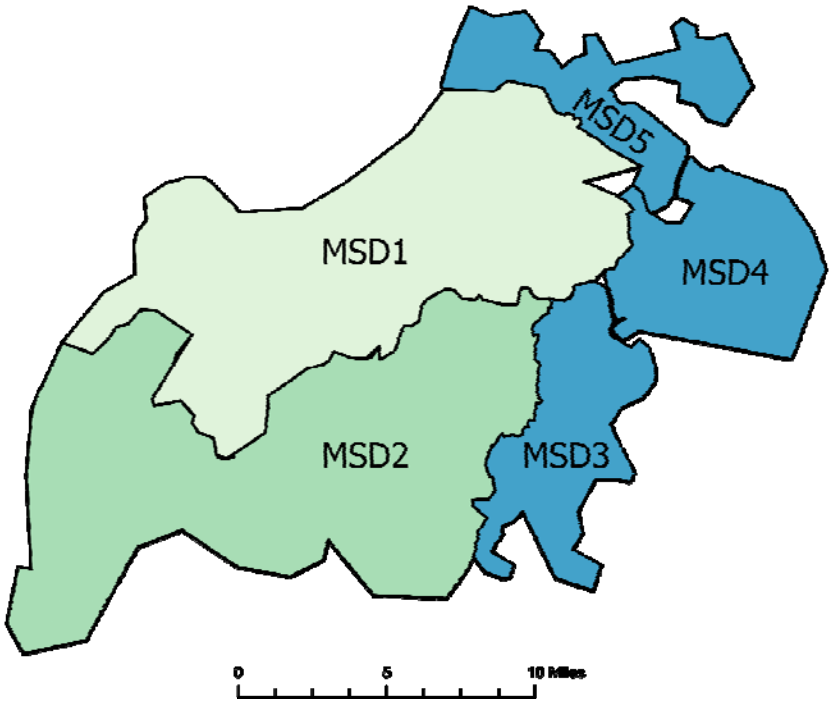

Figure S1. **Studied wastewater treatment plant sewersheds, Jefferson County, Kentucky (USA).**

Table S2. **Characteristics of studied wastewater treatment plant sewersheds of Jefferson County, KY (USA).**

Site name	Income (USD$)[a]	Population[a]	Area (km^2)	Combined sewer
MSD01 Morris Forman Water Quality Treatment Center (MFWQTC)	54,138	349,850	280	Yes
MSD02 Derek R. Guthrie Water Quality Treatment Center (DRGWQTC)	53,577	295,910	332	No
MSD03 Cedar Creek Water Quality Treatment Center (CCWQTC)	76,606	55,928	80	No
MSD04 Floyds Fork Water Quality Treatment Center (FFWQTC)	113,699	32,460	88	No
MSD05 Hite Creek Water Quality Treatment Center (HCWQTC)	106,769	31,269	67	No

[a] Based on 2018 U.S Census Bureau American Community Survey (ACS). Income is mean median household.

S3. Population vaccination model (SVI_2RT)

The equation shown in (1) describes the time-evolution of the proportions of individuals who are susceptible (S), vaccinated (V), infected with Alpha variant(I_1), infected with Delta variant (I_2) removed (R), and seropositive (T). We assume the total initial population of susceptibles is large with a small initial fraction of infected. The model equations are

$$\begin{aligned}
\dot{S}_t &= -\beta S_t I_t^{(1)} - \beta^* S_t I_t^{(2)} - \alpha_t S_t, \\
\dot{V}_t &= \alpha_t S_t - \tilde{\beta} I_t^{(1)} V_t - \tilde{\beta}^* I_t^{(2)} V_t \\
\dot{I}_t^{(1)} &= \beta S_t I_t^{(1)} + \tilde{\beta} I_t^{(1)} V_t - \gamma I_t^{(1)}, \\
\dot{I}_t^{(2)} &= \beta^* S_t I_t^{(2)} + \tilde{\beta}^* I_t^{(2)} V_t - \gamma I_t^{(2)} \\
\dot{R}_t &= \gamma I_t - \delta R_t, \\
\dot{T}_t &= \delta R_t,
\end{aligned} \tag{1}$$

with the initial condition $S_0 = 1 - \rho - \epsilon - \psi > 0, V_0 = 0$, $I_0^{(1)} = \rho > 0$, $I_0^{(2)} = \rho/100$, $R_0 = \epsilon > 0$, and $T_0 = \psi > 0$.

Here, β and $\tilde{\beta}$ are the rates of infection of respectively, unvaccinated and vaccinated, and β^* and $\tilde{\beta}^*$ are the rates of infection according to Delta variant. As our compartment model has two infection compartments, it is called the variant competition model.[1] The observed data in this analysis do not have any information about infection from the Delta variant, and an increase in the number of parameters makes model estimation difficult and may lead to identifiability problems. So, we set β^* and $\tilde{\beta}^*$ at the values 20% higher than β and $\tilde{\beta}$.[2] The function of α_t represents a changing rate of vaccination over time. However, the vaccination process can not follow a stochastic process but may be changed according to a policy or vaccine supply, so set the vaccination rate $\alpha(t)$ to match the empirical percentage of the vaccinated population in Jefferson County at the end of August 2021. Additionally, γ is the rate of recovery, and δ is the rate at which antibodies build to a detectable level after recovery. The SVI_2RT model parameters to be estimated are given by the vector $\theta = (\beta, \tilde{\beta}, \gamma, \delta, \rho, \epsilon, \psi)$.

To obtain the serial estimates of incidence and prevalence from the observed seropositivity levels in four waves of testing, we adapt the idea of an ODE-based survival model proposed recently.[3, 4] According to that model, the scaled quantities $S_t, V_t, I_t^{(1)}, I_t^{(2)}, R_t, T_t$ may be considered as respective probabilities of a randomly selected individual in a large population, being either susceptible, vaccinated, infected with different virus variant, recovered, or seroprevalent at time t. Consequently, we consider the results $Z(t)$ of all individual antibody-based tests conducted at times t as independent Bernoulli variables:

$Z(t) \sim \mathrm{Ber}(T_t^*)$,

where $T_t^* = sensT_t + (1 - spe)(1 - T_t)$ is the specificity adjusted probability of a positive test. For our analysis, both $sens$ and spe are additional parameters to be estimated. We assigned the informative priors to $sens$ and spe from available clinical data.

Assuming at time t, n_t individuals are tested with k_t having positive results, the corresponding log-likelihood function is

$\ell_t(\Theta) \propto k_t \log(T_t) + (n_t - k_t)\log(1 - T_t)$, (2)
where $\Theta = (\beta, \tilde{\beta}, \gamma, \delta, \rho, \epsilon, \psi, spe, sens)$ is the vector of parameters to be identified.
Given the testing data at $m \geq 1$ time points $t_1, \ldots, t_m$, we then aim to find parameter values θ that maximizes the posterior log-likelihood function
$\tilde{\ell}\,(\Theta) \propto \sum_{i=1}^{m} \ell_{t_i}\,(\Theta) + \log p(\Theta)$, (3)
where $p(\Theta)$ is the prior distribution on Θ to be determined from our previous work.[3] Hence, we seek the values of Θ that maximize our posterior log-likelihood function (3). Note the entire system (1) must be solved for each parameter combination.

S3.1 Incidence, prevalence, and seroprevalence estimation

Posterior serial estimates of the relative rates of incidence, prevalence, and seropositivity were obtained from the SVI_2RT model as the time-dependent vector
$\mathrm{Pred}_t = (-\dot{S}_t, V_t, I_t^{(1)}, I_t^{(2)}, T_t)$. (4)
Here $(S_t, V_t, I_t^{(1)}, I_t^{(2)}, T_t)$ is the family of trajectories of (1) evaluated at the posterior distribution of the vector Θ. In practice, the distribution of Pred_t is approximated by taking a random sample of size m from the converged MCMC sampler. In our case $m = 2000$. To obtain daily incidence rates (Inc_d) we have used the approximation $\dot{S}_t \approx S_{t+1} - S_t$ and consequently took $\mathrm{Inc}_d = S_d - S_{d+1}$ where d corresponds to a specific day of interest. The estimated prediction counts were obtained by multiplying the rates in Pred_t by the appropriate population numbers.

Table S3.1. **Posterior mean estimates of the SVI_2RT model parameters in sewersheds of Jefferson County, KY (USA).** The area-specific Hamiltonian Markov chain Monte Carlo (MCMC) posterior estimates are based on seropositivity data aggregated across Jefferson County and stratified by sewersheds. The corresponding 95% credible bounds are provided in parenthesis. The results are based on MCMC implemented via *Rstan* library, with a 6000- and 2000-step burn-in.

	Jefferson County Aggregated	**MSD1**	**MSD2**	**MSD3–5**
$\boldsymbol{\beta}$	0·494 (0·389, 0·572)	0·493 (0·391, 0·575)	0·419 (0·309, 0·503)	0·459 (0·347, 0·542)
$\boldsymbol{\alpha}$	0·009 (64%)	0·010 (67%)	0·007 (55%)	0·013 (76%)
$\tilde{\boldsymbol{\beta}}$	0·420 (0·321, 0·502)	0·423 (0·3254, 0·500)	0·412 (0·384, 0·587)	0·445 (0·335, 0·532)
$\boldsymbol{\gamma}$	0·463 (0·376, 0·532)	0·453 (0·371, 0·520)	0·494 (0·384, 0·587)	0·461 (0·371, 0·533)
$\boldsymbol{\delta}$	0·104 (0·067, 0·136)	0·104 (0·066, 0·136)	0·105 (0·069, 0·138)	0·104 (0·066, 0·135)
$\boldsymbol{\rho}$	$3{\cdot}750 \times 10^{-4}$ ($6{\cdot}385 \times 10^{-5}$, $7{\cdot}426 \times 10^{-4}$)	$4{\cdot}781 \times 10^{-4}$ ($9{\cdot}191 \times 10^{-5}$, $9{\cdot}221 \times 10^{-4}$)	$4{\cdot}382 \times 10^{-4}$ ($6{\cdot}034 \times 10^{-5}$, $9{\cdot}113 \times 10^{-3}$)	$4{\cdot}494 \times 10^{-4}$ ($7{\cdot}366 \times 10^{-5}$, $9{\cdot}099 \times 10^{-4}$)
$\boldsymbol{\epsilon}$	$1{\cdot}506 \times 10^{-3}$ ($1{\cdot}396 \times 10^{-4}$, $3{\cdot}471 \times 10^{-3}$)	$1{\cdot}638 \times 10^{-3}$ ($1{\cdot}170 \times 10^{-4}$, $3{\cdot}618 \times 10^{-3}$)	$1{\cdot}635 \times 10^{-3}$ ($1{\cdot}802 \times 10^{-4}$, $3{\cdot}739 \times 10^{-3}$)	$1{\cdot}649 \times 10^{-3}$ ($1{\cdot}526 \times 10^{-4}$, $3{\cdot}619 \times 10^{-3}$)
$\boldsymbol{\psi}$	0·0223 (0·0182, 0·0253)	0·0222 (0·0183, 0·0253)	0·0223 (0·0184, 0·0254)	0·0223 (0·0184, 0·0254)
Specificity	0·946 (0·934, 0·954)	0·958 (0·941, 0·969)	0·926 (0·905, 0·941)	0·929 (0·898, 0·948)
Sensitivity	0·637 (0·539, 0·705)	0·638 (0·540, 0·703)	0·648 (0·556, 0·714)	0·643 (0·551, 0·711)

Table S3.2. **The prior distribution specifications for the SIRT model.** Parameters were given Gamma prior distributions, with hyper-parameters (*a*, *b*) listed in the table below.

Gamma (a, b)	β	$\tilde{\beta}$	γ	δ	ρ	ε	ψ	**Specificity**	**Sensitivity**
a	40·97	40·97	41·80	24·29	2·18	1·74	112·5	21·7	71
b	92·32	92·32	90·32	232·00	4648	1039·09	5035·15	3·83	38·3

Table S3.3. **Summary of the effects of the vaccination and Delta variant mutation in sewersheds of Jefferson County, KY (USA).** Percentage reduction due to vaccination effect or excess due to Delta variant on estimates of wastewater concentration (WW) and incidence rate (Incidence). In parenthesis, we give lower and upper bounds of 95% credible (incidence) or confidence interval (wastewater). Some values were rounded down to 0 for simplicity.

		Jefferson County Aggregated	MSD1	MSD2	MSD3–5
Vaccination effect with Delta variant	WW	179 (162, 200)	119 (86, 157)	262 (238, 284)	135 (124, 146)
	Incidence	133 (0, 147)	88 (0, 94)	173 (27, 173)	80 (0, 80)
Vaccination effect without Delta variant	WW	1370 (1272, 1493)	603 (479, 818)	823 (801, 849)	928 (848, 1027)
	Incidence	564 (0, 573)	323 (0, 330)	363 (0, 364)	491 (13, 491)
Delta variant effect without vaccination	WW	1157 (1049, 1291)	524 (429, 689)	1393 (1235, 1590)	741 (665, 835)
	Incidence	632 (1913, 605)	340 (664, 333)	758 (0, 760)	516 (18, 516)
Delta variant effect without vaccination	WW	159 (134, 185)	98 (73, 126)	478 (389, 581)	101 (89, 113)
	Incidence	143 (51, 144)	99 (25, 99)	308 (8, 308)	101 (0, 101)

Table S3.4. **Summary of the Bayesian regression results in sewersheds of Jefferson County, KY (USA).** Dispersion (σ) is the standard deviation of the error term of the linear regression.

Sewershed	Parameters	Linear regression model Posterior mean (95% credible interval)
Jefferson County Aggregated	Intercept	$-5{\cdot}563 \times 10^{-4}$ ($-9{\cdot}903 \times 10^{-4}$, $-1{\cdot}250 \times 10^{-4}$)
	Slope	0·453 (0·374, 0·529)
	Dispersion (σ)	$6{\cdot}434 \times 10^{-4}$ ($4{\cdot}589 \times 10^{-4}$, $9{\cdot}191 \times 10^{-4}$)
	Correlation	0·916 (0·764, 0·976)
MSD1	Intercept	$-8{\cdot}684 \times 10^{-4}$ ($-1{\cdot}757 \times 10^{-4}$, $4{\cdot}119 \times 10^{-5}$)
	Slope	0·357 (0·230, 0·481)
	Dispersion (σ)	$1{\cdot}143 \times 10^{-4}$ ($8{\cdot}237 \times 10^{-4}$, $1{\cdot}642 \times 10^{-3}$)
	Correlation	0·754 (0·463, 0·913)
MSD2	Intercept	$-2{\cdot}328 \times 10^{-4}$ ($-1{\cdot}001 \times 10^{-3}$, $5{\cdot}280 \times 10^{-4}$)
	Slope	0·932 (0·741, 1·114)
	Dispersion (σ)	$1{\cdot}339 \times 10^{-3}$ ($9{\cdot}603 \times 10^{-4}$, $1{\cdot}948 \times 10^{-3}$)
	Correlation	0·770 (0·376, 0·944)
MSD3–5	Intercept	$-3{\cdot}978 \times 10^{-4}$ ($-7{\cdot}405 \times 10^{-4}$, $-3{\cdot}951 \times 10^{-5}$)
	Slope	0·284 (0·236, 0·331)
	Dispersion (σ)	$5{\cdot}407 \times 10^{-4}$ ($3{\cdot}843 \times 10^{-4}$, $7{\cdot}697 \times 10^{-4}$)
	Correlation	0·896 (0·542, 0·976)

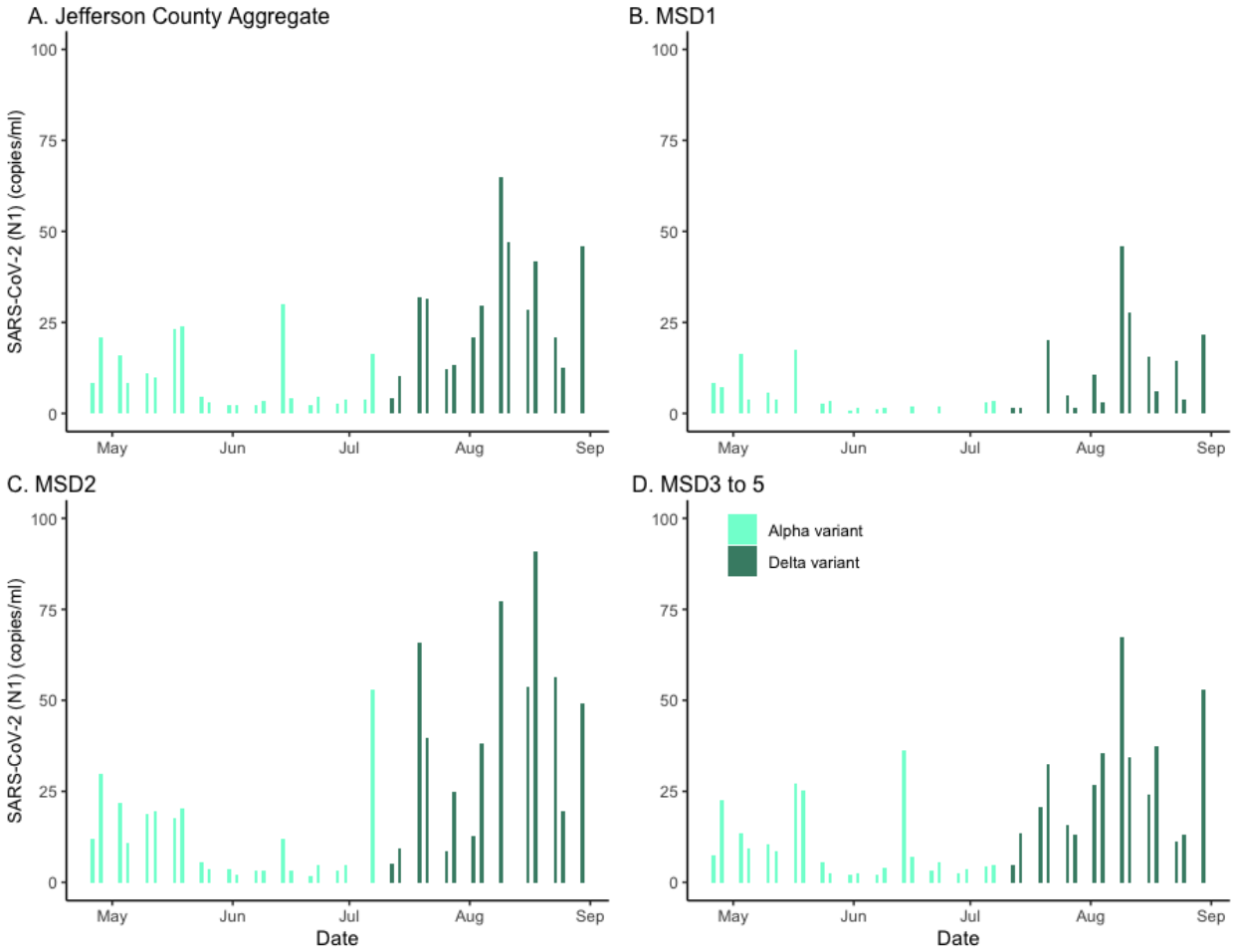

Figure S3.1. **SARS-CoV-2 (N1) wastewater concentration in sewersheds of Jefferson County, KY (USA).** The wastewater concentrations during Alpha and Delta variants are represented in bars (light green for Alpha, dark green for Delta). The panels compare aggregated concentration for Jefferson County (Panel A) as well as stratified by sewershed (Panels B–D).

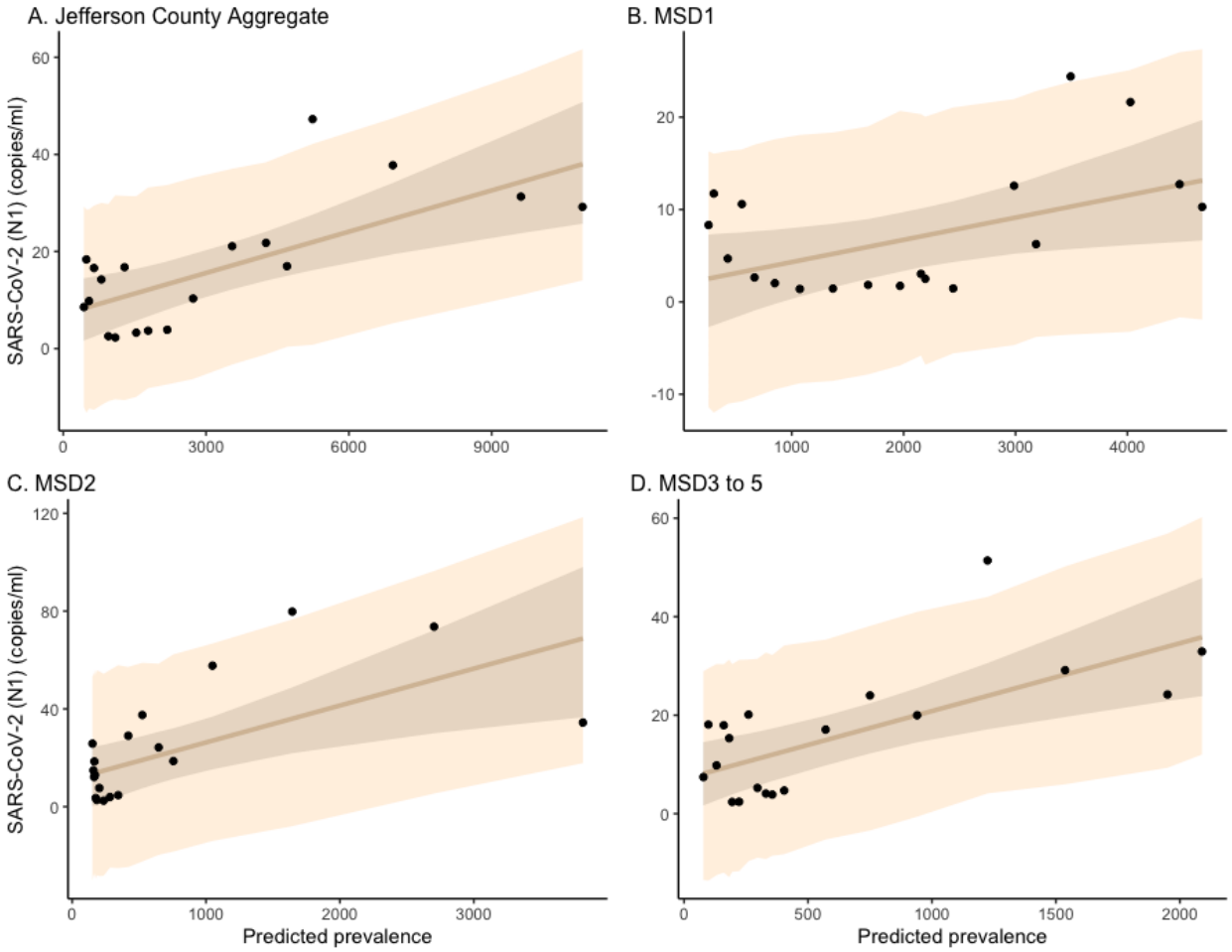

Figure S3.2. **Prevalence versus SARS-CoV-2 (N1) wastewater concentration in sewersheds of Jefferson County, KY (USA).** Bayesian regression between predicted weekly prevalence of SARS-CoV-2 infections and wastewater in the entire Jefferson County (Panel A) as well as stratified by sewershed (Panels B–D). Straight line is the fitted Bayesian regression line. The darker shade marks the 95% credible interval and lighter shade marks the 95% prediction interval.

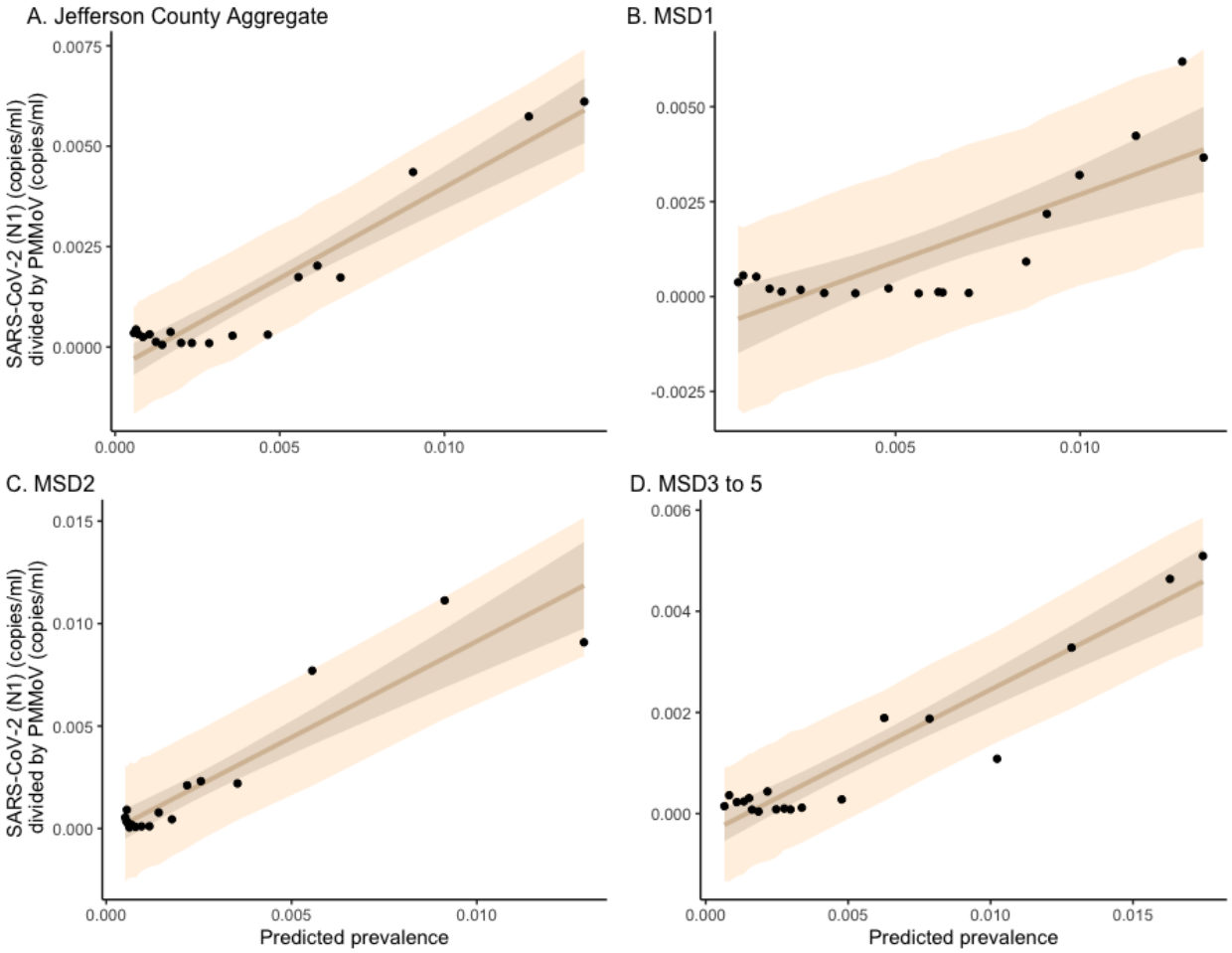

Figure S3.3. **Prevalence versus wastewater SARS-CoV-2 (N1) normalized by pepper mild mottle virus concentration in sewersheds of Jefferson County, KY (USA).** Bayesian regression between predicted weekly prevalence of SARS-CoV-2 infections and wastewater in the entire Jefferson County (Panel A) as well as stratified by sewershed (Panels B–D). Straight line is the fitted Bayesian regression line. The darker shade marks the 95% credible interval and lighter shade marks the 95% prediction interval.

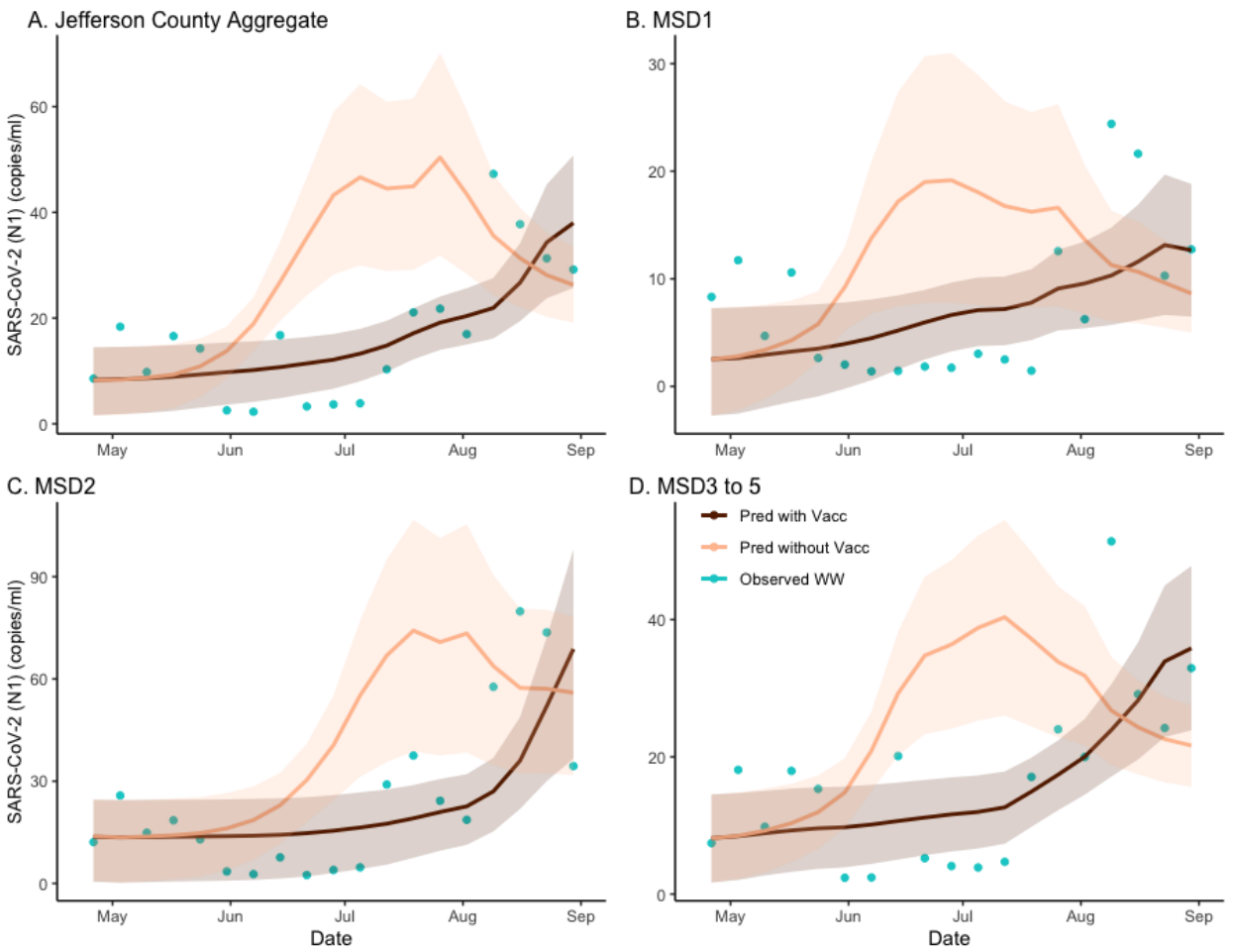

Figure S3.4. **The estimated effect of vaccination on SARS-CoV-2 wastewater concentration in sewersheds of Jefferson County, KY (USA).** The deep brown line is the regression-based fit to the wastewater concentration and the light brown line is the prediction of wastewater concentration using synthetic prevalence from the model with the Delta variant effect zeroed out. The shaded areas represent 95% credible intervals. The blue dots are observed weekly average wastewater concentration. The panels compare the mutation effect on wastewater concentration for Jefferson County (Panel A) as well as stratified by sewershed (Panels B–D).

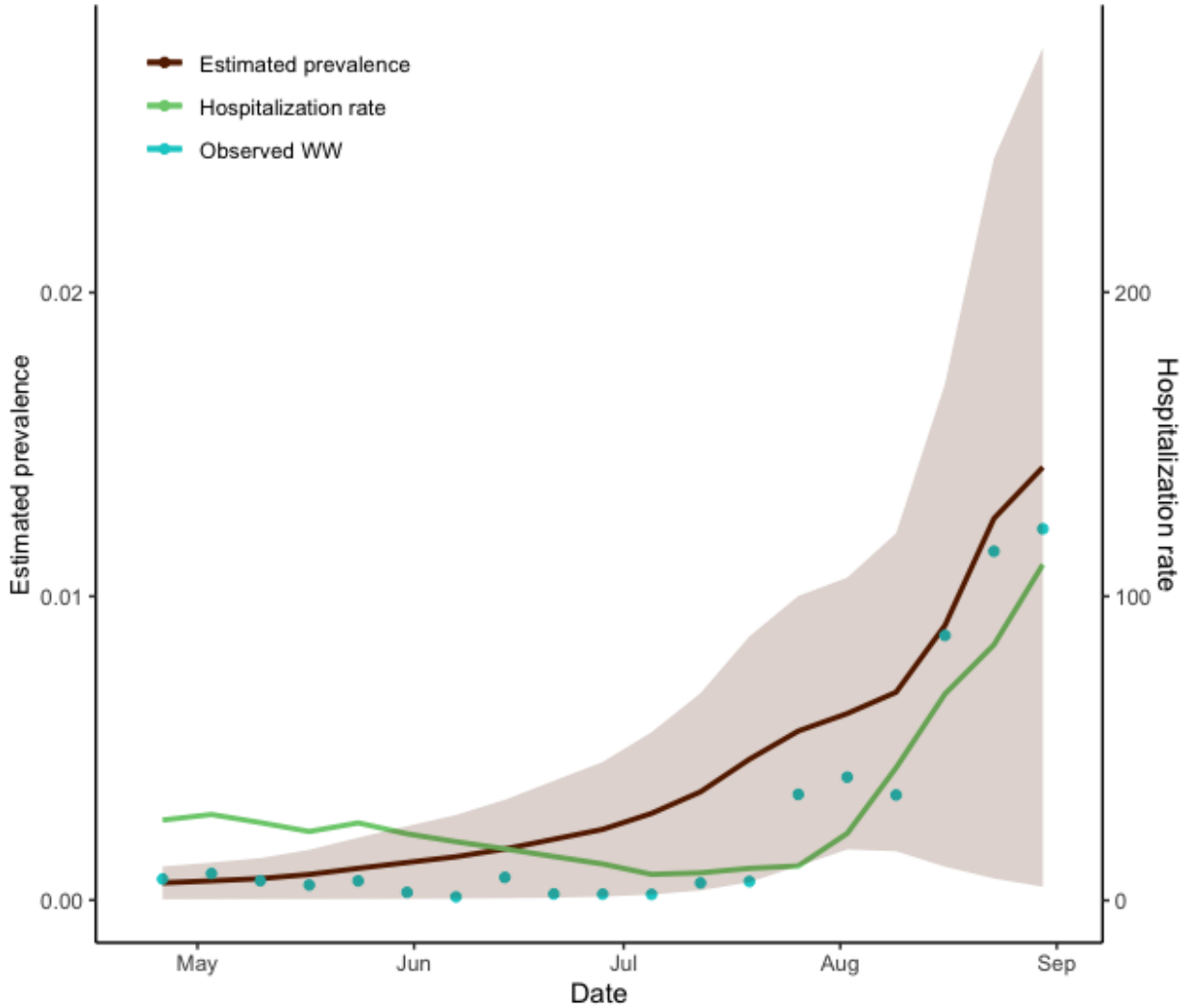

Figure S3.5. **SARS-CoV-2 prevalence and hospitalizations versus SARS-CoV-2 wastewater concentration normalized by pepper mild mottle virus, Jefferson County, KY (USA).** Relationship among observed wastewater concentration, the hospitalization rate, and estimated prevalence. The dark brown line represents the estimated prevalence, and the shaded area is the 95% credible interval of MCMC simulation. The green line is the weekly average of daily hospitalization rate of Jefferson County, and the blue dots represent the weekly average of wastewater concentrations. The Pearson correlation coefficient of estimated prevalence and wastewater concentration is 0·916 (95% CI=(0·764, 0·976)) and that of hospitalization rate and wastewater concentration is 0·720 (95% CI =(0·224, 0·953)).

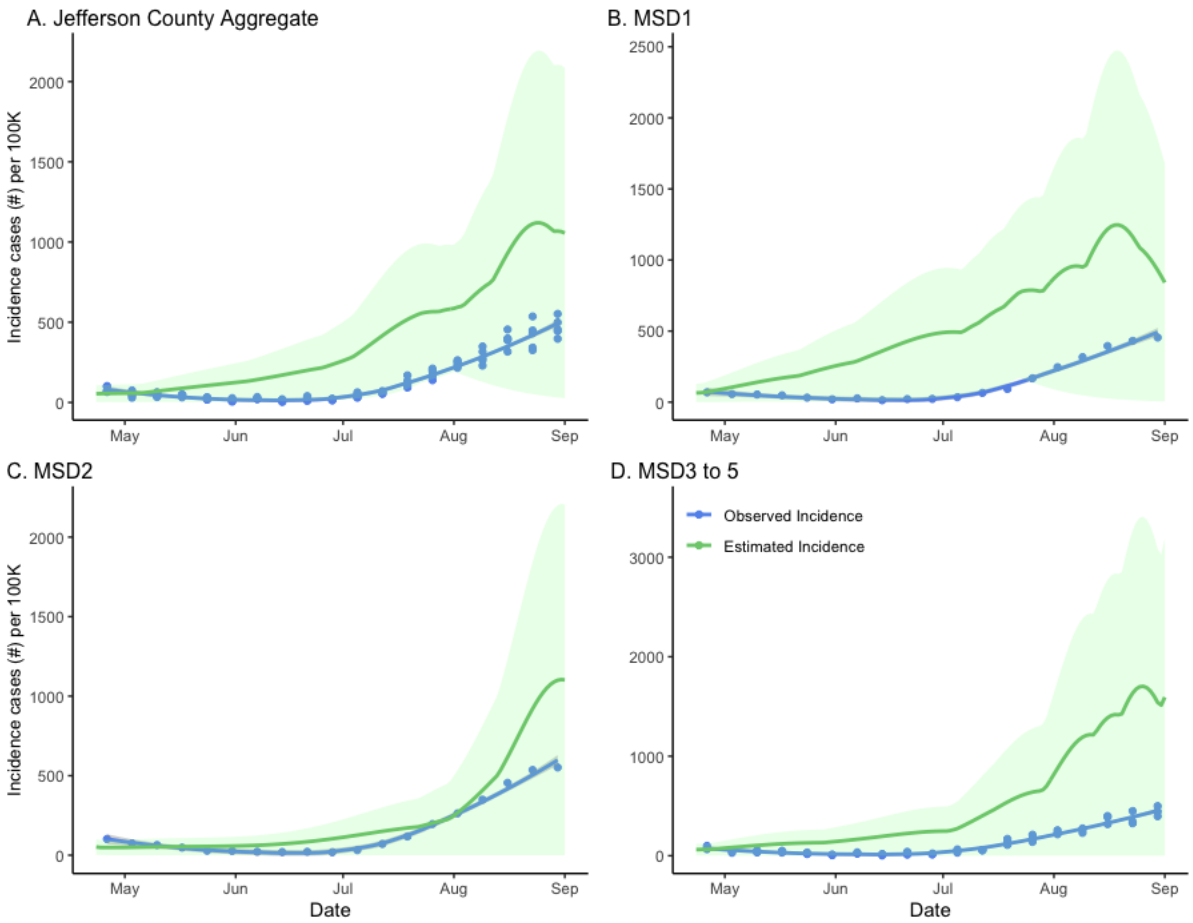

Figure S3.6. **Clinical versus estimated incidence in sewersheds of Jefferson County, KY (USA).** Posterior density and credibility bounds (green curve) of the weekly aggregated incidence rate as predicted by the model compared to official weekly incidence for Jefferson County (blue dots and trend line) as reported by the Jefferson County Health Department. The panels compare aggregated incidence for Jefferson County (Panel A) as well as stratified by sewershed (Panel B–D). The model plots are based on Hamiltonian MCMC samples, with 6000 steps and 2000 steps burn-in period.

S3.2 Details on regression model for wastewater concentration

To relate the SVI_2RT model predictions to the serial wastewater measurements of SARS-CoV-2 concentrations, the Bayesian linear regressions were performed based on aggregated county data and data stratified by sewershed area.

To obtain the linear regressions, the procedure was as follows: Let $\tilde{I}_{tj}$ be the model estimated percentage prevalence corresponding to the same week and sewershed area. W_{tj} was defined in Eq. (5). The linear and NB regression models are given by:

$$W_{tj} = \beta_0 + \beta_1 \tilde{I}_{tj} + e_{tj}, \quad (5)$$

In the Bayesian linear regression models, non-informative priors were assigned. Specifically, the non-informative Cauchy distribution was assigned to the regression coefficients, and the non-informative gamma prior was assigned to the dispersion parameter of the error term. The summary of the posterior estimates of all regression parameters is presented in Table S3.4, and fitting and prediction using the regression model are represented in Figure 2.

S3.3 Time lag-dependency between wastewater concentration and hospitalization rate

It takes a certain period for the patient to be admitted to the hospital to receive treatment. To identify the time lag-dependency between wastewater concentration and hospitalization rate, a simple linear regression analysis was performed using time-lagged variable as a predictor. Let W_{t-d} be the weekly aggregated average wastewater concentration at week t the aggregated Jefferson County, and d represents a time lag. H_t represents the hospitalization rate at time t. The regression model with time lag dependent variable is given by:

$H_t = \beta_0 + \beta_1 W_{t-d} + e_t$. (6)

In this model, we changed the time lag d from 1 to 4 so that the maximum period from a shred of evidence of the community spread of COVID-19 in wastewater to reach a burden to hospitalization is about a month. Of note, hospitalizations data is available daily while wastewater is weekly

Additionally, we performed a simulation study using this regression model how to check how much the hospitalization rate changes according to the vaccination rate. We changed the vaccination rate so that the vaccination percentage of the community was 0% and predicted the serial estimates Pred_t in Eq. (4). And then, we predicted the wastewater concentration using a linear regression model and used them as the predictor in the regression model (5).

S3.4 Calculation of effects based on factual and counter-factual scenarios

All effects of the factual and counterfactual (zero vaccinated or no Delta variant). Data are calculated using the area under the curve using factual (empirical) data and counterfactual (synthetic) data. The equation to estimate the effect is given as:

$$\left|\frac{\text{Area under counterfacutual data}}{\text{Area under factul data}} - 1\right|$$

References

1. Boyle L, Hletko S, Huang J, et al. Selective sweeps in SARS-CoV-2 variant competition. *Proc Natl Acad Sci USA* 2022; **119**(47): e2213879119. doi:10.1073/pnas.2213879119
2. Yang W, Shaman J. Development of a model-inference system for estimating epidemiological characteristics of SARS-CoV-2 variants of concern. *Nat Commun* 2021; **12**.1: 1–9. doi:10.1038/s41467-021-25913-9
3. Smith T, Holm RH, Keith RJ, et al. Quantifying the relationship between sub-population wastewater samples and community-wide SARS-CoV-2 seroprevalence. *Sci Total Environ* 2022; **853**: 158567. doi:10.1016/j.scitotenv.2022.158567
4. KhudaBukhsh WR, Choi B, Kenah E, Rempala GA. Survival dynamical systems: individual-level survival analysis from population-level epidemic models. *Interface focus* 2020; **10**(1): 20190048. doi:10.1098/rsfs.2019.0048

S4. Variant detection

Table S4.1. **Periods of Alpha and Delta variant dominance in in sewersheds of Jefferson County, KY (USA).**

Catchment site	Alpha Dominant		Delta Dominant	
	Begin Date	**End Date**	**Begin Date**	**End Date**
MSD01	3/30/21	5/17/21	7/12/21	8/30/21
MSD02	3/30/21	5/24/21	7/12/21	8/30/21
MSD03	3/30/21	6/21/21	7/19/21	8/30/21
MSD04	3/30/21	7/5/21	7/19/21	8/30/21
MSD05	3/30/21	6/28/21	7/26/21	8/30/21