water

MDPI

Article

Wastewater Surveillance of SARS-CoV-2 in Minnesota

Mark J. Osborn [1], Shannon Champeau [1], Carolyn Meyer [1], Mason Hayden [1], Laura Landini [1], Stacey Stark [2], Stephanie Preekett [3], Sara Vetter [4], Zachary Zirnhelt [4], Stephanie Meyer [4], Daniel Huff [4], Timothy W. Schacker [3] and Charles R. Doss [5,*]

[1] Department of Pediatrics, University of Minnesota, Minneapolis, MN 55455, USA; osbor026@umn.edu (M.J.O.); schampea@umn.edu (S.C.); meyer167@umn.edu (C.M.); hayde183@umn.edu (M.H.); lplandini@gmail.com (L.L.)
[2] U-Spatial, University of Minnesota-Duluth, Duluth, MN 55812, USA; slstark@d.umn.edu
[3] Department of Medicine, University of Minnesota, Minneapolis, MN 55455, USA; preekett@umn.edu (S.P.); schac008@umn.edu (T.W.S.)
[4] Minnesota Department of Health, St. Paul, MN 55164, USA; sara.vetter@state.mn.us (S.V.); zach.zirnhelt@state.mn.us (Z.Z.); stephanie.meyer@state.mn.us (S.M.); daniel.huff@state.mn.us (D.H.)
[5] School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA
* Correspondence: cdoss@umn.edu

Abstract: Wastewater-based epidemiology provides an approach for assessing the prevalence of COVID-19 in a sewer service area. In this study, SARS-CoV-2 RNA was measured in 44 wastewater-treatment plants of varying sizes that sampled approximately 67% of the population of Minnesota, from September 2020 through December 2022. Various linear regression models were investigated to predict the weekly case count from SARS-CoV-2 RNA concentrations under various transformation and normalization methods which we validated via cross-validation averaged across all treatment plants. We find that the relationship between COVID-19 incidence and SARS-CoV-2 RNA in wastewater may be treatment plant-specific. We study storage and time-to-analysis for RNA wastewater data and find large effects of storage temperature, indicating that collection methods may have an important effect on the utility and validity of wastewater data for infectious disease monitoring. Our findings are important for any large-scale wastewater surveillance program.

Keywords: wastewater; SARS-CoV-2; COVID-19; storage stability

check for updates

Citation: Osborn, M.J.; Champeau, S.; Meyer, C.; Hayden, M.; Landini, L.; Stark, S.; Preekett, S.; Vetter, S.; Zirnhelt, Z.; Meyer, S.; et al. Wastewater Surveillance of SARS-CoV-2 in Minnesota. *Water* **2024**, *16*, 541. https://doi.org/10.3390/w16040541

Academic Editor: Abasiofiok Mark Ibekwe

Received: 12 January 2024
Revised: 31 January 2024
Accepted: 7 February 2024
Published: 9 February 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has infected millions of people across the globe which resulted in significant health and economic impacts, and demonstrated the importance of having infrastructure available to deploy for monitoring diseases on a massive scale. Mass community testing is costly and the demand for tests can exceed the capacity of testing facilities during an infection surge [1], not to mention that not everyone has access to testing due to economic, geographic, or social restrictions. Furthermore, test results are a lagging indicator of a pandemic's progression, because testing is usually prompted by symptoms, which may take 2 weeks to show up after infection [2]; furthermore, the more readily available home testing is, the less informative official clinical testing data becomes, since those not requiring medical attention will tend to also not show up in public health testing data as much as those who do require attention, creating confounding and bias in the data. Thus, delays may occur between the appearance of symptoms, testing and the reporting of test results [3]. Finally, for COVID-19, it is estimated that as many as 45% of cases are asymptomatic [4–7]. Considering that people only seek medical attention and undergo diagnostic testing if they are symptomatic, the number of confirmed clinical cases may grossly underestimate the prevalence of the disease (and its capacity to spread).

Wastewater-based epidemiology (WWBE) is a method for monitoring presence and trends of an infectious agent that is shed in bodily secretions/excrement in communities. In WWBE for SARS-CoV-2, wastewater is collected from wastewater-treatment plants (WWTPs) and is tested by qRT-PCR for SARS-CoV-2 RNA excreted via feces, urine, and saliva. The presence and levels of viruses in wastewater samples represents a cross section of viral presence and spread in the communities served by those plants [8–10]. WWBE has been successfully employed as a surveillance tool for diseases such as SARS, hepatitis A, and polio [11–13]. With regard to SARS-CoV-2, viral particles are reported to be shed in feces from infected individuals even if they are asymptomatic [14–18]. Recent studies have shown that WWBE is able to predict COVID-19 prevalence even earlier than clinical case data [3,19–22], supporting the idea that WWBE can be used as an early warning system to monitor community prevalence and spread [23,24].

Over the last three years we have gathered SARS-CoV-2 wastewater data, which we present in this paper both to describe the programmatic data to date and to provide more about the specifics of the SARS-CoV-2 pandemic, and to identify best practices for WWBE more broadly. We present data from two complementary data gathering processes here. The primary data is wastewater samples that were collected from 44 total WWTPs across the state of Minnesota between March 2022 and December 2022. (Our team's data gathering began in May of 2020, but the PCR assay changed in March 2022, as discussed below, so we focus on data after that time period.) The SARS-CoV-2 viral load in wastewater were obtained by qRT-PCR. These 44 WWTPs represent a broad sampling of the Minnesota population serving a total of 191 zip codes and a population of 3,825,269 people, which is approximately 67% of the total population of the state. The secondary data were from an experiment on storage and time-to-analysis conditions for qRT-PCR analysis of wastewater data gathered in the summer of 2023. This experiment was undertaken in part because of our findings from our primary data of significant heterogeneity in the data quality and predictive ability from different WWTPs.

Estimating the SARS-CoV-2 RNA concentrations in wastewater (gene copies per litre) is complicated, as the dilution and fecal strength in the wastewater may vary between sampling dates due to random chance (e.g., caused by variation in a factory's runoff). It has been recommended to multiply the viral concentration in wastewater by the flow of the sampled location (the volume of wastewater that passed through the location in a day) to obtain the viral concentrations in gene copies per day, and account for changes in sanitary sewer contributions [25,26]. Normalizing SARS-CoV-2 RNA concentrations by indicators of human fecal waste is also common, because feces in wastewater can have variable levels of SARS-CoV-2 depending upon the amount of water used per toilet flush or body washing [27]. The contribution of SARS-CoV-2 from human sourced water can then be estimated by dividing the measured SARS-CoV-2 concentration by the concentration of the human waste indicator [27]. A typically examined fecal marker is Pepper Mild Mottle Virus (PMMoV) [19,27,28]. Previous studies have shown that PMMoV is the most abundant RNA virus in human feces and it is shed in large quantities in wastewater [29–33]. It is also highly stable in wastewater, and its concentrations showed little seasonal variation [30,31]. Operating on this information, our program from 2019–2022 normalized SARS-CoV-2 RNA wastewater levels to PMMoV [19]. However, in our ongoing surveillance work, we found that normalizing by PMMoV obfuscated the large spike in wastewater measurements during the COVID-19 Omicron wave of December 2021–January 2022, which compelled us to reexamine normalization methods. Our findings below suggest that normalization by PMMoV may increase the variation in wastewater measurements, and we do not find evidence in the current data that it substantially improves prediction.

Our study objectives in conducting these analyses were to: (1) learn how best to make use of large-scale wastewater surveillance RNA measurements, and (2) study how storage conditions affect the quality of such measurements. Our motivation for (2) was driven in part by our results in (1), in which we found nontrivial unexplained heterogeneity between WWTPs. We compare a variety of data pre-processing methods; to benchmark the

effectiveness of a pre-processing method we use predictive ability of the data on clinical case counts. Rather than use statistical significance testing of model coefficients, or coefficient of determination (R^2) from linear regression for assessment of fit, we will use leave-one-out cross-validation, which is appropriate for assessing model performance when prediction is the main goal and when a variety of models and data transformations are being compared. (See the methods section for a full description).

However, we have observed nontrivial heterogeneity in the predictive ability of measurements from differing WWTPs. Researchers have started to investigate how storage conditions affect wastewater RNA and measured concentrations [34]. In a small scale study focusing on one or a small number of WWTPs, the storage and measurement conditions can easily be controlled. As we will discuss in our results section, we observed significant heterogeneity in the relationship of the wastewater measurements with clinical COVID-19 case counts which we attribute to heterogeneous quality of data from the WWTPs. Thus, we also implemented a study on storage and time-to-analysis conditions for measurements from wastewater, to undertake efforts how much these aspects affect quality of measurements in order to define sample collection guiding principles.

Our surveillance study yields the important finding that data from different WWTPs may have very different predictive ability, even when analyzed by the same laboratory. We also conclude that parsimonious models perform the best for prediction (and the optimal model we find, for our data, is presented below). We find in our stability study that maintaining wastewater samples at 4 °C leads to less RNA degradation than freezing and thawing. The results were statistically and practically significant.

2. Materials and Methods

2.1. Viral Load Quantification

The procedure we used for quantifying viral load is as follows: 40 mL of influent was used to isolate total nucleic acids using the Enviro Total Nucleic Acid Kit for Wastewater (Promega, Madison, WI, USA) and eluted in a volume of 40 µL. Five ul was employed for quantitative RT PCR, performed in duplicate, using the TaqPath™ COVID-19 Combo Kit (ThermoFisher, Waltham, MA, USA) with copy number standards on the QuanStudio 5 RT PCR instrument in order to quantitate viral copy number. For the stability study, a large volume was obtained and then aliquoted for defined times over multiple temperatures. Four total samples from a single site were obtained and the RNA was isolated and qRT PCR was done on all samples concurrently.

2.2. Statistical Methods

Wastewater samples were collected from a total of 44 WWTPs across the state of Minnesota (see Figure 1 for a map of WWTP locations). From its inception to March 2022, we employed a polyethylene glycol concentration and phenol chloroform nucleic acid extraction procedure and qRT-PCR for the nucleocapsid gene in the SARS-CoV-2 genome using N1 and N2 primer:probe pairs. As the pandemic progressed so too did more sensitive and effective isolation and detection methods. Accordingly, in March 2022 we pivoted to a column-based total nucleic acid isolation method (Wizard® Enviro Total Nucleic Acid Kit (Promega, Madison, WI, USA)) and a three gene qRT PCR method that measured the nucleocapsid (N), spike (S), and ORF1ab (O) proteins, sampled from 40 WWTPs. Our statistical findings demonstrate the latter assays are more accurate than the former assays. We present in this paper, data from the latter period (March 2022 to December 2022), using the three gene qRT PCR assay. Since the two assays are nontrivially different, they cannot be combined into a single statistical analysis. For data analysis purposes, a weekly level average of measurements was used. One sample was reported to have zero concentration of N, S and O. It was removed in further analyses.

Among the remaining samples, 424 were reported to have zero virus concentration of S. The reason for the loss of signal (S gene target failure) was due to prevalence of a SARS-CoV-2 variant that acquired mutations in the spike gene that prevented the

primer:probe pair from binding. Thus, the S gene was not used in any further analyses. Importantly, our analytical resolution was preserved using the N and Orf1a genes. The concentrations of a human fecal marker, PMMoV, were also determined in each wastewater sample in parallel. PMMoV concentrations below the 5th percentile were replaced with the 5th percentile value, and those above the 95th percentile were replaced with the 95th percentile value to avoid outliers. The influent flow rate was provided by the participating WWTPs. For 13 of the WWTPs, the catchment area for the WWTP was obtained from the Minnesota Pollution Control Agency (MPCA) or in some cases directly from the treatment facility. For the other WWTPs that did not have a digitized catchment area available, the city boundaries were used to approximate this area for these. Many of those were rural WWTPs where the city boundaries are likely to be a good approximation of the WWTP catchment area. Data enrichment services from Esri [35] were used to apportion 2020 census block population to the WWTP catchment areas.

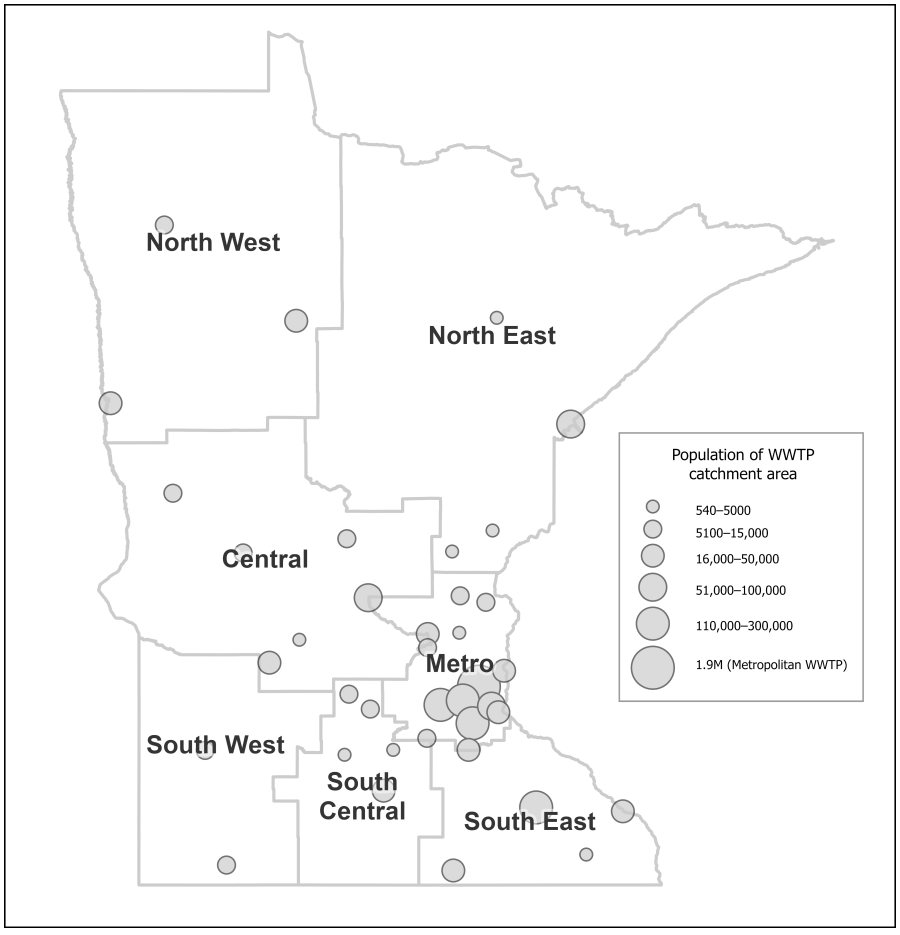

Figure 1. Map of WWTPs across the state of Minnesota.

The weekly number of new infections for each WWTP service area was obtained from the Minnesota Department of Health (MDH) and aggregated to the zip code areas using the residential address for the case. Case counts were summed by zip code weekly (Sunday–Saturday) and the date beginning the week was used to associate with each case count. The case data were then aggregated by WWTP according to the zip codes previously found to be associated with each catchment area. The zip codes were spatially intersected with the sewersheds to identify all of the zip codes associated with each sewershed. Each zip code area was visually inspected to ensure that if a zip code intersected two sewersheds, that zip code would be associated only with the sewershed with the most area in the zip code. Five zip codes intersected more than one WWTP catchment area. To allow a $\log_{10}$ transformation, any zero case counts were replaced with a value of one. The MDH also

provided the number of people who have received at least one dose of COVID-19 vaccine, and the number of people with completed vaccine series in the areas served by each WWTP over time. The complete series could be one, two, or three doses depending on the person's age and which vaccine they received. Being "fully vaccinated" does not include or require further booster doses in the present definition.

The wastewater samples were collected twice per week during the course of the study and were isolated on different days (Monday and Wednesday or Tuesday and Thursday) for different WWTPs. To align the wastewater samples with the clinical case counts and to avoid having to analyze data on irregular time intervals, we aggregated the wastewater measurements by week (Sunday-Saturday). For a given week with any wastewater data, we use the sample average of all the measurements of that week as the aggregated measurement. If there was no wastewater sample in a week then that week was considered missing and was not included in the analysis.

With time-aligned case counts and wastewater measurements, we ran linear regression models (the specific models will be described shortly). We fit separate models for each WWTP. Due to the relatively low number of data points for each WWTP (recall Table 1), nonlinear or machine learning methods are not effective. To assess the model performance, a leave-one-out cross-validation (LOOCV) was used. The models were trained on $n - 1$ observations and validated on the remaining one observation, where n is the sample size. The procedure was repeated n times with each of the n observations used exactly once for validation. The average of the n prediction errors obtained was computed for model comparison. The evaluation metric was the root mean squared errors (RMSE) between the predicted value and the actual value. If the dependent variable was $\log_{10}$-transformed, the transformation was reversed to obtain the predictions on the original scale before the computation of RMSE. This provides an unbiased approach for comparing predictive performance of different models. All statistical analyses were performed using R version 4.2.3 [36].

Table 1. Descriptions of participating WWTPs.

WWTP	Code	No. Weeks w/Data	Population	Avg. Weekly Case Count
Albert Lea	AL	36	20,957	28
Alexandria	AX	35	25,689	41
Bemidji	BI	26	32,741	38
Blooming Prairie	BP	10	3589	4
Blue Lake	BL	28	397,299	674
Cambridge	CB	35	15,529	19
Chisholm	CM	20	5787	9
Eagles Point	EP	29	113,786	208
East Bethel	EB	27	14,880	17
Elk River	ER	34	94,119	133
Empire	EM	28	176,441	287
Fergus Falls	FF	35	18,799	24
Glacial Lakes SSWD	GL	33	5574	6
Glencoe	GC	9	8280	14
Hastings	HS	28	30,027	39
Hinckley	HY	30	5483	7
Hutchinson	HT	24	17,829	32
International Falls	IF	33	9618	20
Lafayette	LY	15	844	1
Lanesboro	LB	24	1849	2
Le Sueur	LS	35	8237	7
Little Falls	LF	36	14,615	26
Mankato	MK	25	63,489	86
Marshall	ML	36	15,550	19
Moorhead	MH	31	42,602	69
Mora	MR	36	9872	15

Table 1. *Cont.*

WWTP	Code	No. Weeks w/Data	Population	Avg. Weekly Case Count
New Prague	NP	36	12,557	14
North Branch	NB	35	16,827	22
Northfield	NF	35	27,040	32
Rochester	RC	36	123,041	242
Rogers	CR	27	14,573	21
Saint Cloud	SD	35	130,901	260
Seneca	SN	28	243,338	430
St. Croix Valley	SV	28	38,958	68
Thief River Falls	TR	36	12,971	17
Twin Cities	TC	28	1,854,291	3261
Western Lake Superior Sanitary District	WL	28	124,459	241
Willmar	WM	36	23,697	35
Winona	WA	34	34,644	59
Worthington	WT	36	14,487	18

Use of SARS-CoV-2 Concentrations

Denote the COVID-19 case count at time t by C_t and denote the SARS-CoV-2 concentrations (either O or N) measured in a wastewater sample at time t by W_t. In this section, we compare different linear regression models for predicting C_t from W_t. As mentioned in Section 1, virus concentrations are commonly normalized by either flow, $W_t^{\text{Flow}} = W_t \cdot \text{Flow}_t$ or a fecal marker such as PMMoV: $W_t^{\text{PMMoV}} = W_t / \text{PMMoV}_t$. However, there have been contradictory findings on whether normalization of virus concentration can improve correlations with cases [28,37,38]. Moreover, it seems that no studies examined using both flow and PMMoV to normalize the virus concentration, as in $W_t^{\text{Flow \& PMMoV}} = W_t \cdot \text{Flow}_t / \text{PMMoV}_t$. (This latter normalization has perhaps less logical coherence but to the extent that flow and PMMoV are both informative, but noisy and imperfect, it could be helpful.) We compare these three normalization approaches. It is also of interest to know whether adding lagged virus concentrations (e.g., W_{t-1} and W_{t-2}) in the model will improve the predictive performance. Finally, it is prevalent in the wastewater literature to use a $\log_{10}$ transformation on the variables to meet assumptions for parametric analysis [38,39]; we compared taking the log transformation of wastewater measurements to not taking it, to validate this standard procedure. In results not shown here for the sake of brevity, we verified as well that because the N and O genes are highly correlated, it is not very important whether one or the other (or an average of the two) is used for prediction. We focus on using solely the O gene here. To summarize, we varied the following factors: 1. Normalization of the virus concentrations (unnormalized, flow, PMMoV or both), 2. The number of lagged values for the virus concentrations (0, 1 or 2), 3. Whether a $\log_{10}$ transformation was used on the dependent and independent variables. The three factors were 'fully crossed', resulting in a total of $4 \times 3 \times 2 = 24$ conditions. The models were fit separately for each WWTP in each condition (we will show, shortly, that different model fits are needed for different WWTPs). To allow for comparisons between WWTPs with different sizes of population served, we divided the case count and virus concentrations by the population size. This ensures that the prediction errors for all WWTPs are theoretically on the same scale. Observations with missing values due to the creation of lagged variables were removed. As was mentioned above, the models were compared by the averaged-over-WWTPs LOOCV.

2.3. Storage and Time-to-Analysis Stability Methods

We conducted a 'stability study' about stability of RNA in wastewater under different conditions. We used a large volume of WW from a single site that was partitioned into samples that were stored at varying conditions (4 °C, −20 °C, and −80 °C) for 2, 5, 7, or 14

days. Nucleic acid for each sample ($n = 4$) was then analyzed by qRT PCR concurrently (for a total of $4 \times 4 \times 5 = 80$ data points).

3. Results

3.1. Description of Surveillance Data

Figure 2 displays the concentrations of N, S and O in three WWTPs, Little Falls, Northfield and Twin Cities, which are chosen as representatives of smaller, midsized, and large population sizes, respectively (see Table 1 below for tabulated populations).

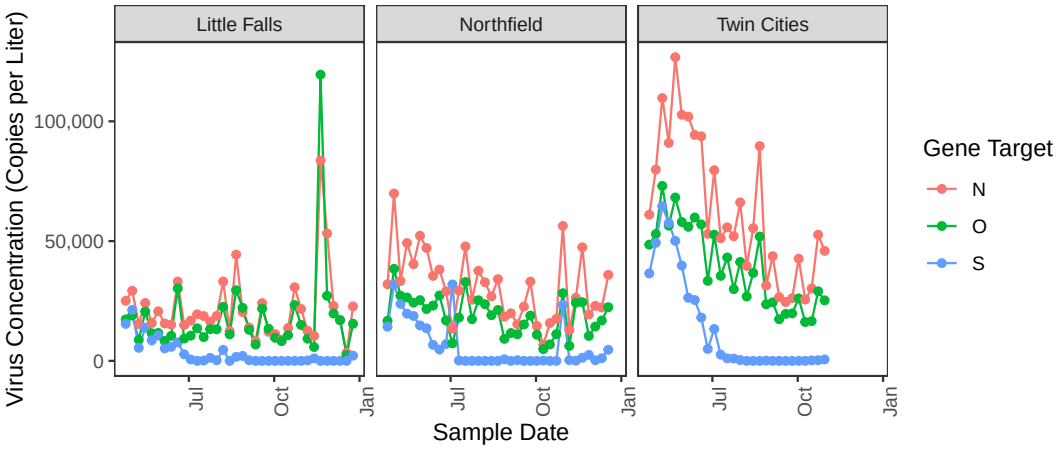

Figure 2. Virus concentrations in Little Falls (small), Northfield (midsized), and Twin Cities (large).

Table 1 provides some descriptions of the participating WWTPs, including the sampling period, the number of weeks where the WWTP was included in the study (because we had clinical case count data and had or could impute wastewater measurements), the size of the population served, as well as the sample mean of the weekly case count.

3.2. Use of SARS-CoV-2 Concentrations

Table 2 displays the (rescaled, for ease of presentation) RMSE (root mean squared prediction error) of the cross-validated RMSE of the linear regression models under different conditions across WWTPs. Lower indicates more accurate prediction. (Recall that the case count variable was divided by the corresponding catchment area population size). Overall the results in Table 2 suggest that taking logs is highly important for performance; this validates the standard practice of taking log transformations not just for parameter estimation but for prediction. We find that the normalization method is not as important especially when logs are taken. When the raw data were used, normalizing by flow alone slightly reduced the RMSE, whereas involving PMMoV in the normalization led to an increase in RMSE. Recall that PMMoV is defined by truncating the highest and lowest quantile values to mitigate the high variability in its measurement; but this was not enough to remove the high level of variability totally. Our interpretation, focusing especially on the non-logged models, is that PMMoV introduces extra variability which causes degradation. For instance, in our data when we normalize by PMMoV, the spike in wastewater RNA levels from the omicron surge that occurred in December 2021 in Minnesota does not appear, because the variability of the PMMoV measurements obfuscates it. Although in some contexts and for some WWTPs it seems that using PMMoV is helpful, our interpretation is that its high level of variability prevent it from being used reliably in normalization, at least at this time. On the other hand, our findings validate the use of flow in that even without a log transformation, normalization by flow is seen to achieve nearly optimal predictive performance.

Including extra lags generally worsened the predictive performance, perhaps surprisingly. These results are in the context of only having a few dozen observations per WWTP, but we argue this is common given that over a longer time horizon the parameters and

prediction models are likely to change in the face of evolving virus and immunity landscape. It is important to note that different models yielded different best fits for individual WWTPs; here we are presenting results averaged over all WWTPs. Results are affected both by high performing and low performing WWTPs. It is worth noting that in some of the 'higher performing' WWTPs, predictive performance was improved by using extra lags. For instance, in the EB, EP, HT, MH, MK, and TC WWTPs, including lagged variables improves performance according to LOOCV.

Overall, the model with the lowest RMSE over all WWTPs is the one where virus concentrations are normalized by flow, with $\log_{10}$ transformation, and without any lags:

$$\log_{10}(C_{t+1}) = \beta_0 + \beta_1 \log_{10}(W_t^{\text{Flow}}) + \epsilon \quad (1)$$

where β_j's are the regression coefficients, and ϵ is the residual. Although it is not clear from our results if the normalization by flow is truly important, this is the model to which we explore various extensions in the upcoming sections.

Table 2. RMSEs ($\times 1000$) of RMSE across WWTPs in all 24 conditions.

Normalization	Scale	Lag	RMSE
Unnormalized	Raw	0	0.72
Unnormalized	Raw	1	0.71
Unnormalized	Raw	2	0.83
Unnormalized	Log	0	0.66
Unnormalized	Log	1	0.67
Unnormalized	Log	2	0.82
Flow	Raw	0	0.64
Flow	Raw	1	0.66
Flow	Raw	2	0.72
Flow	Log	0	0.64
Flow	Log	1	0.64
Flow	Log	2	0.87
PMMoV	Raw	0	0.85
PMMoV	Raw	1	0.92
PMMoV	Raw	2	1.32
PMMoV	Log	0	0.64
PMMoV	Log	1	0.66
PMMoV	Log	2	0.69
Flow & PMMoV	Raw	0	0.85
Flow & PMMoV	Raw	1	0.99
Flow & PMMoV	Raw	2	1.41
Flow & PMMoV	Log	0	0.64
Flow & PMMoV	Log	1	0.66
Flow & PMMoV	Log	2	0.69

3.2.1. Heterogeneity in Fit and Individual WWTP Validation

It would be ideal if it were possible to develop one model using the pooled data from all WWTPs. Unfortunately our results suggest this is not possible; we compared the estimated error variance and the estimated regression coefficients of W_t^{Flow} from the model (1) across WWTPs. We found significantly different regression coefficients and variance estimates, see Table 3 (giving estimated regression coefficients and p-values of $\log_{10}(W_t^{\text{Flow}})$ for each WWTP) and Table 4 (giving estimated σ^2 values for each WWTP). We found that the estimated coefficients ranged from -0.06 to 0.66 (with summary statistics of mean = 0.28, SD = 0.23). The ratio of the largest to smallest variance estimate was larger than 10. This indicates that separate fits are needed for separate WWTPs. More broadly, it suggests the important finding that in any large-scale surveillance plan, measurements from different WWTPs (even when analyzed by the same lab) may not be directly comparable and may require different interpretation.

Table 3. Estimated Coefficients and p-values of $\log_{10}(W_t^{\text{flow}})$ for each WWTP.

Code	Estimated Coefficient	p-Value
AL	0.47	0.00
AX	0.13	0.06
BI	0.10	0.36
BL	0.48	0.00
BP	0.20	0.38
CB	−0.01	0.90
CM	0.09	0.48
CR	0.36	0.07
EB	0.21	0.03
EM	0.59	0.00
EP	0.61	0.00
ER	−0.02	0.72
FF	0.06	0.50
GC	0.03	0.65
GL	0.19	0.05
HS	0.28	0.04
HT	0.09	0.56
HY	0.16	0.17
IF	0.61	0.00
LB	−0.03	0.71
LF	0.14	0.18
LS	0.18	0.15
LY	−0.06	0.21
MH	0.13	0.02
MK	0.66	0.00
ML	0.08	0.56
MR	0.08	0.11
NB	0.31	0.00
NF	0.54	0.00
NP	0.35	0.01
RC	0.65	0.00
SD	0.34	0.00
SN	0.58	0.00
SV	0.59	0.01
TC	0.63	0.00
TR	0.19	0.05
WA	0.35	0.00
WL	−0.06	0.73
WM	0.41	0.01
WT	0.63	0.00

Table 4. Sigma squared estimates for each WWTP.

Code	$\hat{\sigma}^2$
AL	0.028
AX	0.023
BI	0.033
BL	0.012
BP	0.104
CB	0.040
CM	0.056
CR	0.054
EB	0.019

Table 4. *Cont.*

Code	$\hat{\sigma}^2$
EM	0.009
EP	0.010
ER	0.015
FF	0.044
GC	0.027
GL	0.048
HS	0.020
HT	0.028
HY	0.088
IF	0.066
LB	0.067
LF	0.037
LS	0.051
LY	0.022
MH	0.017
MK	0.008
ML	0.035
MR	0.028
NB	0.027
NF	0.044
NP	0.041
RC	0.023
SD	0.012
SN	0.007
SV	0.036
TC	0.008
TR	0.046
WA	0.030
WL	0.034
WM	0.027
WT	0.057

3.2.2. Use of Vaccination Rate

We examine how best to incorporate vaccination data in our linear regression models. Let V_t^{one} be the percentage of people who have received at least one doses of vaccine out of the population served by a WWTP, and V_t^{full} be the percentage of fully vaccinated people. The models to be investigated were built upon the model in Equation (1). In addition to the normalized virus concentration W_t^{Flow}, the models also considered the main effects of V_t^{one} and/or V_t^{full}, as well as two- or three-way interactions between W_t^{Flow}, V_t^{one} and V_t^{full}. The models are listed in Table 5, along with their means and standard deviations of the cross-validated RMSE across WWTPs. The model that includes the interaction between W_t^{Flow} and V_t^{full},

$$\log_{10}(C_{t+1}) = \beta_0 + \beta_1 \log_{10}(W_t^{\text{Flow}}) + \beta_2 \log_{10}(V_t^{\text{full}}) + \beta_3 \log_{10}(W_t^{\text{Flow}}) \cdot \log_{10}(V_t^{\text{full}}) + \epsilon, \quad (2)$$

minimizes the mean RMSE, but the differences are not large (and the results were not sensitive to whether V_t^{one} or V_t^{full} was used). The clearest result is that including too many variables leads to worse predictions. However, in the next subsection we see perhaps surprisingly that including vaccination may add more variability than it removes (especially when predicting two weeks ahead).

Table 5. Means and standard deviations (×1000) of RMSE across WWTPs for different vaccination linear models.

Formula	Mean	SD
$\log_{10}(C_{t+1}) \sim \log_{10}(W_t^{\text{flow}}) + \log_{10}(V_t^{\text{one}})$	0.55	0.21
$\log_{10}(C_{t+1}) \sim \log_{10}(W_t^{\text{flow}}) + \log_{10}(V_t^{\text{full}})$	0.54	0.23
$\log_{10}(C_{t+1}) \sim \log_{10}(W_t^{\text{flow}}) + \log_{10}(V_t^{\text{one}}) + \log_{10}(V_t^{\text{full}})$	2.47	12.26
$\log_{10}(C_{t+1}) \sim \log_{10}(W_t^{\text{flow}}) * \log_{10}(V_t^{\text{one}})$	0.54	0.20
$\log_{10}(C_{t+1}) \sim \log_{10}(W_t^{\text{flow}}) * \log_{10}(V_t^{\text{full}})$	0.53	0.22
$\log_{10}(C_{t+1}) \sim \log_{10}(W_t^{\text{flow}}) * \log_{10}(V_t^{\text{one}}) * \log_{10}(V_t^{\text{full}})$	149.12	938.86

3.2.3. Prediction Accuracy over Different Forecast Horizons

We assess the performance of predictions over different forecast horizons (same week, one week ahead and two weeks ahead). The model in Equation (2) was fitted separately for each WWTP, with $\log_{10}(C_t)$, $\log_{10}(C_{t+1})$ and $\log_{10}(C_{t+2})$ being the dependent variable. The means and standard deviations of the cross-validated RMSE are reported in the first three rows of Table 6. At the current noise level, we do not see a degradation in ability to predict one week ahead versus predicting the current week's case count. The predictions for two weeks ahead are substantially worse than the predictions for the first week. For comparisons, we fitted the model in Equation (1) again with different forecast horizons (Table 6). Table 7 gives results for models that exclude vaccinatino information. Interestingly, when predicting two weeks ahead, the simplest model (excluding vaccination information) does much better in terms of both mean and variance, which shows the value of parsimonious models.

Table 6. Means and SDs (×1000) of RMSE across WWTPs when predicting the case count of the same week, one week ahead and two weeks ahead, for models with vaccination.

Model	Mean	SD
$\log(C_t) \sim \log_{10}(W_t^{\text{flow}}) * \log_{10}(V_t^{\text{full}})$	0.54	0.21
$\log(C_{t+1}) \sim \log_{10}(W_t^{\text{flow}}) * \log_{10}(V_t^{\text{full}})$	0.53	0.22
$\log(C_{t+2}) \sim \log_{10}(W_t^{\text{flow}}) * \log_{10}(V_t^{\text{full}})$	0.75	1.47

Table 7. Means and SDs (×1000) of RMSE across WWTPs when predicting the case count of the same week, one week ahead and two weeks ahead, for models without vaccination.

	Model	Mean	SD
4	$\log(C_t) \sim \log(W_t^{\text{flow}})$	0.59	0.19
5	$\log(C_{t+1}) \sim \log(W_t^{\text{flow}})$	0.61	0.19
6	$\log(C_{t+2}) \sim \log(W_t^{\text{flow}})$	0.60	0.19

3.3. Storage and Time-to-Analysis Stability Study Findings

Although one might expect a priori for measurements from different WWTPs that are studied in the same lab to have broadly similar characteristics, we found above in the analysis of our surveillance data nontrivial heterogeneity in predictive ability (coefficient values and standard deviation estimates) between different WWTPs. Unmeasured variables such as population mobility, access to clinical testing centers, or population willingness to utilize clinical testing could account for some of the variation. Wastewater storage conditions are another possible cause. To shed light on what causes such unexplained large differences between WWTPs, we conducted a stability study, as described in the Materials and Methods section. The summary statistic means (SDs) for each temperature (4 °C, −20 °C, and −80 °C) [averaging across all times] of 39,400 (8560) for baseline, 18,600 (8600) for 4 °C, 2970 (5420) for −20 °C, and 3360 (1550) for −80 °C).

We found that storage temperature had a large and significant impact on the measured RNA. The length of time of storage had a less noticeable and not significant impact. We subtracted each baseline measurement from each storage measurement, yielding a negative number in all cases. We transformed the response by taking the negative and then the logarithm. We then fit a linear mixed regression model with no intercept term, a (linear) time trend term, a factor variable for the three storage conditions, and a random intercept effect for each sampling unit, with no interactions included; that is, we fit

$$\log W_{i,t} = \gamma_i + \sum_{j=1}^{3} \beta_j \times \text{TempIndicator}_j + \beta_4 t + \epsilon_{i,t} \quad (3)$$

for $i = 1, \ldots, 4$, $t \in \{2, 5, 7, 14\}$, TempIndicator_j an indicator/dummy variable for the three temperature storage conditions, γ_i i.i.d. $N(0, \sigma^2)$ random effects, $\sigma^2 > 0$, $\epsilon_{i,t}$ i.i.d. $N(0, \eta^2)$, and β_j the fixed effects of interest $(j = 1, 2, 3)$. The coefficient estimates for each of the three storage conditions were significant; the coefficients for -20 °C and -80 °C were not significantly different than each other, but are different than the coefficient for 4 °C. The negative exponentiated coefficient estimates (p-values) are $-16{,}200$ (2×10^{-12}) for 4 °C, $-32{,}300$ (1×10^{-14}) for -20 °C, and $-32{,}300$ (110^{-14}) for -80 °C. The time trend term was not significant. The means and standard deviations (over the four samples) at each of the 12 different conditions are presented in Table 8. The Day 0 average (sd) is 39,400 (8560). It is easy to see at a glance the effect of storing below 4 °C.

These data provide important guidelines for sample submission and storage whereby 4 °C stored samples showed less degradation than a -20 °C or -80 °C freeze and thaw. Under these principles we advocate long term storage at -80 °C and shipping temperatures under refrigeration (4 °C).

Table 8. Means (standard deviations) of RNA levels (copies/liter) over 4 samples for each of 4 storage lengths at 3 different temperatures. Samples were obtained in the Summer of 2023 over a 6 week period from a single source volume that was then partitioned into aliquots that were stored as indicated.

	Day 1	**Day 5**	**Day 7**	**Day 14**
4 °C	21,900 (6100)	22,000 (11,700)	15,900 (11,200)	14,500 (2700)
-20 °C	900 (200)	4700 (7600)	5600 (7900)	700 (500)
-80 °C	4000 (1100)	2800 (1300)	3600 (2300)	3000 (1600)

4. Discussion

We performed wastewater surveillance for SARS-CoV-2 across the state of Minnesota and undertook efforts to apply statistical analyses to inform how to best interpret and apply these data by studying the predictive relationship of wastewater virus levels to clinical case counts that can be used for public health decision making. For instance, it is clear that if clinical case counts demonstrate a large surge, then the ability of wastewater to demonstrate the same surge is important and beneficial; so in this sense a strong statistical relationship between the wastewater measurements and case counts would be meaningful. Our estimates of error variance and regression coefficients revealed substantial variability across WWTPs in our program. These findings suggest that the relationship between COVID-19 incidence in clinical patients and SARS-CoV-2 RNA in wastewater may be treatment plant-specific, and future work will need to continue investigating how to appropriately normalize data from different plants to allow for cross-plant comparisons. We acknowledge that in the current study, usage of areas with discrepant boundaries for defining case counts and wastewater levels could create data accuracy issues. Future work will assess whether using more refined location data for clinical case counts (rather than binning by zip codes) could aid in improving model performance.

The variation we observed is not accounted for by the size of the catchment area or the size of the corresponding population. This suggests that, at present, COVID-19 WWBE may need to be viewed at the individual plant level for its most impactful application as a regional metric for disease burden. Additionally, studies such as ours involving community level analyses represent challenges in that the available individual level data are rather limited, and the conclusions may not apply in an extremely data rich environment. However, in infectious disease modeling, we believe that true data rich environments (in which necessarily model parameters and immunological landscapes all stay constant over time) are exceedingly rare.

Prompted by variability observed in community-based studies, we set out to determine if we could identify actionable items that would boost accuracy (i.e., diminish variance). Our surveillance study period over the year of 2022 was characterized by mixed SARS-CoV-2 lineages and immunity from either/both exposure or primary and booster vaccines. Future models including vaccination data should adapt updated categories of vaccination classification, for example, a time-since-vaccination variable, to account for the waning of immunity over time. Further considerations are seasonal, regional, climatological, and sample temperature effects [40]. SARS-CoV-2 has year-round disease activity and as the pandemic has progressed, the levels and severity (i.e., hospitalization and deaths) has shown a trend toward higher incidence in the autumn and winter in both the Northern and Southern hemispheres [41,42]. We have not correlated large precipitation events with peaks or troughs in wastewater levels of SARS-CoV-2; however, other studies have observed that rainfall can have a diluting effect that can be calibrated using other chemical indicators [43]. This raises an important consideration for normalizing wastewater data that has been undertaken using both treatment center flow rates and fecal indicators [26,43–47]. Consistent with some studies in the wastewater literature [28,37,38], the current study finds that the predictive performance does not improve after involving PMMoV in the normalization. Interestingly, adding lagged virus concentrations to the model generally worsened the predictive performance. Using the vaccination rate in the model has been shown to improve the predictions. However, the results are not uniformly consistent across different forecast horizons. Due to the relatively low number of data points for each WWTP (Table 1), nonlinear or machine learning methods are not effective, and in fact our very parsimonious linear regression models performed best.

The temperature during the transit of wastewater to treatment sites can vary based on the location and infrastructure of the municipality and represents a variable that is difficult to control. Toward identifying temperature variables we could control, we asked whether the storage condition of the sample from which we quantify viral load would impact analytics. To answer this question, we acquired four large volume samples obtained across four different time periods and partitioned them into aliquots that were stored under room temperature, refrigerated, or frozen conditions for defined lengths of time (Table 8). We observed that freeze-thawing led to greater RNA degradation than maintaining the sample at 4 °C. This has important ramifications and provides guidance for the analytical community whereby samples can be stored onsite under refrigerated conditions and shipped on ice packs. These conditions are easily achievable for any geographic region as opposed to frozen storage that requires dedicated equipment for cold chain storage.

Our multiplex gene analysis has enabled some view of circulating variants due to the polymorphic nature of the spike gene/protein where gain and loss of the amino acids at positions 69 and 70 (H69V70) can lead to presence or absence of a S-gene signal by qRT PCR with the TaqPath™assay [48,49]. Efforts are underway in our lab to expand and more definitively define the wastewater population in regards to SARS-CoV-2 variants using genomic surveillance [37,50,51]. Such approaches can enable correlative relationships between a circulating variant or a mixed occurrence of variants and clinical specimen testing [52] and will enhance our statistical predictive accuracy.

Our program has operated continuously since 2020, and the data presented in this manuscript represent a broad geographic area. Our findings provide new guidance for

statistical modeling and prediction along with storage procedures for wastewater samples and are additive to the fields of epidemiology and pathogen surveillance.

Author Contributions: Conceptualization, C.R.D., M.J.O. and T.W.S.; methodology, C.R.D. and M.J.O.; software, C.R.D. and S.S.; validation, C.R.D. and M.J.O.; formal analysis, C.R.D.; investigation, all authors except C.R.D.; resources, M.J.O. and T.W.S.; data curation, C.R.D. and S.S.; writing—original draft preparation, C.R.D.; writing—review and editing, M.J.O. and T.W.S.; visualization, C.R.D. and S.S.; supervision, C.R.D., M.J.O. and T.W.S.; project administration, S.P., C.R.D., M.J.O. and T.W.S.; funding acquisition, T.W.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded, in part, by a grant from 3M and the Minnesota Department of Health.

Data Availability Statement: Data generated from the study can be found at https://z.umn.edu/ww_dashboard (accessed on 3 January 2024).

Acknowledgments: The authors wish to express their gratitude to Ruth Lynfield, Austin Bell, Lydia Fess and Sydney Kuramoto, as well as to the Metropolitan Council treatment facility.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

WWTP	WasteWater-Treatment Plant
WWBE	WasteWater-Based Epidemiology

References

1. Barasa, E.W.; Ouma, P.O.; Okiro, E.A. Assessing the hospital surge capacity of the Kenyan health system in the face of the COVID-19 pandemic. *PLoS ONE* **2020**, *15*, e0236308. [CrossRef] [PubMed]
2. Lauer, S.A.; Grantz, K.H.; Bi, Q.; Jones, F.K.; Zheng, Q.; Meredith, H.R.; Azman, A.S.; Reich, N.G.; Lessler, J. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Ann. Intern. Med.* **2020**, *172*, 577–582. [CrossRef] [PubMed]
3. Peccia, J.; Zulli, A.; Brackney, D.E.; Grubaugh, N.D.; Kaplan, E.H.; Casanovas-Massana, A.; Ko, A.I.; Malik, A.A.; Wang, D.; Wang, M.; et al. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nat. Biotechnol.* **2020**, *38*, 1164–1167. [CrossRef] [PubMed]
4. Li, R.; Pei, S.; Chen, B.; Song, Y.; Zhang, T.; Yang, W.; Shaman, J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **2020**, *368*, 489–493. [CrossRef] [PubMed]
5. Nishiura, H.; Kobayashi, T.; Miyama, T.; Suzuki, A.; Jung, S.m.; Hayashi, K.; Kinoshita, R.; Yang, Y.; Yuan, B.; Akhmetzhanov, A.R.; et al. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *Int. J. Infect. Dis.* **2020**, *94*, 154–155. [CrossRef] [PubMed]
6. Oran, D.P.; Topol, E.J. Prevalence of asymptomatic SARS-CoV-2 infection: A narrative review. *Ann. Intern. Med.* **2020**, *173*, 362–367. [CrossRef] [PubMed]
7. Post, L.A.; Issa, T.Z.; Boctor, M.J.; Moss, C.B.; Murphy, R.L.; Ison, M.G.; Achenbach, C.J.; Resnick, D.; Singh, L.N.; White, J.; et al. Dynamic public health surveillance to track and mitigate the US COVID-19 epidemic: Longitudinal trend analysis study. *J. Med Internet Res.* **2020**, *22*, e24286. [CrossRef]
8. Tran, H.N.; Le, G.T.; Nguyen, D.T.; Juang, R.S.; Rinklebe, J.; Bhatnagar, A.; Lima, E.C.; Iqbal, H.M.; Sarmah, A.K.; Chao, H.P. SARS-CoV-2 coronavirus in water and wastewater: A critical review about presence and concern. *Environ. Res.* **2021**, *193*, 110265. [CrossRef]
9. Zhang, D.; Duran, S.S.F.; Lim, W.Y.S.; Tan, C.K.I.; Cheong, W.C.D.; Suwardi, A.; Loh, X.J. SARS-CoV-2 in wastewater: From detection to evaluation. *Mater. Today Adv.* **2022**, *13*, 100211. [CrossRef]
10. Yousif, M.; Rachida, S.; Taukobong, S.; Ndlovu, N.; Iwu-Jaja, C.; Howard, W.; Moonsamy, S.; Mhlambi, N.; Gwala, S.; Levy, J.I.; et al. SARS-CoV-2 genomic surveillance in wastewater as a model for monitoring evolution of endemic viruses. *Nat. Commun.* **2023**, *14*, 6325. [CrossRef]
11. Hellmér, M.; Paxéus, N.; Magnius, L.; Enache, L.; Arnholm, B.; Johansson, A.; Bergström, T.; Norder, H. Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks. *Appl. Environ. Microbiol.* **2014**, *80*, 6771–6781. [CrossRef] [PubMed]

12. Manor, Y.; Handsher, R.; Halmut, T.; Neuman, M.; Bobrov, A.; Rudich, H.; Vonsover, A.; Shulman, L.; Kew, O.; Mendelson, E. Detection of poliovirus circulation by environmental surveillance in the absence of clinical cases in Israel and the Palestinian authority. *J. Clin. Microbiol.* **1999**, *37*, 1670–1675. [CrossRef] [PubMed]
13. Ye, Y.; Ellenberg, R.M.; Graham, K.E.; Wigginton, K.R. Survivability, partitioning, and recovery of enveloped viruses in untreated municipal wastewater. *Environ. Sci. Technol.* **2016**, *50*, 5077–5085. [CrossRef] [PubMed]
14. Chan, V.W.S.; Chiu, P.K.F.; Yee, C.H.; Yuan, Y.; Ng, C.F.; Teoh, J.Y.C. A systematic review on COVID-19: Urological manifestations, viral RNA detection and special considerations in urological conditions. *World J. Urol.* **2021**, *39*, 3127–3138. [CrossRef] [PubMed]
15. Chen, Y.; Chen, L.; Deng, Q.; Zhang, G.; Wu, K.; Ni, L.; Yang, Y.; Liu, B.; Wang, W.; Wei, C.; et al. The presence of SARS-CoV-2 RNA in the feces of COVID-19 patients. *J. Med. Virol.* **2020**, *92*, 833–840. [CrossRef]
16. Cheung, K.S.; Hung, I.F.; Chan, P.P.; Lung, K.; Tso, E.; Liu, R.; Ng, Y.; Chu, M.Y.; Chung, T.W.; Tam, A.R.; et al. Gastrointestinal manifestations of SARS-CoV-2 infection and virus load in fecal samples from a Hong Kong cohort: Systematic review and meta-analysis. *Gastroenterology* **2020**, *159*, 81–95. [CrossRef] [PubMed]
17. Parasa, S.; Desai, M.; Chandrasekar, V.T.; Patel, H.K.; Kennedy, K.F.; Roesch, T.; Spadaccini, M.; Colombo, M.; Gabbiadini, R.; Artifon, E.L.; et al. Prevalence of gastrointestinal symptoms and fecal viral shedding in patients with coronavirus disease 2019: A systematic review and meta-analysis. *JAMA Netw. Open* **2020**, *3*, e2011335. [CrossRef]
18. Wong, M.C.; Huang, J.; Lai, C.; Ng, R.; Chan, F.K.; Chan, P.K. Detection of SARS-CoV-2 RNA in fecal specimens of patients with confirmed COVID-19: A meta-analysis. *J. Infect.* **2020**, *81*, e31–e38. [CrossRef]
19. Melvin, R.G.; Hendrickson, E.N.; Chaudhry, N.; Georgewill, O.; Freese, R.; Schacker, T.W.; Simmons, G.E., Jr. A novel wastewater-based epidemiology indexing method predicts SARS-CoV-2 disease prevalence across treatment facilities in metropolitan and regional populations. *Sci. Rep.* **2021**, *11*, 21368. [CrossRef]
20. Ahmed, W.; Angel, N.; Edson, J.; Bibby, K.; Bivins, A.; O'Brien, J.W.; Choi, P.M.; Kitajima, M.; Simpson, S.L.; Li, J.; et al. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. *Sci. Total Environ.* **2020**, *728*, 138764. [CrossRef]
21. Arora, S.; Nag, A.; Sethi, J.; Rajvanshi, J.; Saxena, S.; Shrivastava, S.K.; Gupta, A.B. Sewage surveillance for the presence of SARS-CoV-2 genome as a useful wastewater based epidemiology (WBE) tracking tool in India. *Water Sci. Technol.* **2020**, *82*, 2823–2836. [CrossRef] [PubMed]
22. Randazzo, W.; Truchado, P.; Cuevas-Ferrando, E.; Simón, P.; Allende, A.; Sánchez, G. SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area. *Water Res.* **2020**, *181*, 115942. [CrossRef] [PubMed]
23. Nattino, G.; Castiglioni, S.; Cereda, D.; Della Valle, P.G.; Pellegrinelli, L.; Bertolini, G.; Pariani, E. 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*, 1922–1924. [CrossRef] [PubMed]
24. Wani, H.; Menon, S.; Desai, D.; D'Souza, N.; Bhathena, Z.; Desai, N.; Rose, J.B.; Shrivastava, S. Wastewater-Based Epidemiology of SARS-CoV-2: Assessing Prevalence and Correlation with Clinical Cases. *Food Environ. Virol.* **2023**, *15*, 131–143. [CrossRef] [PubMed]
25. Hasan, S.W.; Ibrahim, Y.; Daou, M.; Kannout, H.; Jan, N.; Lopes, A.; Alsafar, H.; Yousef, A.F. Detection and quantification of SARS-CoV-2 RNA in wastewater and treated effluents: Surveillance of COVID-19 epidemic in the United Arab Emirates. *Sci. Total Environ.* **2021**, *764*, 142929. [CrossRef] [PubMed]
26. Weidhaas, J.; Aanderud, Z.T.; Roper, D.K.; VanDerslice, J.; Gaddis, E.B.; Ostermiller, J.; Hoffman, K.; Jamal, R.; Heck, P.; Zhang, Y.; et al. Correlation of SARS-CoV-2 RNA in wastewater with COVID-19 disease burden in sewersheds. *Sci. Total Environ.* **2021**, *775*, 145790. [CrossRef] [PubMed]
27. Zhan, Q.; Babler, K.M.; Sharkey, M.E.; Amirali, A.; Beaver, C.C.; Boone, M.M.; Comerford, S.; Cooper, D.; Cortizas, E.M.; Currall, B.B.; et al. Relationships between SARS-CoV-2 in wastewater and COVID-19 clinical cases and hospitalizations, with and without normalization against indicators of human waste. *ACS ES&T Water* **2022**, *2*, 1992–2003.
28. Maal-Bared, R.; Qiu, Y.; Li, Q.; Gao, T.; Hrudey, S.E.; Bhavanam, S.; Ruecker, N.J.; Ellehoj, E.; Lee, B.E.; Pang, X. 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. [CrossRef]
29. Hamza, H.; Rizk, N.M.; Gad, M.A.; Hamza, I.A. Pepper mild mottle virus in wastewater in Egypt: A potential indicator of wastewater pollution and the efficiency of the treatment process. *Arch. Virol.* **2019**, *164*, 2707–2713. [CrossRef]
30. Kitajima, M.; Iker, B.C.; Pepper, I.L.; Gerba, C.P. Relative abundance and treatment reduction of viruses during wastewater treatment processes—Identification of potential viral indicators. *Sci. Total Environ.* **2014**, *488*, 290–296. [CrossRef]
31. Kitajima, M.; Sassi, H.P.; Torrey, J.R. Pepper mild mottle virus as a water quality indicator. *NPJ Clean Water* **2018**, *1*, 1–9. [CrossRef]
32. Rosario, K.; Symonds, E.M.; Sinigalliano, C.; Stewart, J.; Breitbart, M. Pepper mild mottle virus as an indicator of fecal pollution. *Appl. Environ. Microbiol.* **2009**, *75*, 7261–7267. [CrossRef]
33. Zhang, T.; Breitbart, M.; Lee, W.H.; Run, J.Q.; Wei, C.L.; Soh, S.W.L.; Hibberd, M.L.; Liu, E.T.; Rohwer, F.; Ruan, Y. RNA viral community in human feces: Prevalence of plant pathogenic viruses. *PLoS Biol.* **2006**, *4*, e3. [CrossRef]
34. Khan, K.; Tighe, S.W.; Badireddy, A.R. Factors influencing recovery of SARS-CoV-2 RNA in raw sewage and wastewater sludge using polyethylene glycol–based concentration method. *J. Biomol. Tech. JBT* **2021**, *32*, 172. [CrossRef] [PubMed]

35. Redlands, C. Environmental Systems Research Institute. ArcGIS Online, ESRI Demographics. Available online: https://www.arcgis.com (accessed on 1 November 2023).
36. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.
37. Duvallet, C.; Wu, F.; McElroy, K.A.; Imakaev, M.; Endo, N.; Xiao, A.; Zhang, J.; Floyd-O'Sullivan, R.; Powell, M.M.; Mendola, S.; et al. Nationwide Trends in COVID-19 Cases and SARS-CoV-2 RNA Wastewater Concentrations in the United States. *ACS ES&T Water* **2022**, *2*, 1899–1909.
38. Feng, S.; Roguet, A.; McClary-Gutierrez, J.S.; Newton, R.J.; Kloczko, N.; Meiman, J.G.; McLellan, S.L. Evaluation of sampling, analysis, and normalization methods for SARS-CoV-2 concentrations in wastewater to assess COVID-19 burdens in Wisconsin communities. *Acs Es&T Water* **2021**, *1*, 1955–1965.
39. Farkas, K.; Pellett, C.; Alex-Sanders, N.; Bridgman, M.T.; Corbishley, A.; Grimsley, J.M.; Kasprzyk-Hordern, B.; Kevill, J.L.; Pântea, I.; Richardson-O'Neill, I.S.; et al. Comparative assessment of filtration-and precipitation-based methods for the concentration of SARS-CoV-2 and other viruses from wastewater. *Microbiol. Spectr.* **2022**, *10*, e01102-22. [CrossRef] [PubMed]
40. McMahan, C.S.; Self, S.; Rennert, L.; Kalbaugh, C.; Kriebel, D.; Graves, D.; Colby, C.; Deaver, J.A.; Popat, S.C.; Karanfil, T.; et al. COVID-19 wastewater epidemiology: A model to estimate infected populations. *Lancet Planet. Health* **2021**, *5*, e874–e881. [CrossRef] [PubMed]
41. Wiemken, T.L.; Khan, F.; Puzniak, L.; Yang, W.; Simmering, J.; Polgreen, P.; Nguyen, J.L.; Jodar, L.; McLaughlin, J.M. Seasonal trends in COVID-19 cases, hospitalizations, and mortality in the United States and Europe. *Sci. Rep.* **2023**, *13*, 3886. [CrossRef] [PubMed]
42. Meo, S.A.; Alsomali, A.H.; Almushawah, A.A.; Halepoto, D.M. Seasonal variations impact on SARS-CoV-2 incidence and mortality in southern and northern hemispheres: Two years pandemic period based study. *J. King Saud Univ. Sci.* **2022**, *34*, 102335. [CrossRef] [PubMed]
43. Saingam, P.; Li, B.; Quoc, B.N.; Jain, T.; Bryan, A.; Winkler, M.K. Wastewater surveillance of SARS-CoV-2 at intra-city level demonstrated high resolution in tracking COVID-19 and calibration using chemical indicators. *Sci. Total Environ.* **2023**, *866*, 161467. [CrossRef]
44. D'Aoust, P.M.; Mercier, E.; Montpetit, D.; Jia, J.J.; Alexandrov, I.; Neault, N.; Baig, A.T.; Mayne, J.; Zhang, X.; Alain, T.; et al. Quantitative analysis of SARS-CoV-2 RNA from wastewater solids in communities with low COVID-19 incidence and prevalence. *Water Res.* **2021**, *188*, 116560. [CrossRef]
45. Li, B.; Di, D.Y.W.; Saingam, P.; Jeon, M.K.; Yan, T. Fine-scale temporal dynamics of SARS-CoV-2 RNA abundance in wastewater during a COVID-19 lockdown. *Water Res.* **2021**, *197*, 117093. [CrossRef] [PubMed]
46. Medema, G.; Been, F.; Heijnen, L.; Petterson, S. Implementation of environmental surveillance for SARS-CoV-2 virus to support public health decisions: Opportunities and challenges. *Curr. Opin. Environ. Sci. Health* **2020**, *17*, 49–71. [CrossRef] [PubMed]
47. Wilder, M.L.; Middleton, F.; Larsen, D.A.; Du, Q.; Fenty, A.; Zeng, T.; Insaf, T.; Kilaru, P.; Collins, M.; Kmush, B.; et al. Co-quantification of crassphage increases confidence in wastewater-based epidemiology for SARS-CoV-2 in low prevalence areas. *Water Res. X* **2021**, *11*, 100100. [CrossRef] [PubMed]
48. Migueres, M.; Lhomme, S.; Trémeaux, P.; Dimeglio, C.; Ranger, N.; Latour, J.; Dubois, M.; Nicot, F.; Miedouge, M.; Mansuy, J.; et al. Evaluation of two RT-PCR screening assays for identifying SARS-CoV-2 variants. *J. Clin. Virol.* **2021**, *143*, 104969. [CrossRef] [PubMed]
49. Shahhosseini, N.; Babuadze, G.; Wong, G.; Kobinger, G.P. Mutation signatures and in silico docking of novel SARS-CoV-2 variants of concern. *Microorganisms* **2021**, *9*, 926. [CrossRef] [PubMed]
50. Brunner, F.S.; Payne, A.; Cairns, E.; Airey, G.; Gregory, R.; Pickwell, N.D.; Wilson, M.; Carlile, M.; Holmes, N.; Hill, V.; et al. Utility of wastewater genomic surveillance compared to clinical surveillance to track the spread of the SARS-CoV-2 Omicron variant across England. *Water Res.* **2023**, *247*, 120804. [CrossRef] [PubMed]
51. Espinosa-Gongora, C.; Berg, C.; Rehn, M.; Varg, J.E.; Dillner, L.; Latorre-Margalef, N.; Székely, A.J.; Andersson, E.; Movert, E. Early detection of the emerging SARS-CoV-2 BA. 2.86 lineage through integrated genomic surveillance of wastewater and COVID-19 cases in Sweden, weeks 31 to 38 2023. *Eurosurveillance* **2023**, *28*, 2300595. [CrossRef]
52. Li, L.; Uppal, T.; Hartley, P.D.; Gorzalski, A.; Pandori, M.; Picker, M.A.; Verma, S.C.; Pagilla, K. Detecting SARS-CoV-2 variants in wastewater and their correlation with circulating variants in the communities. *Sci. Rep.* **2022**, *12*, 16141. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.