Prolonged viral shedding from noninfectious individuals confounds wastewater-based epidemiology

Tin Phan[1], Samantha Brozak[2], Bruce Pell[3], Stanca M. Ciupe[4], Ruian Ke[1], Ruy M. Ribeiro[1], Anna Gitter[5,6], Kristina D. Mena[5,6], Alan S. Perelson[1, 7], Yang Kuang[2], and Fuqing Wu[5, 6]*

[1] Theoretical Biology and Biophysics, Los Alamos National Laboratory, NM 87544, USA.

[2] School of Mathematical and Statistical Sciences, Arizona State University, AZ 85281, USA.

[3] Department of Mathematics and Computer Science, Lawrence Technological University, MI 48075, USA.

[4] Department of Mathematics, Virginia Tech, Blacksburg, VA 24060, USA.

[5] School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA.

[6] Texas Epidemic Public Health Institute, TX, USA.

[7] Santa Fe Institute, Santa Fe, NM 87501, USA.

* Corresponding author: Fuqing Wu

Email: Fuqing.Wu@uth.tmc.edu

Keywords: wastewater-based epidemiology, prolonged viral shedding, mathematical modeling, transmission dynamics.

Abstract

Wastewater surveillance has been widely used to track and estimate SARS-CoV-2 incidence. While both infectious and recovered individuals shed virus into wastewater, epidemiological inferences using wastewater often only consider the viral contribution from the former group. Yet, the persistent shedding in the latter group could confound wastewater-based epidemiological inference, especially during the late stage of an outbreak when the recovered population outnumbers the infectious population. To determine the impact of recovered individuals' viral shedding on the utility of wastewater surveillance, we develop a quantitative framework that incorporates population-level viral shedding dynamics, measured viral RNA in wastewater, and an epidemic dynamic model. We find that the viral shedding from the recovered population can become higher than the infectious population after the transmission peak, which leads to a decrease in the correlation between wastewater viral RNA and case report data. Furthermore, the inclusion of recovered individuals' viral shedding into the model predicts earlier transmission dynamics and slower decreasing trends in wastewater viral RNA. The prolonged viral shedding also induces a potential delay in the detection of new variants due to the time needed to generate enough new cases for a significant viral signal in an environment dominated by virus shed by the recovered population. This effect is most prominent toward the end of an outbreak and is greatly affected by both the recovered individuals' shedding rate and shedding duration. Our results suggest that the inclusion of viral shedding from non-infectious recovered individuals into wastewater surveillance research is important for precision epidemiology.

Introduction

SARS-CoV-2 concentration in wastewater is strongly correlated to case report data ($R > 0.7$) (1), supporting the effort of using wastewater-based epidemiology (WBE) to estimate infection incidence (2, 3). However, gastrointestinal tract viral shedding exhibits great heterogeneity among individuals, and a significant portion of clinically recovered individuals (i.e., individuals who are no longer infectious) shed virus in fecal specimens for weeks to months after the infectious phase (4-7). As an outbreak progresses, the number of recovered individuals accumulates, which results in an increase in their contribution to the overall viral RNA load in wastewater. While the prolonged shedding of SARS-CoV-2 in feces is well documented (4-6), to what extent the persistent shedding impacts the epidemiological inference of wastewater data remains unexplored.

Based on recent studies, we propose the integration of wastewater data with within-host (viral shedding dynamics) and between-host (transmission dynamics) models to demultiplex the viral shedding from infectious and recovered phases (8, 9). The central idea is to use a population-level viral shedding profile, segmented according to the infection state, to connect viral RNA in wastewater with standard transmission dynamics models (SI Appendix). Our goal is to examine the potential impact of the virus shed by the recovered population on the capability of WBE to capture the epidemic landscape.

Results and Discussion

Viral shedding by the recovered population alters the correlation of virus measured in wastewater and case report data. We fit a Susceptible-Exposed-Infectious-Recovered-Virus (SEIR-V) model (Eq. 1, SI Appendix) including a population viral shedding function (Eq. 5, SI Appendix) to viral RNA measurements in wastewater collected from the Greater Boston area from 10/01/2020 to 02/27/2021 (1, 3), under the

assumption that individuals who stop shedding virus are removed from the recovered class. The time frame was chosen to limit the data to a single wave of the COVID-19 pandemic (Alpha variant B.1.1.7) when the vaccination rate was low in the population. For a proof-of principle demonstration, we assume that recovered individuals on average shed virus at 50% of the mean viral shedding rate of infectious individuals for 14 days. The model captures the viral dynamics in wastewater well and the inferred transmission dynamics reflect the trend of the reported incidence (Fig. 1A-B). The model-predicted transmission dynamics, however, precede the reported incidence by about 3 weeks (as compared by peak timing – vertical dashed lines in Fig. 1B). The model-inferred incidence exceeds the reported daily incidence by 3-5 times near the beginning of Oct. and at the end of Feb. and over 10 times around Dec. Both observations are consistent with the expected lead time of wastewater data relative to case report data and the underreported incidence rates (10, 11).

Next, we compared the dynamics of the infectious and the recovered populations. As shown in Fig. 1C, the number of recovered individuals increases quickly and exceeds the infectious population. While the infectious population contributes most of the viral RNA in wastewater near the beginning of this epidemic wave, the recovered population eventually surpassed their viral contribution. Although these quantitative results may vary with different assumed viral shedding rates and shedding durations for the recovered population, they show how the viral shedding from the recovered population becomes more significant as an outbreak progresses.

The increase of virus shed by the recovered population creates a transient period where wastewater data may not directly reflect the incidence and dynamics of active infection.

An intuitive approach to validate this inference is by comparing the correlation between wastewater viral measurements and reported incidence using data spanning different phases of an outbreak. During the early phase of an outbreak, wastewater data should be highly correlated with the reported daily new cases since the infectious population contributes substantially more virus to wastewater. However, the correlation should decrease as the outbreak progresses with the increase of the virus shed by the recovered population. We explicitly demonstrated this point by fitting the same model to increasingly larger data sets (Fig. 1D and 1E). We observe a clear decreasing trend in the correlation coefficient between the model-inferred daily incidence and the reported daily incidence with the inclusion of data post-transmission peak. Additionally, the model-simulated correlations (SI Appendix) between daily viral measurement and incidence show that the inclusion of viral shedding by the recovered population predicts a faster decline of the correlation coefficient as the outbreak progresses (Fig. 1F). Note that the increasing trend near the end is a result of the convergence of daily viral measurement in wastewater and incidence at the end of the outbreak. Thus, the exclusion of the virus shed by the recovered population is likely to have a small effect on the model-inferred transmission dynamics in the early phase of an outbreak (Fig. 1D and 1F). However, when using wastewater surveillance to track emerging cases toward the end of an outbreak or to connect multiple waves, it is necessary to include the virus shed by the recovered population to accurately capture transmission.

Inclusion of recovered shedding predicts an earlier peak time of the outbreak. To better explain the mechanisms behind the changes in the correlation between wastewater data and case report data, we tested the impact of varied shedding rates and duration from the recovered individuals on model-inferred transmission dynamics.

Fig. 2A shows the model fitting to viral load in wastewater under three scenarios: 1) no shedding from recovered individuals, 2) HRMD: high recovered shedding rate (75% of the infectious shedding rate) for a moderate duration of 14 days, 3) MRLD: moderate recovered shedding rate (50% of the infectious shedding rate) with a long duration of 21 days. Model fits under the three scenarios are similar, but the higher shedding rates and longer shedding durations result in slightly slower decreasing trends of the total viral load (Fig. 2A). A remarkable difference is observed in the daily incidence. The inclusion of viral shedding by the recovered population shifts the transmission peak by 6- and 10-day earlier under MRLD and HRMD scenarios, respectively (vertical dashed lines, Fig. 2B). These results suggest that the actual incidence peak time of the Alpha variant outbreak may be much earlier than the clinically reported cases and the change in the correlation is due in part to the differences in the viral shedding dynamics between the infectious and the recovered phases.

The virus shed by the recovered population leads to a delay in the detection of an emerging variant by WBE. Several studies have looked at the minimum number of new cases required for the detection of virus in wastewater (12-14); however, none has looked into this issue in the presence of uncertainty due to the viral shedding from the recovered population. This is especially important during a transition period when a new variant is emerging near the end of an ongoing outbreak (Fig. 2C). The viral shedding from a large number of recovered individuals in wastewater may overshadow the viral shedding from new cases infected with the new variant, therefore undermining a rapid recognition of the emergence of the new variant based on the viral load in wastewater. We used the SEIR-V model to further quantify the relationship between shedding from recovered population and shedding from individuals with the new variant. Assuming the

emerging variant has a similar viral shedding rate to its predecessor, we found that the number of required infectious cases with the new variant to signal an emerging outbreak is an increasing function of both the viral shedding rate and shedding duration of the recovered population (Fig. 2D, warmer colors). Since it takes time to generate the required number of infections for detection, this implies an increasing delay in the detection of an emerging outbreak as more new infections are needed. This delay is minimized if the viral shedding rate of the new variant is significantly higher than its predecessor (data not shown). Thus, tracking the viral shedding from the recovered population is important to mitigate the delay in the detection of new variants based on wastewater data.

The identification and mitigation of this time delay requires an accurate description of the population viral shedding profile, which may be variant-specific. Thus, integrating modeling analysis of routine wastewater surveillance data and genomic sequencing of wastewater samples (15) may better facilitate the identification of emerging variants circulating in the community. While our analysis focuses on the end of an outbreak to highlight the impact of prolonged shedding, similar analyses can be carried out to study the detection delay at different times during the outbreak and may need to account for the viral shedding from both the infectious and recovered population if they are not negligible.

In summary, we show that an understanding of the relative viral shedding during each phase of an infection is crucial to the implementation of precision WBE, which transpires beyond the study of the correlation between viral measurement in wastewater and case report data. Prolonged shedding can alter model-inferred transmission dynamics and

lead to a delay in detection if not accounted for. Our analyses provide a proof-of-principle framework to quantify the impact of prolonged fecal shedding on wastewater-based estimation of new infections and the outbreak progression.

Materials and Methods

See SI Appendix.

Data and Code Availability

The authors declare that the data and code will be provided with the publication of this study.

Author Contributions

T.P., R.K., R.M.R., Y.K., and F.W. designed research; T.P., S.M., B.P., and F.W. performed research; T.P., S.M., R.K., R.M.R., K.D.M., A.S.P., Y.K., F.W. contributed new reagents/analytic tools; T.P., S.M., B.P., S.M.C., R.K., R.M.R., A.G., and F.W. analyzed data; and T.P., S.M., B.P., S.M.C., R.K., R.M.R., A.G., K.D.M., A.S.P., Y.K., and F.W. wrote the paper.

Acknowledgments

This work is supported by Faculty Startup funding from the Center of Infectious Diseases at UTHealth, the UT system Rising STARs award, and the Texas Epidemic Public Health Institute (TEPHI) to F.W. This work was also supported by the Director's postdoctoral fellowship at Los Alamos National Laboratory to T.P.; Y.K. and S.B. are partially supported by the US National Science Foundation Rules of Life program DEB -1930728 and the NIH grant 5R01GM131405-02.

Competing Interest Statement

The authors declare no competing interest.

References

1. F. Wu, et al., SARS-CoV-2 RNA concentrations in wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases. *Sci. Total Environ.* 805, 150121 (2022).
2. P. Kilaru, et al., Wastewater surveillance for infectious disease: a systematic review. *Am. J. Epidemiol.* 192(2), 305-322 (2023).
3. A. Xiao, et al., Metrics to relate COVID-19 wastewater data to clinical testing dynamics. *Water Res.* 212, 118070 (2022).
4. S. Gupta, J. Parker, S. Smits, J. Underwood, S. Dolwani, Persistent viral shedding of SARS-CoV-2 in faeces–a rapid review. *Colorectal Dis.* 22(6), 611-20 (2020).
5. M. Cevik, et al., SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *The Lancet Microbe*. 2(1), e13-22 (2021).
6. A. Natarajan, et al., Gastrointestinal symptoms and fecal shedding of SARS-CoV-2 RNA suggest prolonged gastrointestinal infection. *Med* 3(6), 371-387 (2022).
7. B. Killingley, et al., Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults. *Nature Med.* 28(5), 1031-1041 (2022).

8. T. Phan, et al., A simple SEIR-V model to estimate COVID-19 prevalence and predict SARS-CoV-2 transmission using wastewater-based surveillance data. *Sci. Total Environ.* 857, 159326 (2023).
9. B. Pell, S. Brozak, T. Phan, F. Wu, Y. Kuang, The emergence of a virus variant: dynamics of a competition model with cross-immunity time-delay validated by wastewater surveillance data for COVID-19. *J. Math. Biol.* 86(5), 1-34 (2023).
10. F. Wu, et al., SARS-CoV-2 titers in wastewater are higher than expected from clinically confirmed cases. *mSystems* 5(4), e00614-20 (2020).
11. F. J. Angulo, L. Finelli, D. L. Swerdlow, Estimation of US SARS-CoV-2 infections, symptomatic infections, hospitalizations, and deaths using seroprevalence surveys. *JAMA Netw. Open* 4(1), e2033706-e2033706 (2021).
12. J. Hewitt, et al., Sensitivity of wastewater-based epidemiology for detection of SARS-CoV-2 RNA in a low prevalence setting. *Water Res.* 211, 118032 (2022).
13. P. Y. Hong, et al., Estimating the minimum number of SARS-CoV-2 infected cases needed to detect viral RNA in wastewater: To what extent of the outbreak can surveillance of wastewater tell us? *Environ. Res.* 195, 110748 (2021).
14. M. Nauta, et al., Early detection of local SARS-CoV-2 outbreaks by wastewater surveillance: a feasibility study. *Epidemiol. Infect.* 151, e28 (2023).
15. S. Karthikeyan, et al., Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. *Nature* 609, 101-108 (2022).

Figures

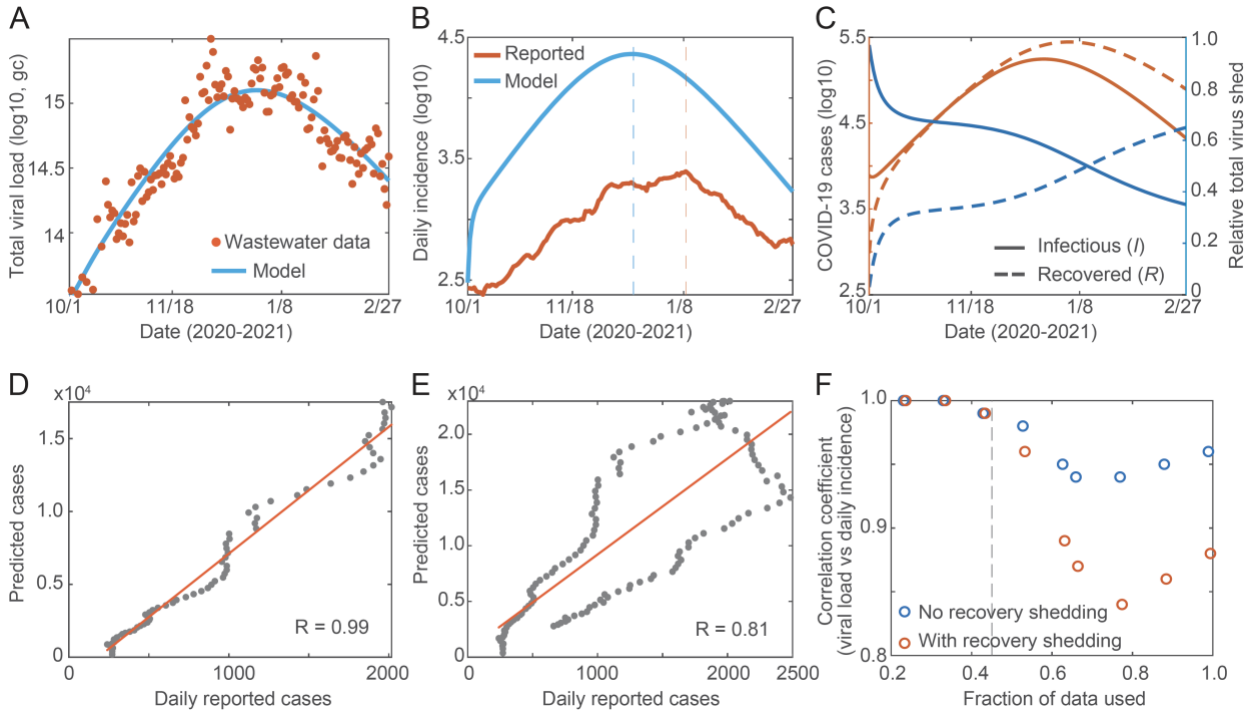

Figure 1. **Viral shedding from the recovered population alters the correlation of virus measured in wastewater and case report data.** Predicted viral dynamics obtained from fitting model Eq. 1 to viral shedding data of both infectious and recovered individuals. (**A**) Theoretical viral dynamics (blue curve) obtained by fitting the daily production of V in Eq. 1 to daily measurements of viral RNA (brown dots) in wastewater for the Greater Boston area. (**B**) Model-predicted daily incidence (blue line) vs. reported incidence (brown curve). The vertical dashed lines are the incidence peak time. (**C**) Left y-axis: model predicted number of infectious (solid brown curve) and recovered individuals (dotted brown curve). Right y-axis: relative contribution to the viral RNA measurements from the infectious (solid blue curve) and recovered (dotted blue curve) populations. (**D-E**) The correlation coefficient between predicted cases and reported

cases decreases as more data are used for the fitting (within one wave). D only uses pre-peak data. E contains increasingly more data after the peak. (**F**) The model-predicted correlation between daily viral measurements in wastewater and daily incidence with and without the recovered shedding in the model. If the fraction of data used is 1, the data for one entire outbreak is used for the correlation. The vertical dashed line indicates the transmission peak.

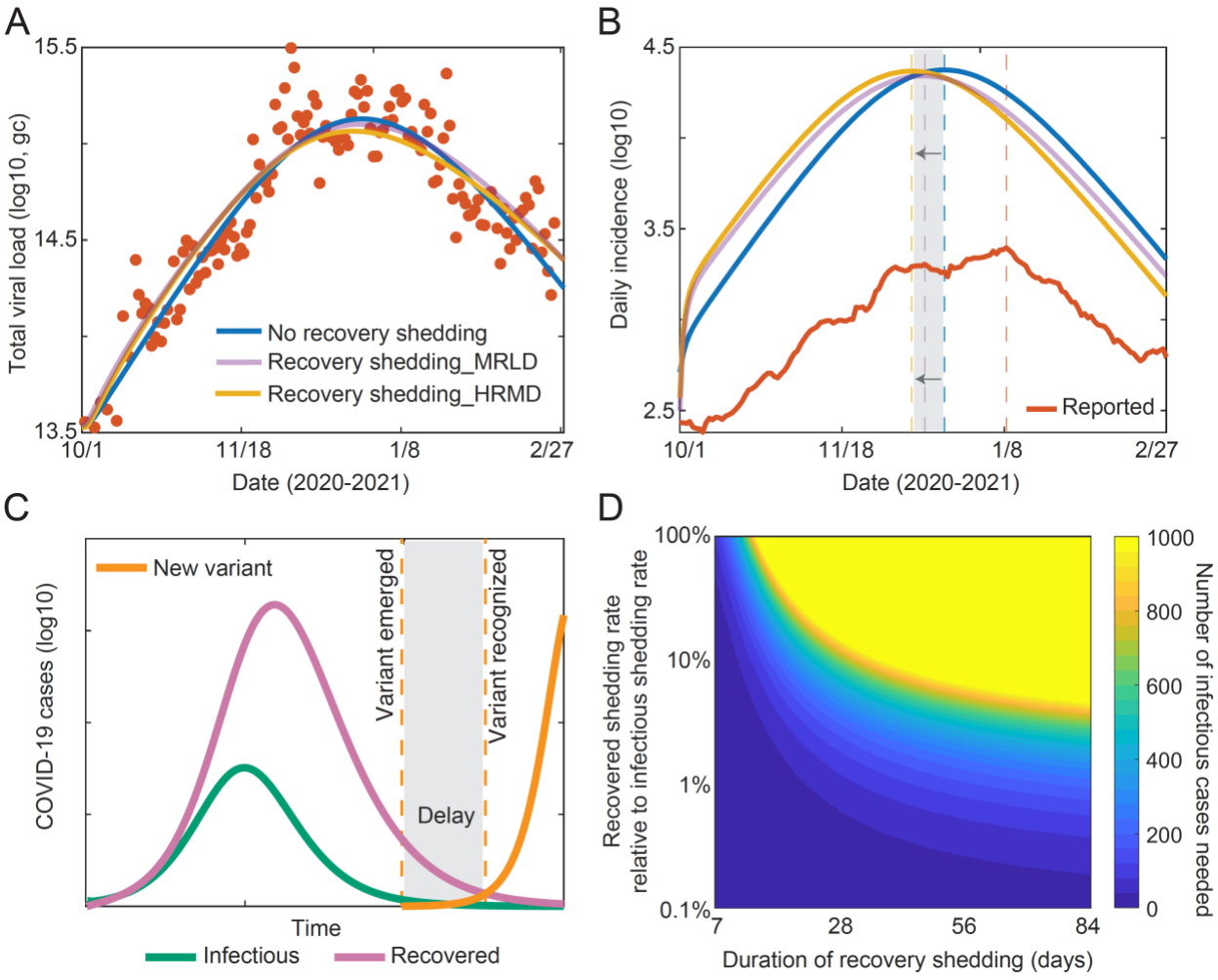

Figure 2. The inclusion of recovered shedding predicts an earlier peak time of the outbreak and **a delay in the detection of an emerging variant by WBE.** (**A**) Viral RNA in wastewater under different recovered shedding scenarios. Each model was fit to data from 10-01-2020 to 02-27-2021, which was chosen to restrict the outbreak to a single variant (brown data points). (**B**) Model inferred daily incidence under three recovered shedding scenarios. The vertical dashed curves are the corresponding peak times. (**C**) A schematic of tracking the emergence of a new variant near the end of an outbreak when most of the virus measured comes from the recovered population. (**D**) The number of new infectious cases needed to produce a sufficiently large viral signal, which depends

259 on the recovered shedding rate and duration (Eqs. 9-11, SI Appendix). Warmer colors
260 refer to a higher number of cases. This snapshot was taken when the outbreak reached
261 99.9% of its final size (as depicted by **C**) to show the potential of WBE to capture an
262 emerging outbreak.

Supplementary Materials

Prolonged viral shedding from noninfectious individuals confounds wastewater-based epidemiology

Tin Phan[1], Samantha Brozak[2], Bruce Pell[3], Stanca M. Ciupe[4], Ruian Ke[1], Ruy M. Ribeiro[1], Anna Gitter[5], Kristina D. Mena[5,6], Alan S. Perelson[1, 7], Yang Kuang[2], and Fuqing Wu[5, 6,*].

*Corresponding author: Fuqing Wu

Email: Fuqing.Wu@uth.tmc.edu

This PDF file includes:

Methods

S1. SEIR-V model with temperature variation.

We modified the SEIR-V model in Phan et al. (1) to include viral shedding from both the infectious and recovered populations. We also initially examined the viral shedding from the exposed class, which turned out to be negligible, due to the low shedding rate and short duration of the exposed class (not shown). Its implications are discussed in S6.

Equation 1

$$S' = -\lambda IS$$

$$E' = \lambda IS - kE$$

$$I' = kE - \delta I$$

$$R' = (1-\epsilon)\delta I - \sigma R$$

$$V' = \alpha\big(1-\gamma(T)\big)(\beta_I I + \beta_R R).$$

S is the susceptible population, E is the exposed population, I is the infectious population, and R is the recovered population that still sheds virus. Susceptible individuals are infected at a rate λI. Exposed individuals take on average $1/k$ days before becoming infectious. Infectious individuals recover or die after an average duration of $1/\delta$ days. The disease-induced mortality rate is ϵ. However, since SARS-CoV-2-induced mortality is small, we set $\epsilon = 0$ for simplicity. Recovered individuals shed virus for an average duration of $1/\sigma$ days. The cumulative viral RNA in wastewater (V) is contributed by both infectious and recovered individuals in this model at rates β_I and β_R, respectively, as opposed to the model proposed by Phan et al. (1). α is the average fecal load per day. $\gamma(T)$ is the fraction of virus lost in wastewater between shedding and collection (for measurement), which is assumed to be temperature-dependent (T). The temperature variation over the duration of the study is fitted to average daily temperature data (1) in Celsius.

Equation 2

$$T(t) = 3.6249\sin(0.0202t - 4.4665) + 16.2298.$$

Where t is measured in days and $t = 0$ is 10/01/2020, see Fig. S1 in (1). The temperature-adjusted half-life is assumed to be an exponential decay of the form:

Equation 3

$$\eta(T) = \eta_0 Q_{10}^{-(T(t)-T_0)/10},$$

with η_0 being the half-life (hours) at ambient temperature T_0 and Q_{10} $(= 2.5)$ being the temperature-dependent rate of change (1, 2). The temperature-dependent decay rate $\xi(T)$ is approximated by a first-order decay function, which gives: $\xi(T) = \frac{\ln 2}{\eta(T)}$. Assume viral RNA decays exponentially $\left(V' = -\xi(T)V \text{ or } V(t) = V_0 e^{-\xi(T)t}\right)$, then the fraction of virus decay between shedding and collection is given by:

Equation 4

$$\gamma(T) = \frac{V_0 - V(t_{arrive})}{V_0} = 1 - e^{-\xi(T)t_{arrive}}.$$

Here, t_{arrive} is the time between shedding and collection, which we take to be 24 hours. The loss fraction γ is approximated between 0.03-0.28 for varying travel times and temperature, see SI in (1). However, this loss fraction could also be a function of the biochemical composition of the wastewater and microorganism community in the sewer. For example, an extremely harsh environment may result in a shorter viral half-life, which implies that the sample may be mostly influenced by viruses shed closer to the time of sampling.

Finally, we remark that $V(t)$ is a proxy variable for the cumulative virus shed in wastewater. Its daily difference ΔV_t, which represents the virus measured on day t is the quantity of interest, e.g.,

$\Delta V_t = V(t) - V(t-1)$. Similarly, we estimate the daily incidence by looking at the cumulative number of infectious individuals. This is done by creating another proxy variable $C(t)$ with $C'(t) = kE$ and taking the daily difference ΔC_t to be the daily incidence on day t, e.g., $\Delta C_t = C(t) - C(t-1)$. Note that while ΔC_t is a good approximation, it does not perfectly correspond to the daily report data, which varies based on many factors such as report time and symptom onset.

S2. The population level viral shedding function.

The viral dynamics within an infected individual is relatively well defined (3, 4) and can be described by standard models of viral dynamics (5). Here, to simplify, we describe the viral levels in the gastrointestinal tract, which is proportional to the viral shedding rate, using a phenomenological function introduced by Phan et al. (1):

Equation 5

$$f(t) = \frac{\omega_1 t}{\omega_2^2 + t^2},$$

which peaks at $\frac{\omega_1}{2\omega_2}$ when $t = \omega_2$. Here, ω_1 (log10 viral RNA copy per g) is a magnitude modifier and ω_2 (day) influences the timing and magnitude of the viral shedding peak. Note that the viral shedding rate represents the viral concentration (per g of fecal matter) per day. This function has been shown to be a good approximation for the viral shedding dynamics (1). With this function, the total shedding between times t_1 and t_2 is given by

Equation 6

$$\int_{t_1}^{t_2} f(t)\, dt = \frac{\omega_1}{2} \log\left(\frac{\omega_2^2 + t_2^2}{\omega_2^2 + t_1^2}\right)$$

Where $0 \leq t_1 < t_2$ and log denotes the natural logarithm. We remark that the unit of $f(t)$ is log10 viral RNA copy per g per day, meaning the integral $\int f(t)dt$ has unit of log10 viral RNA copy per g.

The average exposed E, infectious I, and recovered with shedding R phases of an infection are classified based on the time since infection (S1). Thus, we can define $t_E, t_I,$ and t_R (or $1/k$, $1/\delta$, and $1/\sigma$) to be the average time an infected person spent in these infection phases, respectively. The average viral shedding rate during the infectious phase β_I can be determined by calculating the average viral shedding over the infectious duration:

Equation 7

$$\beta_I = \frac{1}{t_I}\int_{t_E}^{t_E+t_I} f(t)\,dt = \frac{1}{t_I}\frac{\omega_1}{2}\log\left(\frac{\omega_2^2 + (t_E+t_I)^2}{\omega_2^2 + t_E^2}\right)$$

Note that the unit of β_I is log10 viral RNA copy per g per day. Based on the best-fit to the viral shedding data (1), we take $\omega_1 = 71.97$ log10 viral RNA copy per g and $\omega_2 = 4$ days. The estimated value of β_I with $t_E = 3$ days and $t_I = 8$ days is 7.65 log10 viral RNA copy per g per day. The viral shedding rate for the recovered population β_R is assumed to be a fraction of β_I. For instance, the example in Fig. 1 assumes $\beta_R = 0.5 \times \beta_I$ and $t_R = 14$ days.

S3. Data fitting.

When fitting the SEIR-V model to the viral RNA data in wastewater, we minimize the difference (SSE) between the measurement collected every 24-hour and the total virus produced at each time data point. Here, $\hat{V}(t_d)$ is the total virus (e.g., viral RNA concentration × total flow) measured on day t_d, where t_d is the time of the measurement. $V(t)$, as given by Eq. 1, is a proxy of the

cumulative viral RNA in wastewater. Thus, the viral RNA produced daily, given by $\int_{t_d-1}^{t_d} V'(s)ds$, or ΔV_t, is the comparable quantity to the viral RNA measured every 24-hour.

Equation 8

$$SSE = \sum_{t_d} \left(\log \int_{t_d-1}^{t_d} V'(s)ds - \log \hat{V}(t_d) \right)^2.$$

The computational minimization procedure is done using the function $fmincon$ and $multistart$ (500 random initial guesses) in MATLAB.

The best-fit in Fig. 1A gives $\lambda = 9.36 \times 10^{-8}$ per day per person, $\alpha = 126$ g, and $E(0) = 145$ people. Note that the initial for the infected population is approximated by $I(0) = \frac{V(0)}{\alpha\beta(1-\gamma)}$, where $V(0)$ is taken as the first viral RNA measurement and $E(0)$ and $R(0)$ are assumed to be 0. The parameters k and δ are fixed to 1/3 and 1/8 per day, respectively (3), and σ is fixed to 1/14 per day for the scenario in Fig. 1.

S4. Theoretical correlation between daily viral measurement and incidence with and without recovered shedding.

We use the model to demonstrate how recovered shedding may affect the correlation between viral measurements and incidence. The methods that are often used to relate daily viral measurements to incidence data vary in literature (6, 7). However, here, we will simply look at the correlation of daily viral measurement (ΔV_t) and incidence (ΔC_t). To produce Fig. 1I, we use the best fit parameters for Fig. 1A (given in S3). The model without recovered shedding has $\beta_R = 0$.

S5. Simulation study of the impact of recovered shedding on the capability of WBE.

One aspect of our study is to understand how the population average viral shedding profile affects the sensitivity of wastewater surveillance data to track new cases over the course of an outbreak. Since the viral shedding function $f(t)$ is fixed in structure and was fitted to only one data set, it may not be able to fully capture the true dynamics of viral shedding. Thus, we want to examine how variations in β_R and σ (or $1/t_R$) affect the model inferred transmission dynamics. To focus our analysis on the impact of viral shedding from the recovered individuals, we vary only the recovered shedding rate β_R and the duration σ while keeping all other parameters in the model (Eq. 1) fixed as follow: $k = 1/3$ per day, $\delta = 1/8$ per day, $\lambda = 4 \times 10^{-7}$ per person per day, $\alpha = 125$ g, $\gamma = 0.1$, $\beta_I = 7.65$ log10 viral RNA copy per g per day, $S(0) = 1 \times 10^6$ people, $E(0) = 1$ person, $I(0) = 1$ person, and $R(0) = 1$ person.

The net virus shed by the R class varies over the course of an outbreak. Here, we limit the analysis to the end of an outbreak, where virus shed by the recovered population contributes close to 100% of viral measurements. In this scenario, to detect a significant change in the virus in wastewater, we need enough new infections (or infectious cases) that can generate at least 5% (assumed) of the viral shed by the recovered population on day t_d denoted by

Equation 9

$$C_R(t_d) = \int_{t_d-1}^{t_d} \alpha(1-\gamma)\beta_R R(t)dt.$$

Let $N(t_d)$ be the minimum number of new infections needed to match the viral signal from the recovered population, then $N(t_d)$ must satisfy:

Equation 10

$$N(t_d)\alpha(1-\gamma)\beta_I > 0.05\ C_R(t_d).$$

Solving for N, we have

Equation 11

$$N(t_d) > \frac{0.05\, C_R(t_d)}{\alpha(1-\gamma)\beta_I}.$$

For reporting purposes (Fig. 2C-D), we round up N to the nearest integer. The heat map (Fig. 2D) is generated by calculating N for the average recovered shedding rate β_R varying between 0.001 to 1 of β_I and the average recovered shedding duration σ varying between $1/7$ to $1/84$ per day (or an average recovered shedding duration from 1 week to 12 weeks). The relative time (99.9%) of the outbreak is defined as the time when the cumulative recovered population given by $\int_0^t \delta I(s)ds$ reaches 99.9% of its final size $\left(\lim_{t\to\infty} \int_0^t \delta I(s)ds\right)$.

S6. Further considerations.

Our analyses ignore an intrinsic detection delay with WBE due to the low viral shedding of the exposed phase. Exposed individuals likely shed virus at a very low rate compared to infectious and recovered individuals. Thus, it is logical to assume that WBE is unlikely to detect the virus shed by an exposed individual during an ongoing outbreak, leading to an intrinsic detection delay of new cases using WBE. This detection delay may not be longer than 5 days, which is the average time from infection to symptom onset, signaling high respiratory tract viral load (3). This issue could be studied in detail if an accurate population level viral shedding dynamics is available. However, to our knowledge, there has not been reported data on virus shed from the gastrointestinal tract during the exposed phase of infection. Thus, even the population level viral shedding function $f(t)$ used here may not accurately reflect the early viral shedding dynamics. Furthermore, positive detection of SARS-CoV-2 virus was possible in only about 50% of samples from infected individuals (8, 9), which suggests the possibility of gastrointestinal tract infection being an inherent stochastic event. These important factors, among others, were not considered

in our study due to limited data but should be analyzed further to better understand and improve the utility of WBE.

SI References

1. T. Phan, et al., A simple SEIR-V model to estimate COVID-19 prevalence and predict SARS-CoV-2 transmission using wastewater-based surveillance data. *Sci. Total Environ.* 857, 159326 (2023).
2. O. E. Hart, R. U. Halden, Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and challenges. *Sci. Total Environ.* 730, 138875 (2020).
3. B. Killingley, et al., Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults. *Nature Med.* 28(5), 1031-1041 (2022).
4. R. Ke, et al., Daily longitudinal sampling of SARS-CoV-2 infection reveals substantial heterogeneity in infectiousness. *Nat. Microbiol.* 7(5), 640-652 2022..
5. R. Ke, C. Zitzmann, D. D. Ho, R. M. Ribeiro, A. S. Perelson, In vivo kinetics of SARS-CoV-2 infection and its relationship with a person's infectiousness. *Proc. Natl. Acad. Sci. U.S.A.* 118(49), e2111477118 (2021).
6. F. Wu, et al., SARS-CoV-2 RNA concentrations in wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases. *Sci. Total Environ.* 805, 150121 (2022).
7. A. Xiao, et al., Metrics to relate COVID-19 wastewater data to clinical testing dynamics. *Water Res.* 212, 118070 (2022).
8. S. Gupta, J. Parker, S. Smits, J. Underwood, S. Dolwani, Persistent viral shedding of SARS-CoV-2 in faeces–a rapid review. *Colorectal Dis.* 22(6), 611-20 (2020).
9. A. Natarajan, et al., Gastrointestinal symptoms and fecal shedding of SARS-CoV-2 RNA suggest prolonged gastrointestinal infection. *Med* 3(6), 371-387 (2022).