Estimated transmission dynamics of SARS-CoV-2 variants from wastewater are robust to differential shedding

David Dreifuss[1,2], Jana S. Huisman[1,2,3,*], Johannes C. Rusch[4], Lea Caduff[4], Pravin Ganesanandamoorthy[4], Alexander J. Devaux[4], Charles Gan[4], Tanja Stadler[1,2], Tamar Kohn[5], Christoph Ort[4], Niko Beerenwinkel[1,2,*], Timothy R. Julian[4,6,7,*]

[1]Department of Biosystems Science and Engineering, ETH Zurich, CH-4058 Basel, Switzerland;
[2]SIB Swiss Institute of Bioinformatics, CH-1015 Lausanne, Switzerland;
[3]Physics of Living Systems, Massachusetts Institute of Technology, Cambridge, MA, United States of America;
[4]Eawag, Swiss Federal Institute of Aquatic Science and Technology, CH-8600 Dübendorf, Switzerland;
[5]Laboratory of Environmental Chemistry, School of Architecture, Civil and Environmental Engineering, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland;
[6]Swiss Tropical and Public Health Institute, CH-4051 Basel, Switzerland;
[7]University of Basel, CH-4055 Basel, Switzerland

*Corresponding Authors

Abstract

The COVID-19 pandemic has accelerated the development and adoption of wastewater-based epidemiology. Wastewater samples can provide genomic information for detecting and assessing the spread of SARS-CoV-2 variants in communities and for estimating important epidemiological parameters such as the growth advantage of the variant. However, despite demonstrated successes, epidemiological data derived from wastewater suffers from potential biases. Of particular concern are differential shedding profiles that different variants of concern exhibit, because they can shift the relationship between viral loads in wastewater and prevalence estimates derived from clinical cases. Using mathematical modeling, simulations, and Swiss surveillance data, we demonstrate that this bias does not affect estimation of the growth advantage of the variant and has only a limited and transient impact on estimates of the effective reproduction number. Thus, population-level epidemiological parameters derived from wastewater maintain their advantages over traditional clinical-derived estimates, even in the presence of differential shedding among variants.

Main

In the context of the COVID-19 pandemic, wastewater-based surveillance has become widespread and has proved to be a reliable data source to inform disease trajectories[1]. The viral loads of SARS-CoV-2 in wastewater have been shown to have a strong positive correlation with COVID-19 incidence in the region connected to the studied sewershed, and they are routinely

used as an epidemiological indicator. Tracking the dynamics of SARS-CoV-2 RNA in wastewater can reliably estimate the effective reproductive number (*Re*)[2], a critical parameter describing disease dynamics and informing public health policy. Further, wastewater samples enable the detection and tracking of the progression of genomic variants using PCR-based methods[3], including quantitative and digital PCR (dPCR)[4], as well as next-generation sequencing (NGS) methods[5–7]. Both approaches are able to reliably quantify the relative abundances of genomic variants, with estimates coinciding with those obtained from traditional methods based on clinical samples[4–7]. Wastewater genomics allows early detection of variants, as shown, for example, by the detection of the first confirmed case of Omicron BA.1 in Switzerland in wastewater[8]. For epidemiology, wastewater samples provide accurate population-level data at largely reduced costs and simplified logistics. Therefore, wastewater-derived data is increasingly incorporated into national surveillance strategies.

Tracking the relative abundance of a newly introduced genomic variant over time provides insights into its dynamics and growth advantage over the current dominant strain, which is a key parameter for models predicting future disease trajectories. Early, accurate estimates of this parameter are crucial as they can inform on the epidemic potential of a new variant before an increase in case numbers is observed[9]. The growth advantage informs about the increase in transmissibility and immune evasion of a variant[10]. Generally, it is assumed that a constant multiplicative growth advantage or disadvantage holds for a newly introduced variant, which in the case of an advantage results in logistic growth of its relative abundance[11]. It has been shown that both NGS and dPCR analysis of wastewater samples can provide timely and accurate estimates of the growth advantage of a variant, requiring orders of magnitude fewer samples than traditional surveillance methods relying on samples of infected individuals[4,5].

However, wastewater-based quantification of SARS-CoV-2 and its genomic variants is influenced by shedding load profiles, which describe the average amount of viral RNA each infected individual contributes to the sewer during the course of infection[2]. Although some rare prolonged low levels of fecal shedding have been reported[12], shedding decay throughout the course of infection is believed to be fast enough so that wastewater concentration is indicative of COVID-19 incidence rather than prevalence[13]. The different variants of concern that have emerged and spread since the beginning of the pandemic have all exhibited a difference in average viral loads in the respiratory tract, but the overall shape of the shedding profile was conserved[14]. For the Omicron BA.1 variant of concern, lower fecal shedding compared to Delta[15] has been suspected, possibly due to changes in tropism[16]. Such differences in shedding profiles can decouple the established relationship between estimates of incidence based on wastewater from those derived from clinical test data. As a consequence, population-level incidence based on historical relationships between SARS-CoV-2 concentrations in wastewater and case data will be underestimated (if shedding from the new variant is lower) or overestimated (if shedding from the new variant is higher). To enable accurate wastewater-based estimates of disease trajectory for all variants, including new and uncharacterized ones, analytical approaches need to be robust to shifts in shedding load profiles. Such methods will aid perennial integration of wastewater-derived epidemiological data into public health policy.

In this study, we derived, for a class of common wastewater-derived estimates of variant growth advantage and *Re* based on both NGS and dPCR, mathematical expressions of their bias under shifts in shedding load profiles. The mathematical derivations are supported by simulations, as well as by analysis of sequencing and dPCR data collected during the introduction and transmission of Omicron BA.1 in Switzerland. We demonstrate the invariance of growth advantage estimates under differential shedding and show that wastewater-derived *Re* estimates are affected only transiently and with a limited magnitude.

Results

Growth advantage estimates are invariant to differences in shedding

The growth advantage of an emerging variant is commonly estimated by fitting a logistic growth curve to the time series of variant relative abundance. For wastewater-derived estimates, we derived a closed-form expression for the bias in the estimate of growth advantage for the case in which the variant is shed at a different rate and demonstrated that it is zero (Figure 1 A, B, C). However, the estimate of the midpoint of the logistic growth curve has an additive bias of $-log(c)/a$, where a is the growth rate of the curve and c is the rate at which the variant is differentially shed (Figure 1 C, Figure 2 A).

Differences in shedding have a transient and limited effect on *Re* estimates

We next focused on the effective reproduction number (*Re*). We derived a closed-form expression to describe the bias in *Re* that arises from a lower (or higher) shedding of the variant and demonstrated that this bias term can be significantly different from zero, but only for a brief period. Assuming a mean generation interval time g, the additive bias term has the form

$$g\ log\ \{\frac{1+f(t+1)(c-1)}{1+f(t)(c-1)}\}.$$

The bias vanishes for $f(t+1) \approx f(t)$, i.e., early in the introduction of the new variant or later when it has established itself as the most abundant strain. The estimate will experience a temporary bias as the variant grows (Figure 1D), with the severity of the bias depending on the extent of undershedding (Figure 2B). Importantly, the introduction and spread of a competitive variant, will still be indicated by a rise in the apparent *Re*. However, this rise can be delayed shortly, potentially setting back the day on which the estimate of the *Re* will be observed to cross the *Re*=1.0 threshold.

Increased variability of generation time does not bias estimates of growth advantage and *Re*

The closed-form expression of the bias in *Re* derived above assumes an exponentially-distributed generation interval time (the time between infections in a chain of infections) with mean and standard deviation *g*. Generally this assumption is considered unrealistic, and a preferred, less restrictive assumption is that the generation interval time follows a Gamma distribution, which can account for different degrees of dispersion[17]. We investigated whether

dropping the exponential distribution assumption and varying the level of dispersion in the generation time interval distribution might affect our findings. We simulated viral variant incidence time series from a stochastic model where the generation interval time is sampled from a Gamma distribution with a mean of 4.8 days (Figure 2 C). From these time series, we computed the *Re* values using the method implemented in *EpiEstim*[18]. Varying the variance (while keeping the mean constant) of the generation interval time distribution did not add bias to the growth advantage estimates, and they remained unbiased (Figure 2 D). Similarly, the pattern of transient bias of the *Re* estimates was not affected by increasing the variance of the generation interval time distribution (Figure 2 E).

Growth advantage estimates of BA.1 in Switzerland from wastewater and from clinical data

In Switzerland, SARS-CoV-2 viral loads in wastewater are monitored since 2020, and variants are monitored in wastewater since 2021[1]. During the progression of Omicron, there was an apparent decoupling of the daily measured SARS-CoV-2 load (viral genome copies per day) from the daily reported new cases in all six of the sewersheds monitored at that time (Figure 3 A, B, Supplementary Figure 1). We suspected that the discrepancy might be due to differential shedding of Omicron versus Delta.

Despite this discrepancy, the estimates of the logistic growth rates based on wastewater NGS and dPCR analysis of wastewater were very similar to the estimates based on clinical sequencing data (Figure 3 C, Supplementary Figure 1). Specifically, the logistic growth rates based on wastewater sequencing ranged from 0.14 (0.12 – 0.17) in Laupen to 0.23 (0.19 – 0.27) in Altenrhein. The estimates based on dPCR analysis of wastewater ranged from 0.14 (0.12 – 0.16) in Laupen to 0.23 (0.19 – 0.27) in Chur. In clinical sequencing, the logistic growth rates estimates ranged from 0.12 (0.08 – 0.17) in Zürich to 0.21 (0.15 – 0.27) in Lugano. In contrast, estimates of the midpoint parameter of the growth curve were higher when computed from wastewater data compared to those computed using clinical data (Supplementary Figure 2, Z), suggesting that the bias stemming from possible undershedding was absorbed in the nuisance parameter t_0. Logistic growth curves fitted to the data from the six sewersheds indicated a similar progression of the relative abundances of BA.1 across Switzerland (Supplementary Figure 1).

Estimates of the relative abundance of BA.1 produced using the S:L452R dPCR probe were in general consistently higher compared to those generated using the S:HV69-70 deletion probe or those generated using wastewater NGS (Supplementary Figure 1). This apparent bias is reflected by the estimates of the midpoint parameter of the curve, which were consistently lower (Supplementary Figures Y, Z). However, this bias has no discernible effect on the estimates of the growth advantage of the variant (Figure 3 C, Supplementary Figure 1).

[1] https://www.eawag.ch/en/department/sww/projects/sars-cov2-in-wastewater/

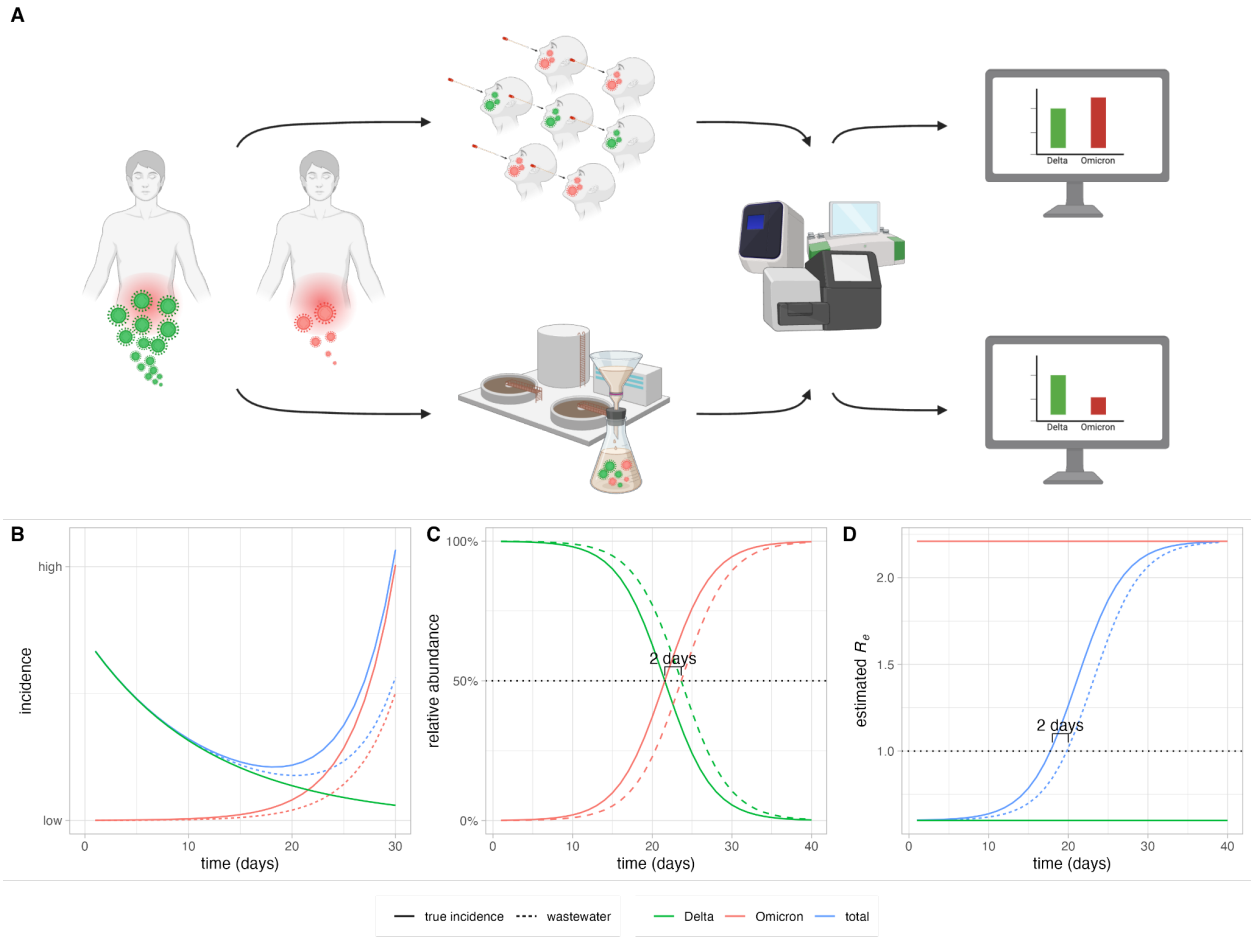

Figure 1: Effect of differential shedding on estimates of incidence, growth advantage and *Re* of SARS-CoV-2 and its variants. **A:** Different variants (Delta, green and Omicron, red) are shed in variable amounts. The incidence of SARS-CoV-2 and its variants in the population is quantified from clinical samples (top path) and from wastewater samples (bottom path) using PCR-based and NGS methods. The difference in shedding can lead to a biased wastewater-based quantification. **B:** Simulated time series: the dominant strain (Delta, green) has a stable *Re* of 0.6, and the incidence is steadily declining. A new variant (Omicron, red) with a growth advantage over Delta resulting in an *Re* of 2.2, is introduced and increases in absolute prevalence. The generation interval time is constant and equal to 4.8 days. For illustration, we assume here that the new variant is shed 50% less, which decouples the historic relationship between community prevalence (solid line) and concentration in wastewater (dashed line), leading to underestimation of its incidence from wastewater-derived data. This in turn also leads to underestimating the total incidence of the virus (blue). **C:** The differential shedding results in a time-shift of 2 days in the growth and decay curves of relative abundance of the variants, but does not alter the growth or decay rates. The estimates of the growth advantage of the variant are not affected. **D:** The differential shedding results in a transient bias in the estimated *Re*. The *Re*=1.0 threshold is estimated to be crossed 2 days later than without undershedding. The variant-specific *Re* for both variants do not suffer any bias.

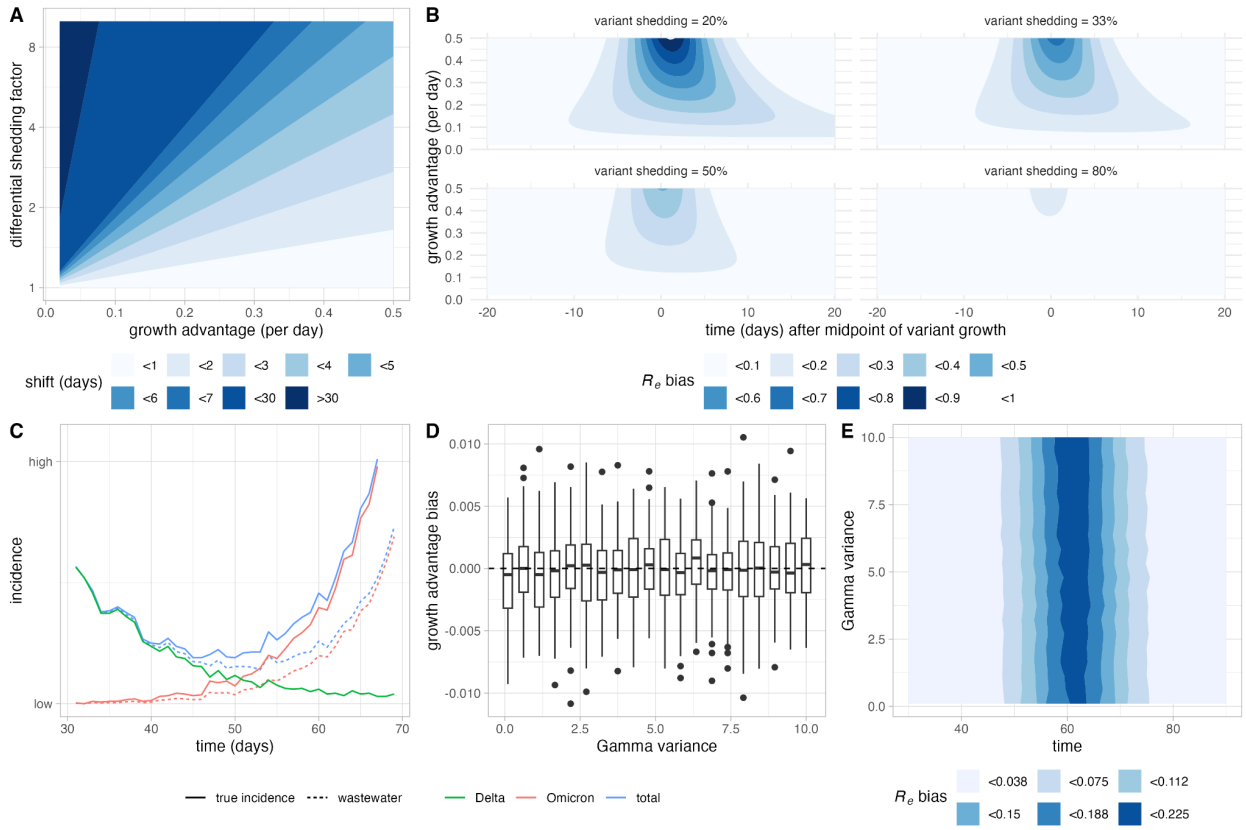

Figure 2: Dependence of the bias in midpoint and *Re* estimation on shedding difference, growth advantage, and generation interval time distribution. **A:** The time-shift in the growth and decay curves depends both on the magnitude of differential shedding and on the growth advantage of the variant. **B:** The magnitude of the bias of the *Re* estimate depends on the amount of undershedding. For variants with a higher growth advantage, the bias rises and decreases more sharply. **C:** Simulated time series, where a variant with *Re*=0.6 (Delta, green) is replaced by a newly introduced variant with *Re*=2.2 (Omicron, red). The generation time is sampled from a Gamma distribution with mean 4.8 days and variance 5 days^2. The new variant is assumed to shed 50% less, leading to underestimation of its incidence (solid line) from wastewater concentrations (dashed line) and underestimation of the total incidence of the virus (blue). **D:** For increasing variance levels of the generation interval time distribution (20 values linearly spaced between 0.1 and 10.0 days, but constant mean of 4.8 days), we produced growth advantage estimates from simulated data with and without undershedding and calculated the difference to estimate the bias stemming from undershedding. At each variance level, we repeated the simulation 100 times. Boxplots hinges represent the median and quartiles, with whiskers extending to the largest and smallest values no further than 1.5 times the interquartile range (data points beyond these limits are plotted individually). On average, the bias is zero, independently of the variance of the generation interval time distribution. **E:** Using the same simulations as in B, the bias in *Re* over time averaged over the simulation runs is shown. The transient pattern of the bias is not affected by the variance of the generation interval time distribution.

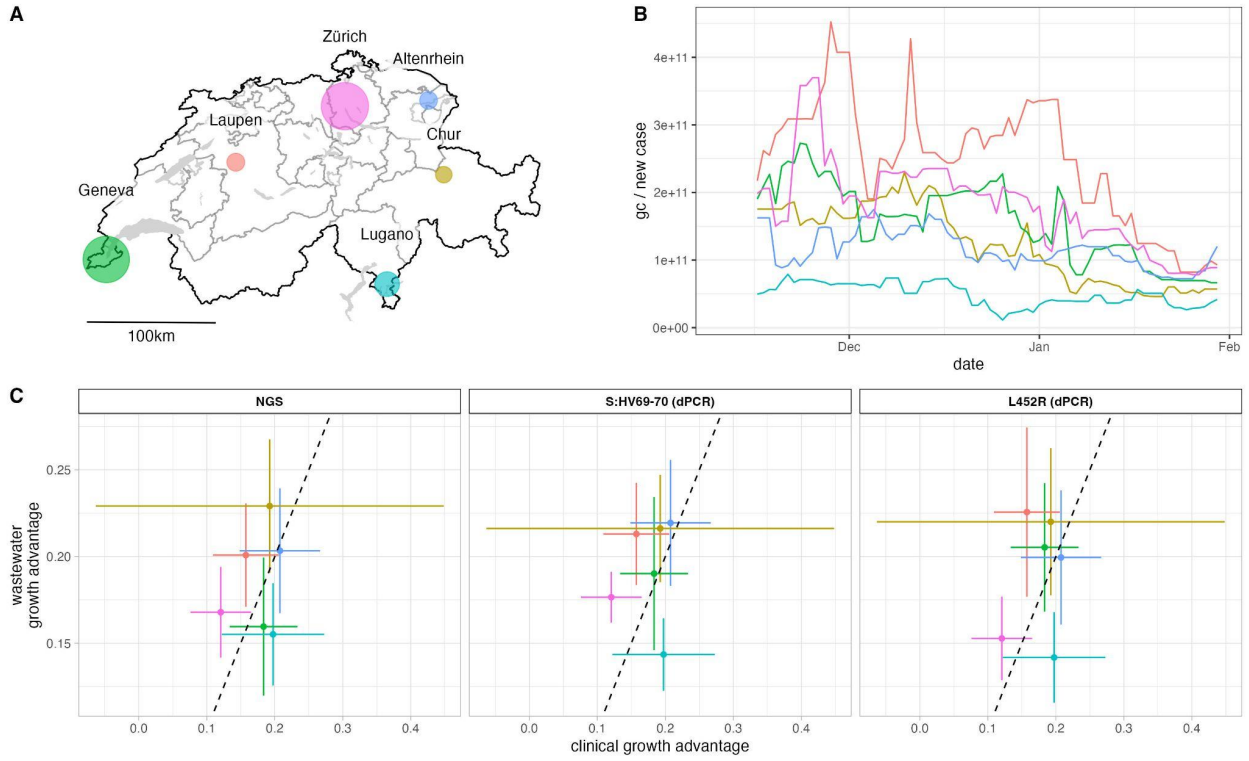

Figure 3: Concordant estimation of the growth advantage of Omicron BA.1 in Switzerland using clinical and wastewater-derived data. **A:** Map of Switzerland indicating the location of all six wastewater treatment plants (WWTPs) sampled in this study. The discs are scaled to represent the amount of inhabitants in the respective catchment areas. **B:** SARS-CoV-2 genome copies in wastewater (7-day rolling median) per new confirmed case in the catchment area (7-day rolling median), throughout the spread of BA.1. **C:** Growth advantage estimates for BA.1 in the six WWTP regions, based on wastewater NGS (left) as well as wastewater ddPCR duplex assays targeting the S:HV69-70 deletion (center) and S:L452R substitution (right), compared with estimates derived from clinical sequencing from the cantons surrounding the WWTPs. Error bars represent 95% Wald confidence intervals adjusted for overdispersion.

Discussion

A current caveat of wastewater-based epidemiology is the possible impact of differential shedding. The introduction and spread of a new variant that is shed more or less than the current one can introduce bias in the real time estimates of viral incidence and prevalence, as well as estimates of variant relative abundances. Beyond these quantities, wastewater-based epidemiology can provide multiple other metrics of central importance to public health strategies. Metrics describing growth of the virus and its variants are of prime importance and are expected to be less sensitive to biases affecting normalization. In the case of genomic epidemiology, a parameter that is crucial to estimate when a new variant emerges is its growth advantage relative to the currently circulating strain(s)[11]. Here, we have shown that if there is differential shedding by a constant factor, then no bias is introduced in the estimation of the growth advantage. A parameter describing the current disease dynamics, which is critical to assess the epidemiological situation and inform policy, is the effective reproduction number (*Re*). We have shown here that the bias of *Re* estimation stemming from differential shedding is transient, i.e., the estimator will recalibrate itself after a short period of time. We derived a formula for the bias based on the mean generation interval time, the growth advantage and the difference in shedding.

Clinical data is generally considered the gold standard to obtain epidemiological insight into the dynamics of a pathogen and its variants. However, estimates based on clinical data are themselves subject to biases, which are not encountered in wastewater-derived data, such as non-random testing and sequencing. Also, they require orders of magnitude more samples. Our findings demonstrate that wastewater-based epidemiology remains a valid (and advantageous) tool to estimate competitive advantages between multiple variants of a pathogen and to assess the current disease dynamics, even in the presence of differential shedding. Moreover, it follows that a large number of possible biases stemming from the protocol used for quantification (e.g., underestimating a variant prevalence due to the dPCR probe used) will also not impact the estimation of the growth advantage. This robustness makes integrating various data sources into a large scale program easier and more robust to differences in methodology among participating labs. Lastly and beyond wastewater-based epidemiology, differential clinical testing between variants (for example due to differential asymptomatic infection rates) should, following the same reasoning, introduce a bias in the quantification of a variant that will not impact variant growth advantage estimates from clinical data.

Tracking the progression of a variant using dPCR relies on the availability of an assay targeting a mutation separating it from the currently circulating variants. Using an already established assay can reduce the lead time needed to track the progression of a new variant. We tracked the spread of BA.1 using not only the rise of a BA.1 signature mutation (S:HV69-70), but also using the decline of a Delta signature mutation (L452R). Despite yielding different estimates of the relative abundance of Omicron BA.1, the same estimates of the growth advantage were obtained using the different dPCR assays. This result further illustrates the robustness of estimating the growth advantage of a variant to quantification biases.

More generally, we suggest that if there is a potential introduction of a new source of bias in the quantification of a pathogen or its variants, it is essential to understand the impact of this bias on the estimation of relevant quantities. As we showed here for differential shedding, it might turn out that this bias has no bearing on the answers to relevant questions regarding growth. In other cases, it might also be that even though the bias indeed propagates to the estimates of some quantity of interest, it does so in such a small magnitude relative to the already present statistical error that it can be ignored in practice. If feasible and even if based on strong assumptions, closed-form expressions can clarify the nature of the bias. Additionally, simulations can help test the assumptions of the closed-form expressions or replace them entirely when they are not available. Here, we investigated the effects of a scaling of the shedding profile, but other sources of biases (e.g., changes in mean generation time, shifts in shedding load distributions, etc.) should be rigorously studied to further test the robustness of wastewater-based genomic epidemiology.

Methods

Bias in growth advantage estimates due to differences in shedding

We derive a closed-form expression for the bias in growth advantage due to lower or higher shedding of a variant, and show that it is zero. Let $X(t)$ and $Y(t)$ be the incidence of two variants in the population through time t. Let $f(t)$ be the relative abundance of X in the population

$$f(t) = X(t)/(X(t) + Y(t))$$

The logit transform of $f(t)$ is then:

$$logit\ (f(t)) \ = log\ \{f(t)/[1 - f(t)]\} \ = log\ \{X(t)\} - log\ \{Y(t)\}$$

Let $X'(t)$ and $Y'(t)$ be the measured incidence (derived from viral genome copies) of X and Y in the wastewater through time t. Let $g(t)$ be the shedding load profile describing the magnitude of viral particles shed at time t. $Y'(t)$ and $Y(t)$ are related by a convolution with the shedding load profile $g(t)$, such that

$$Y'(t) = Y(t) * g(t) = \int_{\tau \in T} g(t - \tau) Y(\tau) d\tau$$

If X is on average shed less (or more) by a constant c relative to Y, such that the shedding load profile is scaled by c, then

$$X'(t) = X(t) * cg(t) = c(X(t) * g(t))$$

Where the second equality follows from associativity of the convolution product with scalar multiplication. For the fraction $f'(t) = X'(t)/(X'(t) + Y'(t))$ of $X'(t)$ concentration in wastewater, we have

$$logit\ (f'(t)) \ = log\ \{X'(t)\} - log\ \{Y'(t)\} \ = log\ \{c(X(t) * g(t))\} - log\ \{Y(t) * g(t)\}$$

$$= log\ \{X(t) * g(t)\} - log\ \{Y(t) * g(t)\} + log\ \{c\} = logit\ (f(t)) + log\ \{c\}$$

Thus, the logit-transformed estimates of wastewater-based prevalences are shifted by the constant $log\{c\}$, the logarithm of the constant factor by which X is differentially shed.

If $f(t)$ is assumed (as is commonly the case) to follow a logistic growth with rate a (the relative growth advantage of the variant over the dominant strain) and midpoint t_0, then $f'(t)$ will follow a logistic growth with the same rate, with the bias from undershedding being absorbed by the midpoint parameter t_0. To see this, let a' and $t_0{}'$ be the growth rate and midpoint observed in wastewater. Then

$$a'(t - t_0{}') = logit\ (f'(t)) = logit\ (f(t)) + log\ \{c\} = a(t - t_0) + log\ \{c\}\ = a(t - t_0 + log\ \{c\}/a)$$

We conclude that $a' = a$ and that $t_0{}' =\ t_0 - log\{c\}/a$, i.e., the observed growth advantages are equal, and the midpoints differ by a constant depending on c and a.

Bias in *Re* estimates due to differences in shedding

Next, we analyze the effective reproduction number *Re*. We derive a closed-form expression for the bias in *Re* estimation due to lower shedding of a variant and show that it is significant for a short period of time. We construct our derivation using the approximation presented by Bettencourt and Ribeiro, which is based on the SIR model and thus assumes an exponential distribution of the generation interval time[17,19]. The approximation relates the incidence through time $I(t)$ to the reproduction number $R(t)$ and mean generation time g by

$$I(t+1) = I(t)e^{\frac{R(t)-1}{g}}$$

Solving for $R(t)$ and substituting *f(t)* we obtain

$$R(t)\ =\ 1 + g\ log\ \{\frac{f(t+1)I(t+1)\ +\ (1 - f(t+1))I(t+1)}{f(t)I(t)\ +\ (1 - f(t))I(t)}\}$$

If X is differentially shed by a factor *c*, then the observed reproduction number $R'(t)$ is affected as

$$R'(t)\ =\ 1 + g\ log\ \{\frac{cf(t+1)I(t+1) + (1-f(t+1))I(t+1)}{cf(t)I(t) + (1-f(t))I(t)}\} = 1 + g\ log\ \{\frac{I(t+1)}{I(t)}\} + g\ log\ \{\frac{1+f(t+1)(c-1)}{1+f(t)(c-1)}\}$$

i.e., the observed reproduction number suffers an additive bias of $g\ log\ \{\frac{1+f(t+1)(c-1)}{1+f(t)(c-1)}\}$.

Stochastic simulations of variant emergence

In the derivations of closed form expressions for the biases in growth advantage and *Re* due to reduced or increased shedding, we assumed an exponentially-distributed generation time interval. We simulated data to assess the impact of underdispersed or overdispersed generation time intervals on the estimates of *Re* and growth advantage. We simulated two distinct time series of SARS-CoV-2 infections, each for a different variant: the first variant started from 1500 cases with a constant Re of 0.6, and the second started from a single case with a constant Re of 2.2. The simulations largely followed the simulation framework used in Huisman et al.[2,20]. Infections were simulated forward in time using the renewal equation framework described by Cori et al[2120,21]. We relaxed the exponentially-distributed generation time to a Gamma-distributed generation time, with constant mean but varying variance. The overall *Re* was then estimated

from the simulated infection incidence time series, once considering the two time series untouched and once having the second time series scaled down by 50% to simulate undershedding. The *Re* values were computed using the package *EpiEstim*[18].

Wastewater sampling and processing

Wastewater-based surveillance for SARS-CoV-2 was conducted in six sewersheds (Altenrhein, Chur, Laupen, Lugano, Geneva, Zurich) across Switzerland from 24 November 2021 through 10 January 2022 (Figure 3 A, Supplementary Figure 1). Twenty-four–hour flow composite samples were collected daily from wastewater influent at each of the six sites and stored at 4°C for up to 5 days before being transported on ice for processing at a central laboratory (Eawag, Dübendorf, Switzerland). Processing included total nucleic acid extraction from 40 ml samples (Wizard Enviro Total Nucleic Acid Extraction Kit, CN A2991, Promega Corporation, USA) with an elution volume of 80ul and subsequent inhibitor removal using OneStep PCR Inhibitor Removal columns (CN D6030, Zymo Research, USA). From these samples, a subset were analyzed for variants of concern using drop-off RT-dPCR assays targeting signature mutations for Delta (S:L452R, n = 74 samples) and Omicron BA.1 (S:HV69-70, n = 79), based on Caduff et al. (2022)[4]. Almost all (n = 280) samples were also analyzed using NGS to identify Delta versus Omicron BA.1, based on Jahn et al. (2022)[5]. RNA extracts were stored at -80°C for up to 1 week prior to sequencing and up to 3 months prior to RT-dPCR analysis.

Drop-off RT-dPCR assays for detection of signature mutations

Drop-off RT-dPCR assays targeted the S:HV69-70 deletion indicative of Omicron BA.1, as previously described[4], and the S:L452R mutation indicative of Delta (lineage B.1.617.2).The assay targeting S:L452R includes two hydrolysis probes binding to a single amplicon: a universal probe targeting a conserved region on the amplicon and a variant-specific probe that binds to the S:L452R mutation (Supplementary Table 1, Supplementary Figure 3). In the digital PCR, generated droplets with dual fluorescence indicates the presence of an amplicon from the mutation (i.e., Delta), whereas single fluorescence indicative of only the universal probe indicates an amplicon without the S:L452R mutation (i.e., Omicron). For the S:HV69-70 assay, the variant-specific probe only binds when S:HV69-70 is present[22], and therefore dual fluorescence in a given droplet indicates presence of an amplicon without the mutation (i.e. Delta), whereas single fluorescence of only the universal probe indicates the presence of the mutation (i.e., Omicron).

Clinical VOC data

For each canton surrounding a WWTP, we downloaded counts of infected individuals binned by variant of sequenced PCR-positive clinical samples through the LAPIS API of Cov-Spectrum [23]. The data was restricted to sequences originating from the Viollier lab (8525 sequences, Supplementary Figure 1), which sends a random subset of their PCR-positive samples out for sequencing. For the Geneva sewershed, we also compared estimates to data on the qPCR S-gene target failure (SGTF) data from the Geneva University Hospital (HUG) available at https://www.hug.ch/laboratoire-virologie/surveillance-variants-sars-cov-2-geneve-national (5794

tests, Supplementary Figure 1),. The SGTF qPCR is based on the detection of S:HV69-70; failed amplification of a clinical sample previously positive for the N1 gene target is used as a proxy to indicate that the clinical sample is Omicron BA.1.

Data Analysis

We analyzed the data using the R statistical programming language and the R package WWdPCR[2] [4]**.** The package was used to obtain maximum likelihood estimates (MLE) and confidence intervals of the logistic growth rate of the Omicron BA.1 variant in each region. Confidence intervals for the logistic growth parameter were computed assuming a quasibinomial (for the clinical and wastewater sequencing data) or quasimultinomial (for the dPCR data) model of the counts to account for overdispersion, but without allowing underdispersion (i.e. overdispersion factors <1 were not considered). Confidence bands for the fitted values were computed on the logit scale using the Delta method and then back-transformed to the linear scale, to optimize their coverage and ensure they were constrained to the [0,1] range.

The logistic growth model was fitted separately using the wastewater S:L452R dPCR data, the wastewater S:HV69-70 dPCR data, the wastewater sequencing data, and the clinical sequencing data. For Geneva, we fit the model also on the SGTF data. For the S:L452R dPCR, the dual fluorescence droplets are indicative of Delta, so we assumed that the single fluorescence droplets indicated Omicron BA.1. The growth advantage of BA.1 was calculated using S:L452R data assuming a negative logistic decay rate of the mutated fraction. For wastewater sequencing, we proceeded as previously described in Jahn et al. [5], and we estimated the relative abundance of BA.1 using the observed fractions of reads bearing mutations characteristic of BA.1, while excluding mutations also present in B.1.617.2*. Amplicons suffering potential differential dropout rates or altered amplification due to mutations in the primer regions were discarded.

References

1. Naughton, C. C. *et al.* Show us the data: global COVID-19 wastewater monitoring efforts, equity, and gaps. *FEMS Microbes* **4**, xtad003 (2023).
2. Huisman, J. S. *et al.* Wastewater-based estimation of the effective reproductive number of SARS-CoV-2. *Environ. Health Perspect.* **130**, 57011 (2022).
3. Carcereny, A. *et al.* Dynamics of SARS-CoV-2 Alpha (B.1.1.7) variant spread: The wastewater surveillance approach. *Environ. Res.* **208**, 112720 (2022).

[2] https://github.com/cbg-ethz/WWdPCR

4. Caduff, L. *et al.* Inferring transmission fitness advantage of SARS-CoV-2 variants of concern from wastewater samples using digital PCR, Switzerland, December 2020 through March 2021. *Euro Surveill.* **27**, (2022).
5. Jahn, K. *et al.* Early detection and surveillance of SARS-CoV-2 genomic variants in wastewater using COJAC. *Nat Microbiol* **7**, 1151–1160 (2022).
6. Karthikeyan, S. *et al.* Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. *Nature* **609**, 101–108 (2022).
7. Amman, F. *et al.* Viral variant-resolved wastewater surveillance of SARS-CoV-2 at national scale. *Nat. Biotechnol.* **40**, 1814–1822 (2022).
8. Bagutti, C. *et al.* Wastewater monitoring of SARS-CoV-2 shows high correlation with COVID-19 case numbers and allowed early detection of the first confirmed B.1.1.529 infection in Switzerland: results of an observational surveillance study. *Swiss Med. Wkly* **152**, w30202 (2022).
9. Vöhringer, H. S. *et al.* Genomic reconstruction of the SARS-CoV-2 epidemic in England. *Nature* **600**, 506–511 (2021).
10. Viana, R. *et al.* Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* **603**, 679–686 (2022).
11. Chen, C. *et al.* Quantification of the spread of SARS-CoV-2 variant B.1.1.7 in Switzerland. *Epidemics* **37**, 100480 (2021).
12. Natarajan, A. *et al.* Gastrointestinal symptoms and fecal shedding of SARS-CoV-2 RNA suggest prolonged gastrointestinal infection. *Med (N Y)* **3**, 371–387.e9 (2022).
13. Hoffmann, T. & Alsing, J. Faecal shedding models for SARS-CoV-2 RNA among hospitalised patients and implications for wastewater-based epidemiology. *bioRxiv* (2021) doi:10.1101/2021.03.16.21253603.
14. Puhach, O., Meyer, B. & Eckerle, I. SARS-CoV-2 viral load and shedding kinetics. *Nat. Rev. Microbiol.* **21**, 147–161 (2023).

15. Prasek, S. M. *et al.* Variant-specific SARS-CoV-2 shedding rates in wastewater. *Sci. Total Environ.* **857**, 159165 (2023).
16. Miyakawa, K. *et al.* Reduced Replication Efficacy of Severe Acute Respiratory Syndrome Coronavirus 2 Omicron Variant in 'Mini-gut' Organoids. *Gastroenterology* **163**, 514–516 (2022).
17. Gostic, K. M. *et al.* Practical considerations for measuring the effective reproductive number, Rt. *PLoS Comput. Biol.* **16**, e1008409 (2020).
18. Thompson, R. N. *et al.* Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics* **29**, 100356 (2019).
19. Bettencourt, L. M. A. & Ribeiro, R. M. Real time bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS One* **3**, e2185 (2008).
20. Huisman, J. S. *et al.* Estimation and worldwide monitoring of the effective reproductive number of SARS-CoV-2. *Elife* **11**, (2022).
21. Cori, A., Ferguson, N. M., Fraser, C. & Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **178**, 1505–1512 (2013).
22. Vogels, C. B. F. *et al.* Multiplex qPCR discriminates variants of concern to enhance global surveillance of SARS-CoV-2. *PLoS Biol.* **19**, e3001236 (2021).
23. Chen, C. *et al.* CoV-Spectrum: Analysis of Globally Shared SARS-CoV-2 Data to Identify and Characterize New Variants. *Bioinformatics* (2021) doi:10.1093/bioinformatics/btab856.