Public Health Agency of Canada | Agence de la santé publique du Canada

Normalization practices for SARS-CoV-2 data in wastewater-based epidemiology

A Technical Report prepared for PHAC

Thomas Maere, Jean-David Therrien and Peter VanRolleghem

Université Laval

Canada

FACULTÉ DES SCIENCES ET DE GÉNIE

Département de génie civil et de génie des eaux

Cité universitaire
Québec, Canada, G1V 0A6

Technical report

Normalization practices for SARS-CoV-2 data in wastewater-based epidemiology

AUTHORS/COAUTHORS	TEAM/GROUP
Thomas Maere	model*EAU*
Jean-David Therrien	model*EAU*
Peter Vanrolleghem	model*EAU*

WRITTEN BY		LATEST REVISION BY	
Name	Thomas Maere	**Name**	Peter Vanrolleghem
Date	31/03/2022	**Date**	31/03/2022

DOCUMENT FILE NAME	Normalization practices for SARS-CoV-2 data in wastewater-based epidemiology - Technical Report
# PAGES	16
REVISION NO.	1

model*EAU*, Département de génie civil
et de génie des eaux – Université Laval
Pavillon Adrien-Pouliot – local 2974
1605 Avenue de la Médecine
Québec (QC) G1V 0A6, Canada

Téléphone: +1 (418) 656-5085
Télécopieur: +1 (418) 656-2928

peter.vanrolleghem@gci.ulaval.ca
http://modelEAU.fsg.ulaval.ca

INDEX

LIST OF TABLES

1. PROBLEM STATEMENT AND OBJECTIVES

The amount of genetic SARS-CoV-2 material found in wastewater samples, generally expressed as gene copies per ml, might not always allow for a reliable assessment of the epidemiological situation in the sewershed. It is essential to understand the different types of factors that might influence wastewater measurements of SARS and the possible ways to cope with these factors. This report aims to give a structured overview of these influencing factors and discuss the various coping strategies and techniques that are generally addressed in literature as "normalization".

2. NATURE OF THE VIRAL SIGNAL IN WASTEWATER

The public health clinical case data for COVID-19 in larger cities during an epidemiological wave have a characteristically smooth behaviour consistent with epidemiological theory and infection model simulations. The case data of a wave typically follow a bell-shaped curve where the value on a particular day is determined mainly by the value on the previous day and the current trend in infections. On average, the wastewater viral signal correlates reasonably well with the long-term infection trends. Still, it has a specific short-term spikey behaviour that is generally not seen in clinical case data. It is still a matter of study if these spikes are real (e.g. due to variability in fecal shedding of viral material per person, variability in the timing and location of cases in the sewershed, population mobility, super spreading events) or unwanted variations (e.g. related to processes happening in the sewer, sampling errors or measurement issues). In the case of big cities and a well-established epidemiological wave, one can reasonably expect that the real causes of variability are mostly evened out following the law of large numbers for stochastic events, leaving mostly the unwanted variability related to various sewer processes and sampling and measurement errors. For small sewersheds or low clinical case numbers, the distinction between real and unwanted variation is much more challenging to make, however. Even more so since in these conditions, with only minute quantities of genetic material in the wastewater, the employed biomolecular techniques are pushed to their limit and produce less reliable results.

3. THE CONCEPT OF NORMALIZATION IN WASTEWATER-BASED EPIDEMIOLOGY

In data science, normalization refers to transforming or scaling all dataset values into the [0, 1] range. Z-score normalization or standardization rescales the data to have the properties of a standard normal distribution ($\mu=0$ and $\sigma=1$). In the dictionary of standard epidemiology, the term normalization doesn't exist. In contrast, the term standardization refers to a set of techniques based on weighted averaging, used to remove as much as possible the effects of differences in age or other confounding variables in comparing two or more populations. In current wastewater-based epidemiology (WBE) literature, the use of the term "normalization" is ambiguous and covers various topics. It generally includes the use of some type of auxiliary parameter (e.g., sewershed size, flow, …) or biomarker (e.g., wastewater strength, pharmaceuticals, crAssphage, PMMoV, …) that is measured alongside the variable of interest (i.e., SARS-CoV-2) and used to assess, correct or scale the value of that variable.

The measurement of SARS-CoV-2 involves many steps, and errors can occur at each stage of the sampling-transport-extraction-amplification chain. The use of different techniques for each step in

this measurement chain significantly impacts the obtained measurement results as well. The measurement of a variable that is supposed to be constant in wastewater discharge for the population served and subjected to all the same measurement steps as SARS-CoV-2, can be used as an internal control standard and give a first indication of the validity of the SARS-CoV-2 values. Depending on the actual result and the expected value range of the biomarker, one can decide to accept or reject the measured SARS-CoV-2 value. The application of a biomarker to check the validity of SARS-CoV-2 measurements in the lab is generally not considered a type of normalization but rather a part of the quality assurance and control procedure (QA/QC; Shik et al., 2022) required to produce reliable data. However, when applying a correction procedure to the SARS-CoV-2 signal based on the expected values of a biomarker, WBE literature often talks about normalization. Here, normalization is defined as a technique to correct unwanted variability in the viral wastewater signal by comparing the detected viral material to another substance excreted by humans whose daily loading at the sewershed level is known or stable over time.

The idea of using a biomarker is that it can describe one or a combination of multiple factors that influence the signal of interest. The behaviour of the biomarker should be well known throughout its journey from the moment of shedding until its measurement in the lab. Furthermore, its behaviour should resemble the behaviour of the target of interest for the factor of influence one wants to correct. It helps if the load of the biomarker into the sewershed is known precisely, but it is not a prerequisite if the load appears constant (i.e., a constant load is expected and the data show it). The biomarker can be of the same nature as the target of interest (e.g., genetic material of another virus in the case of SARS-CoV-2), but this is not necessary if the factor that influences the viral signal affects other types of compounds equally (e.g., wastewater strength to describe dilution).

Literature shows that comparing SARS-CoV-2 wastewater data from different studies is challenging. Not only are there differences in the epidemiological situation and its management in various cities and countries, but each sewershed has its peculiarities, and the techniques used for each study are different. In this context, several approaches are mentioned to better compare data between studies. The comparison of diverse sampling and measurement techniques has been the subject of several studies and gives an idea of their equivalence. The differences between labs employing the same methods could be assessed through ring testing aliquots of the same sample. Currently, no standard measurement technique nor standard QA/QC procedure exists. The influence of local sewershed characteristics (e.g., type, size, age) is still rather vague, but it is assumed that the results must be expressed per person equivalent to make any comparison possible. When dynamic population estimates are used to account for variations in population in the sewershed - which can also be done using biomarkers - literature uses the term “population normalization”. Combining the results of several regions into one metric is called aggregation.

4. DESCRIPTION OF THE INFLUENCING FACTORS ALONG THE WBE PATHWAY

4.1. Shedding dynamics

A considerable amount of people will shed viral material through their feces once infected. However, the exact percentage of people who shed is unknown, just as it is unknown who will shed, how much one will shed, and how long. One can hypothesize that this depends on individual genetics, age and immunization status, the COVID-19 variant in question and the severity of the infection (Hoffmann and Alsing, 2021). An additional source of uncertainty is related to the timing and location of each

shedding event (i.e., 5-10 sec toilet flush), which can be considered infrequent, discrete pulse loads of virus into the sewer system at certain moments in the week.

There are no known normalization techniques available to correct for the uncertainty related to shedding dynamics. Hence, this shedding uncertainty will hamper all other types of analysis with wastewater data, such as finding the best correction procedure for other factors influencing measurement variability. Through shedding and sewer modelling, it is possible to understand the range of uncertainty that shedding dynamics bring about and in which conditions its effects are critical (Wade et al., 2022). As mentioned before, with the law of large numbers, one would expect less influence of shedding dynamics for bigger cities with lots of cases and considerable dispersion and mixing in the sewer pipes.

4.2. Sewage origin, transport and processes

The amount and composition of sewage vary in time with diurnal, daily, seasonal and yearly dynamics. Sewage consists of a liquid and particulate fraction, each with a myriad of biological and chemical components. Both fraction's transport and dispersion dynamics in the sewer are different as particles can settle and resuspend depending on the flow regime. Sewage and sewersheds differ from city to city. Distinctions can be made based on sewage origin (i.e., domestic, industrial), sewershed type (separate, combined or hybrid storm drain and sewage pipe layouts), size (affecting the hydraulic and particulate retention time), location (temperature), topography (gravitational or pumped flows), age (groundwater infiltration through leaks), construction rules (flowrates, combined sewer overflows, retention tanks). All the above factors lead to varying levels of impact that the different physical, chemical and biological processes in sewers can have on the measured SARS-CoV-2 signal. While normalization procedures can alleviate the influence of some of these processes (e.g., dilution by rain, groundwater or snowmelt; dispersion and mixing), it might not necessarily be the case for other factors (e.g., genetic material degradation) for which the importance should be quantified through targeted studies and sewer modelling.

4.3. Sampling

The representativity of the obtained results greatly depends on the chosen sampling methodology. WBE studies for COVID-19 show that the full range of sampling types are being used (grab, time- and flow composite, passive sampling). From experience with tracking contaminants of emerging concern (CEC) in wastewater (Ort et al., 2010), it is recognized that a well-executed flow-proportional composite sampling strategy (i.e., well-placed intake, frequent sub-samples, large sample volume, refrigeration) will give the most realistic view on the total viral loads coming through the sewer network. While frequent grab sampling or time-proportional composite sampling might provide a good indication of the situation for SARS-CoV-2, these strategies entail structural biases that cannot be solved by biomarker normalization since that biomarker would have to exhibit the exact same dynamic pattern as the compound of interest. The more erratic, dynamic, and exaggerated the diurnal pattern of the compound of interest is, the greater the bias may be. As such, upstream catchment locations will be more susceptible to sampling errors given the reduced mixing time and dispersion of the pulse loads of virus into the total flow of wastewater. Sewer modelling can estimate the range of bias and variability expected in small to large catchments throughout an epidemic wave.

As the concentration fluctuates throughout the day, infrequent grab samples could indicate either very high or very low values depending on sampling time. In upstream catchments with only a few

cases, a viral pulse might be missed altogether. In these situations, and when it is infeasible to install composite samplers, a passive sampling strategy might be helpful to indicate the presence or absence of cases. Several studies indicate the semi-quantitative nature of passive sampling for SARS-CoV-2 with higher measured values related to higher wastewater concentrations and longer exposure times. However, back-calculation to the original concentration or load proves difficult as the sorption and desorption kinetics are not well defined (yet). Long exposures over multiple days also increase the risk of saturation of the passive sorption material (Schang et al., 2021). Biomarker normalization might help to improve the quantitative nature of passive sampling, but only if the biomarker behaves similarly.

An important last aspect of the sampling strategy is the target to be sampled. While most studies target the whole wastewater matrix, some studies focus on solids and seek those solids in the sludge underflow of the primary clarifier. The latter seems to improve the detection limit as SARS-CoV-2 tends to be mainly associated with solids. However, it comes with the complication of back-calculating the results to daily loads on catchment level. Primary clarifiers typically manage to collect about half of the solids in the raw influent (unless chemically enhanced), but their particle separation and concentration performance is dynamic and depends on diurnal flow dynamics. Normalization through solids-related biomarkers and primary clarifier modelling should be helpful to some extent.

4.4. Analysis

As mentioned earlier, neither standard SARS-CoV-2 measurement technique nor standard QA/QC procedure for wastewater currently exists. Differences in results have been reported for different storage, extraction and analysis methods, making it difficult to compare results. One of the more significant sources of variation is the targeted fraction of the wastewater matrix. Samples are centrifuged and filtered to various degrees, leading to different combinations of the original fractions present in wastewater (i.e., solutes, small particles, bigger particles). These varying fractions may alter the relation of the obtained results to the actual viral content in the original samples and likely also impact the efficiency and repeatability of the RNA extraction and amplification steps.

Furthermore, one must keep in mind the variable degree to which factors along the WBE pathway (shedding, sewer, lab) can affect each wastewater fraction (e.g., particle sedimentation and resuspension in sewers). The equivalence and standardization of procedures should be the subject of dedicated studies and can be enforced through ring testing and certification. Using a biomarker as internal control substance could help increase comparability if its partitioning in the wastewater matrix is known and stable and if it reacts in similar ways to the target of interest throughout the measurement procedure. Trust in the data from each lab can be increased through meticulous consistency in the used methodology, ample technical and biological replicates, and high frequency (daily) measurements that make time-series analysis possible.

5. DIFFERENCES BETWEEN WASTEWATER-BASED AND CLINICAL COVID-19 DATA

To assess the quality of the collected wastewater data and the validity of applied normalization procedures, the raw or corrected wastewater data are generally compared to the available clinical COVID-19 data. This comparison is made either visually or by using of some kind of correlation metric. There are several differences in the nature of the wastewater and clinical data that one has

to keep in mind. First, the clinical dataset is of the binary data type. That is, a clinical data point only has two possible values: “infected” or “not-infected”. In contrast, viral wastewater data can be considered semi-continuous (i.e. continuous behaviour at large concentrations and discrete behaviour at low concentrations). Part of the infected people periodically shed large amounts of virus in the wastewater. This shedding results in a semi-continuous range of concentrations at sampling points further down the sewer system through mixing and dispersion. When many COVID-19 cases are present within a larger city-scale sewershed, the time series values for both data sources will show a relatively smooth continuous pattern. However, with only a small number of COVID-19 cases, the binary nature of the public health data will become apparent, and the wastewater results will start to show more considerable variability. The latter is not only related to the stochastic shedding process (quantity, timing, location) but also the stochastic nature of the sampling process. The chance of having a certain number of viral gene copies in a sample is Poisson-distributed. Stochasticity thus becomes non-negligible when approaching the limit of quantification (LOQ) and detection (LOD) of the employed biomolecular techniques, resulting in a wastewater viral signal error that is inversely related to its absolute value (Wade et al., 2022).

When comparing the wastewater viral signal to clinical case data, one has several public health metrics to choose from, such as new cases, active cases, positivity rate, hospitalizations and deaths. There is no one golden standard that can be considered the truth. Normalization aims to connect the wastewater data to the truth, however, at times the truth is unknown. In some instances, the wastewater signal might even be a better indicator of the current epidemiological situation in the population. Firstly, public health testing for COVID-19 has seldom been a randomly executed process. It depends on local public health recommendations (that vary over time) and generally focuses on people with symptoms, people in contact with confirmed COVID-19 cases through contact tracing, and people at risk. The positivity rate can give an indication of the randomness of the screening strategy, whereby the WHO assumes values above 10% to be indicative of insufficiently comprehensive testing. Hospitalizations, intensive care admissions and deaths should, in principle, not be affected by the type of public health screening scheme or the willingness of people to get tested. However, differences can be expected based on the immunity status (vaccinated and recovered), age distribution of the population, hospital admission policy (because of COVID or with COVID; disease severity) and quality of care. However, the information that these parameters deliver trails behind the actual COVID transmission status within the population and lags considerably compared to the wastewater data. The amount of lag depends on the COVID variant and differs for each case and has, as such, a particular stochastic nature to it that will become more apparent in low-case situations. This further complicates the comparison between public health and wastewater data. The active case parameter is flawed as it requires continued surveillance and regular testing of each case. When this testing capacity is lacking, a new case is often considered active for a standard amount of time. The new case parameter is the most frequently used in literature to compare with wastewater data, which makes sense as it is the parameter with the smallest lag. The onset of symptoms should, in principle, be an earlier indicator than newly detected cases, but it is a less reliable parameter as it depends on declarations of patients and because a considerable amount of people is asymptomatic. In fact, studies have shown that with an extensive enough screening effort, the lag between active cases and wastewater data is negligible. Since the literature on fecal shedding indicates a peak release of the virus at the beginning of the infection, this also explains why a high correlation between wastewater and new cases can be expected (Hoffmann and Alsing, 2021).

The above-mentioned issues cannot be solved by any normalization procedure, but should be taken into consideration when analyzing WBE datasets, i.e., by assessing the suitability of the selected datasets and envisioned analysis procedures for answering the research question at hand.

6. NORMALIZATION IN PRACTICE IN THE CANADIAN CONTEXT

When looking at the public WBE dashboards across Canada available through the COVIDPoops19 website (last checked on 22 March 2022), one notices a mix of ways to present SARS-CoV-2 results (Table 1). At the beginning of the pandemic, most dashboards offered the option to view the raw daily results for one or both N1 and N2 genes, expressed in gene copies per ml of wastewater (gc/ml). More recently, however, several dashboards have removed access to the raw data and only show transformed data. This change indicates a growing consensus that raw viral data (gc/ml) might not give the best possible picture of the epidemiological situation. The more common data transformations are: taking a 3 to 7-day (moving) average; averaging or summing the N1 and N2 gene results; multiplying by flow; dividing by PMMoV, plotting on a log10 scale. Among the less common transformations, one finds dedicated smoothing algorithms, extraction efficiency adjustments, and data aggregation from several locations. E-mail responses from various Canadian research groups on normalization mentioned the use of catchment population numbers to execute population normalization and expression of the viral signal per “wet mass” of solids.

Smoothing the viral signal makes sense where one expects a smooth behaviour in line with epidemiological theory (i.e., for larger catchments and well-developed community transmission). However, a smooth signal is not a priori expected (i.e., small catchments, low-case numbers). The methods employed in Canada range from simple weekly averaging to moving mid-point averaging over different periods as well as dedicated smoothing algorithms. Selecting the best method and the calibration of that method are likely case-specific as they depend on sewershed characteristics, measurement frequency, and other factors. However, the selection procedure could be standardized, as in Arabzadeh et al. (2021), who tested 13 different data filtering techniques within a defined optimization framework with performance metrics to reduce noise in the signal and improve correlation with clinical case data.

Targeting multiple genes will certainly improve confidence in the viral data. The results from each gene can be compared to the others and help highlight any specific gene assay issues. When the results for different gene assays are expected to be similar, averaging or summing the results for these genes is expected to reduce noise in the data. When different assays show consistent differences, it is probably advisable to analyze the results for each gene separately. The use of extraction efficiency by means of an external viral control standard to correct the viral signal is not commonly executed in literature since the obtained recovery rates in wastewater are generally very low and variable. This would result in large and uncertain correction factors, which should be handled with care. More common is the use of a range of expected recoveries to detect outliers that might require a re-analysis of the sample.

The use of log10 scaling is in accordance with CDC guidelines, just as the use of flow, population numbers and biomarkers. The use of concepts like “wet mass” should probably be avoided in the context of standardizing measurement methods, as it is not clearly defined and likely variable from location to location. Multiplying viral concentrations from a 24h flow-proportional sample with the corresponding average daily flow gives the daily load of virus that passes at that sampling location. Remark that any other type of sampling for this calculation would result in biased daily viral loads. The recommendation that it is better to use daily viral loads than concentrations for comparing with

Table 1: Overview of how WBE data are presented on dashboards throughout Canada

Northwest Territories COVID-19 (NWT)	Average (unspecified genes) (gc/ml)
Centre for Health Informatics Alberta (AB)	Daily & 3-day moving average of: N1 (copies/day), N2 (copies/day), average N1 & N2 (copies/day). The daily load is calculated with the flow.
Newfoundland and Labrador (NFL)	SARS (unspecified gene) (gc/mg), SARS (unspecified gene) (gc/ml)
CentrEau-COVID (QC)	N1 (gc/ml), N1 (gc/ml) divided by PMMoV (gc/ml)
Simcoe-Muskoka Health District (ON)	7-day moving average of: N1 (gc/ml), N1 (gc/ml) divided by PMMoV (gc/ml)
613covid (ON)	Daily & 7-day moving average of: sum of N1 & N2 (gc/ml) divided by PMMoV (gc/ml)
Leeds, Grenvill and Lanark District (ON)	7-day moving average (unspecified viral signal)
KFL&A Public Health (ON)	Daily & moving geometric mean (range unspecified) of: SARS (unspecified gene) (gc/ml)
Durham Region COVID-19 Data Tracker (ON)	Daily and smoothed (LOWESS locally weighted regression, range unspecified) of SARS (unspecified gene) (gc/ml) on log10 scale
COVID-19 in York Region (ON)	7-day moving average of: SARS (unspecified gene) (gc/ml) divided by PMMoV (gc/ml), SARS (unspecified) (gc/ml)
ONTARIO DASHBOARD (ON)	Province-wide, weighted mean of standardized, biomarker-normalized (PMMoV) SARS (N1 & N2) results from several locations (units unknown).
Region of Waterloo (ON)	7-day moving average for: SARS (unspecified gene) (gc/ml); SARS (unspecified gene) (gc/ml) divided by PMMoV (gc/ml)
Thunder Bay District Health Unit (ON)	Daily and 7-day moving average for: N1 (gc/ml) divided by PMMoV
Global Institute for Water Security (SK)	7-day average for: viral extraction efficiency adjusted SARS (unspecified gene) (gc/100ml)
MetroVancouver (BC)	SARS (unspecified gene) (gc/ml), SARS (unspecified gene) (copies/day). The daily load is calculated with the flow.

daily clinical cases comes intuitively and does not need further explanation. Multiplying by flow is a standard procedure in wastewater engineering for calculating daily loads for various parameters. This practice is often called “flow normalization” in WBE as it helps take dilution effects into account. It does, however, not correct for secondary flow effects in sewers (e.g. settling and resuspension of particles). The analysis of flow data should thus be extended and taken to another level as described in section §7. To compare viral levels across sampling locations, it is common in the literature to perform population scaling by accounting for the number of people in the sewershed and expressing the data in viral load per person per day. Population normalization is when dynamic population

estimates are used to correct for population variations in the sewershed. The latter can be achieved in several ways using population numbers and flow, or human-related biomarkers, as will be touched upon later in section §7. The aggregation of viral signals from multiple locations into one metric falls outside the scope of this report.

It is interesting to note that the range of auxiliary parameters used for data transformations in the Canadian context has so far been limited to flow and PMMoV. The experiences of different Canadian research groups with PMMoV have revealed a varying degree of satisfaction, from “very pleased” to “not using it anymore”. Depending on the sewershed, there is mention of sporadic daily anomalies, continued biases over medium-long periods (snowmelt), and persistent seasonal variations. For some research groups, normalization with PMMoV does not remove any variability from the viral signal. However, it is still considered useful to check for sample integrity and extraction performance (i.e., deviating behaviour, outliers). The observed PMMoV values appear to be location-specific, and tend to show more erratic behaviour for small catchments where individual choices in diet are likely not sufficiently averaged out. Given these mixed results for PMMoV, it makes sense to start looking more intensely at various common wastewater quality parameters like NH4, COD, TSS, etc., for normalization. The latter have a clear link with human fecal discharge, their measurement methods are well defined and deliver repeatable results with smaller error bars, and their sampling methodology is perfectly compatible with the requirements needed for SARS-CoV-2 (e.g., refrigeration and analysis within a few days). Moreover, it was already shown that a combination of these common water quality parameters helped improve SARS-CoV-2 case predictions of a machine learning model compared to using SARS-CoV-2 and PMMoV viral measurements solely (Therrien et al., 2022).

7. USED BIOMARKERS AND AUXILIARY PARAMETERS IN THE INTERNATIONAL CONTEXT

Studies on best normalization practices have only recently appeared in the literature. In a comprehensive survey of uncertainties related to SARS-CoV-2 measurements for WBE in the UK, Wade et al. (2022) explore the use of flow, ammonia, ortho-phosphate, crAssphage and antidepressants to normalize the viral signal. While flow is considered the most critical parameter for correcting temporal signal variations due to dilution, it is not deemed sufficient for population normalization and obtaining dynamic population estimates. For the latter, it is suggested to use biochemical markers such as ammonia and orthophosphate by comparing actual and expected daily discharge values per capita per site. When flow data are not available to correct for temporal variability, one can use indirect normalization techniques employing biochemical markers as a proxy for flow under the assumption they are present in the sewershed at constant load. Dilution effects were found to have a minor impact on SARS-CoV-2 concentrations in wastewater, with significant changes occurring only during heavy rainfall events or discharges from other sources. Concerning the use of antidepressants, the metabolites (and not the parent compounds) showed promising results as suitable population markers. Fecal biomarkers like crAssphage are considered valuable, but are not the first choice for normalization.

In the English program, when flow data were not available, crAssphage was initially used to help normalize the SARS-CoV-2 results to account for dilution by industrial wastewater and rainfall. However, it created an extra workload and delayed the workflow. It was, therefore, subsequently dropped in favour of other indicators of fecal load. A multi-biomarker approach (e.g., ammonia nitrogen and orthophosphate) was proposed to provide more robust dilution estimates, since it is

impossible to distinguish between a decrease in flow and an increase in marker load using a single marker (e.g., due to a one-off industrial or agricultural discharge). The model can identify outliers such as one-off discharges and estimate flow variability if at least one marker is quantified, albeit with larger error bars. When no marker is quantified (missing data), the model predicts average flows with substantial error bars. Critical in the approach is that the natural variability in biomarker loads is accounted for by assigning variable importance to different markers. For instance, more priority should be given to ammonia nitrogen since crAssphage gene copy concentrations exhibit more natural variability, though both can inform on dilution estimates. In the Scottish program, when flow data are not available for a particular site, a cross-site linear mixed model trained on flow, NH_4, and population data of multiple sites is used. If neither flow nor NH_4 is available (missing data), a spline function using recent ammonia trends is employed (fitted on national trends plus site-specific effects). In the case of substantial combined sewer overflows (CSO) normalization through NH_4 data is preferred over flow data.

Langeveld et al. (2021) have investigated three normalization methods using flow, electric conductivity (EC) and crAssphage for multiple sites. They propose flow normalization supported by a quality check using conductivity monitoring as the advocated normalization method in case detailed flow monitoring is or can be made available. CrAssphage loads were found to be very consistent over time and space, and direct CrAssphage-based normalization could be applied reliably for populations of approximately 5000 and above. In contrast to Wade et al. (2022), flow is considered sufficient to execute population normalization by estimating the proportion of domestic sewage in the total flow when assuming a fixed domestic dry weather flow discharge per person. The EC and crAssphage normalization initially require an in-depth analysis of historic flow, EC and crAssphage data but do not require new daily flow measurements afterward. Large differences in correction factors were found, indicating the need for normalization to compare results for different regions. The EC method is sensitive to road de-icing and should be used with care in a Canadian context. The combined use of the three parameters provides an additional level of quality control and could point out occasional or persistent site-specific issues with the wastewater flow (pump failures, CSO, etc.) or quality (salts, industrial compounds, etc.). In the absence of flow data (current and historical) to determine the proportion of domestic wastewater through one of the methods above, direct SARS-CoV-2/crAssphage normalization is suggested as a valid option to produce the same trends. The crAssphage results appear seasonably stable, comparable between different sites even though the variation between individual stool samples is considerable, environmentally persistent, and show no significant build-up in sediments in the sewer network, unlike other parameters like COD or TSS.

Part of the normalization efforts of Wade et al. (2022) is described in more detail in Sweetapple et al. (2021). When flow and population data are available, NH_4 and PO_4 can be used to obtain dynamic population estimates and normalize for population. In the absence of flow and population data, direct normalization (dividing by NH_4 or PO_4) can be used to obtain the same trends, similarly to Langeveld et al. (2021)'s use of crAssphage. However, the direct normalized signal is only locally valid (i.e. for each specific site). The impact of the population normalization approach was found to account for dilution and population variations, but overall, not considered that impactful for the studied sites and the studied period (lockdown). While not insignificant in the short term, normalization did not greatly alter long-term trend analysis. Correlations with public health data were generally weaker for smaller, near-to-source sites. The population-normalized approach was found to make the comparison between sites possible but was deemed insufficient as the sole factor of importance. The data show a high degree of variation in the per capita daily loads of NH_4 and PO_4 between different sites, possibly related to industrial discharges and to uncertainty in assigning

census population data to sewersheds. Furthermore, other site differences that might have an impact, such as hydraulic retention time or sampling technique, might need to be incorporated more explicitly. The inclusion of site-specific sewershed, wastewater quality and sampling campaign metadata for better WBE-based prevalence predication was attempted using a machine learning approach and showed promising results (Morvan et al., 2021).

Alabzadeh et al. (2021) employed NH_4 for population normalization, but did not discuss any potential benefits as it was not the focus of the study. The authors did remark a certain threshold of viral load for small catchments before a clear relation of the wastewater signal with population infection dynamics is seen, and also state that infection clusters seem to have a larger effect in smaller catchments. Nagarkar et al. (2022) have investigated the normalization of the N1/N2 SARS-CoV-2 wastewater signal for three different-sized sewersheds using the matrix recovery efficiency (Betacoronavirus-1 strain OC43), flow and three fecal markers: crAssphage, PMMoV and HF183. Direct normalization was used just by dividing the respective signals. The results showed crAssphage and HF183 to behave similarly and differently from PMMoV. Furthermore, a clear difference was noticed in the normalization behaviour for the different sewersheds. However, firm conclusions on best normalization practices are unfortunately hampered by the limited dataset length and weekly sampling frequency. Greenwald et al. (2021) compared crAssphage, PMMoV, Bacteroides HF183 16S ribosomal RNA and human 18S ribosomal RNA for direct normalization. CrAssphage and PMMoV were the most consistent biomarkers; however, they did not strengthen the correlation with the clinical data. The PMMoV normalization even led to a significant deterioration of the correlation. Despite the sound data analysis procedure that was followed, the limited weekly sampling frequency is, unfortunately, hampering strong conclusions. Schoen et al. (2022) suggest a minimum 4-day sampling frequency for WBE purposes, based on a model-based analysis of artificial frequency-reduced datasets of SARS/PMMoV values from daily primary clarifier sludge grab samples.

8. ADDITIONAL NORMALIZATION PROCEDURE CONSIDERATIONS

As is so often the case, the quality of the raw data is primordial to reaching meaningful conclusions. Since post hoc data correction is cumbersome and often impossible, any trustworthy assessment of COVID prevalence based on WBE data should start with a well-crafted sampling campaign and measurement procedure. Ideally, samples should be taken daily, with 24h flow-composite samplers. The type of sewershed, the sampling location in the sewershed and the sampling target (solids, liquid) will determine to a great extent the limits of what can be achieved with the data. Uncertainty surrounding the lab measurements can be kept to a minimum by employing rigorous QA/QC procedures (Chik et al., 2022), and estimates of data quality can be obtained via a weighted score of several QA/QC parameters (Kantor et al., 2022). These estimates could be considered in further data analysis in a direct or indirect way.

Before data analysis, several authors log10-transform their SARS-CoV-2 datasets. This procedure is often utilized when working with heavily right-skewed data and alters the relative importance of small and large values. For this data transformation to work, zero values in the WBE dataset must be substituted with a constant. Generally, half the LOD value is used. However, these data points are most likely best omitted when checking the normalization procedures (Wade et al., 2022). An assumption in WBE is that SARS-CoV-2 data are log-normally distributed. However, the latter might not be the case, especially for small values close to the LOD and LOQ, as discussed earlier in section §5. Small values pose a particular problem in WBE as they come with high uncertainty yet are of high interest to public health when trying to detect the start of a new epidemiological wave. A specific

additional problem in low case settings is the masking of clinical data (i.e. not mentioning the exact number of cases below a certain threshold, e.g. 5 cases per day) out of ethical concerns on the traceability of patients. Another issue at low prevalence is asymmetric behaviour between clinical and WBE data (i.e. positive case detection in wastewater but not clinically, or the inverse), resulting in very low correlation scores. For all these reasons, it might be worthwhile to evaluate data from low case situations separately in a dedicated statistical framework to increase confidence in the decisions that can be made based on those data.

Data smoothing and outlier detection algorithms have to be evaluated for each specific dataset and situation (see section §6), just as is the case for the various possible normalization procedures. Smoothing and normalization improvements are generally noticeable in the form of a better correlation of the adjusted WBE data with clinical case data. The relation between WBE data and clinical case data can change in both magnitude and time lag (e.g., due to another virus variant, changes in population vaccination status, clinical testing regime). One should, therefore, carefully select the period of data to use (e.g., one wave at a time) and keep in mind the inherent limitations of the dataset (e.g., different data behaviour expectations for big cities vs. small municipalities; section §5).

In the literature, the degree of correlation is very often checked with the Pearson correlation coefficient or a variant thereof like the Bayesian Pearson's r (Nagarkar et al., 2022) or Kendall's Tau-b, a non-parametric ranked correlation coefficient adapted for left-censored data (i.e., datasets with data below a lower limit of detection; Greenwald et al., 2021). With a weighted Pearson correlation coefficient, data uncertainty could be incorporated. The Pearson correlation coefficient measures the linear correlation of variables and thus ignores many other types of relationships or correlations. It is an intuitive and easy-to-understand statistic that does not require its variables to have the same unit or range of values. It is, however, affected by noise and outliers, and its statistical inference is sensitive to the data distribution being normal, especially with small datasets. By employing the Pearson correlation coefficient for WBE datasets, one knowingly disregards the statistical properties of the data (not-randomly selected, autocorrelation, not normally distributed). Yet, using the coefficient still allows one to reflect on the benefits of normalization (Arabzadeh et al., 2021). Alternatively, linear regression analysis is also often used with the (adjusted) coefficient of determination indicating the goodness of fit of the linear regression model. The statistical properties of the data are generally disregarded here as well. Non-linear statistical regression models, or other types of descriptive and predictive models (e.g. machine learning) can also be used to assess the benefit of normalization by biomarkers. The goodness of fit can be expressed through a wide variety of scores. Conversely, residuals may be tested for autocorrelation, homoscedasticity, their statistical distribution, etc.

9. CONCLUSIONS

An overview has been presented of the reasons, the employed practices as well as the problems surrounding the normalization of the SARS-CoV-2 viral signal from wastewater samples. When selecting biomarkers and auxiliary parameters to correct for unwanted influences in the WBE data, it makes sense to first assess the uncertainty related to the normalization parameter used and consider the processes by which this parameter itself is influenced. It also makes sense to approach normalization in a stepwise manner. For instance, flow can be accurately measured and is unambiguously related to dilution, so its use for dilution correction is self-evident. Population estimates for population normalization can be obtained with many parameters. However, each has

a different degree of inherent uncertainty (e.g., source, stability, sewer transport, measurement) that allows ranking them by their suitability: flow > common water quality parameters > CEC > viral indicators like crAssphage and PMMoV. Using several parameters at once (in a combined way) to normalize a specific unwanted effect reduces uncertainty in the correction procedure compared to using just one parameter. Specific combinations of parameters (e.g., soluble indicator, particulate indicator, viral indicator, etc.) should be investigated for their ability to normalize other unwanted effects that have not been addressed thoroughly in literature so far (sewer transport processes, sampling error, lab measurement quality, etc.).

10. REFERENCES

Arabzadeh, R., Grünbacher, D. M., Insam, H., Kreuzinger, N., Markt, R., & Rauch, W. (2021). Data filtering methods for SARS-CoV-2 wastewater surveillance. Water Science and Technology, 84(6), 1324–1339. https://doi.org/10.2166/wst.2021.343

Chik, A., Abbey, A.-M., Flemming, C., Fletcher, T., Ho, J., Oke, M., Pileggi, V., Raby, M., Simhon, A., Thomas, J., Weir, S., D'Aoust, P., Brown, S., Chao, T.-C., Daigle, J., Delatolla, R., Dhiyebi, H., Donaldson, M., Edwards, E., & Xie, Y. (2022). Protocol for Evaluations of RT-qPCR Performance Characteristics: Technical Guidance. https://doi.org/10.13140/RG.2.2.10627.02080

Chik, A. H. S., Glier, M. B., Servos, M., Mangat, C. S., Pang, X. L., Qiu, Y., D'Aoust, P. M., Burnet, J. B., Delatolla, R., Dorner, S., Geng, Q., Giesy, J. P., McKay, R. M., Mulvey, M. R., Prystajecky, N., Srikanthan, N., Xie, Y., Conant, B., & Hrudey, S. E. (2021). Comparison of approaches to quantify SARS-CoV-2 in wastewater using RT-qPCR: Results and implications from a collaborative inter-laboratory study in Canada. Journal of Environmental Sciences (China), 107, 218–229. https://doi.org/10.1016/j.jes.2021.01.029

Greenwald, H. D., Kennedy, L. C., Hinkle, A., Whitney, O. N., Fan, V. B., Crits-Christoph, A., Harris-Lovett, S., Flamholz, A. I., Al-Shayeb, B., Liao, L. D., Beyers, M., Brown, D., Chakrabarti, A. R., Dow, J., Frost, D., Koekemoer, M., Lynch, C., Sarkar, P., White, E., … Nelson, K. L. (2021). Tools for interpretation of wastewater SARS-CoV-2 temporal and spatial trends demonstrated with data collected in the San Francisco Bay Area. Water Research X, 12. https://doi.org/10.1016/j.wroa.2021.100111

Hoffmann, T., & Alsing, J. (2021). Faecal shedding models for SARS-CoV-2 RNA amongst hospitalised patients and implications for wastewater-based epidemiology. MedRxiv, 2021.03.16.21253603. https://doi.org/10.1101/2021.03.16.21253603

Kantor, R. S., Greenwald, H. D., Kennedy, L. C., Hinkle, A., Harris-Lovett, S., Metzger, M., Thornton, M. M., Paluba, J. M., & Nelson, K. L. (2022). Operationalizing a routine wastewater monitoring laboratory for SARS-CoV-2. PLOS Water, 1(2). https://doi.org/10.1371/journal.pwat.0000007

Langeveld, J., Schilperoort, R., Heijnen, L., Elsinga, G., Schapendonk, C. E. M., Fanoy, E., de Schepper, E. I. T., Koopmans, M. P. G., de Graaf, M., & Medema, G. (2021). Normalisation of SARS-CoV-2 concentrations in wastewater: the use of flow, conductivity and CrAssphage. MedRxiv, 2021.11.30.21266889. https://doi.org/10.1101/2021.11.30.21266889

Morvan, M., lo Jacomo, A., Souque, C., Wade, M., Hoffmann, T., Pouwels, K., Singer, A., Bunce, J., Engeli, A., Grimsley, J., O'Reilly, K., & Danon, L. (2021). Estimating SARS-CoV-2 prevalence from large-scale wastewater surveillance: insights from combined analysis of 44 sites in England. https://doi.org/10.21203/rs.3.rs-770963/v1

Nagarkar, M., Keely, S. P., Jahne, M., Wheaton, E., Hart, C., Smith, B., Garland, J., Varughese, E. A., Braam, A., Wiechman, B., Morris, B., & Brinkman, N. E. (2022). SARS-CoV-2 monitoring at three sewersheds of different scales and complexity demonstrates distinctive relationships between wastewater measurements and COVID-19 case data. Science of the Total Environment, 816. https://doi.org/10.1016/j.scitotenv.2021.151534

Ort, C., Lawrence, M. G., Reungoat, J., & Mueller, J. F. (2010). Sampling for PPCPs in Wastewater Systems: Comparison of Different Sampling Modes and Optimization Strategies. Environmental Science & Technology, 44(16), 6289–6296. https://doi.org/10.1021/es100778d

Schang, C., Crosbie, N. D., Nolan, M., Poon, R., Wang, M., Jex, A., John, N., Baker, L., Scales, P., Schmidt, J., Thorley, B. R., Hill, K., Zamyadi, A., Tseng, C.-W., Henry, R., Kolotelo, P., Langeveld, J., Schilperoort, R., Shi, B., … McCarthy, D. T. (2021). Passive Sampling of SARS-CoV-2 for Wastewater Surveillance. Environmental Science & Technology, 55(15), 10432–10441. https://doi.org/10.1021/acs.est.1c01530

Schoen, M. E., Wolfe, M. K., Li, L., Duong, D., White, B. J., Hughes, B., & Boehm, A. B. (2022). SARS-CoV-2 RNA wastewater settled solids surveillance frequency and impact on predicted COVID-19 incidence using a distributed lag model. MedRxiv, 2022.02.21.22270864. https://doi.org/10.1101/2022.02.21.22270864

Sweetapple, C., Wade, M. J., Grimsley, J. M. S., Bunce, J. T., Melville-Shreeve, P., & Chen, A. S. (2021). Population normalisation in wastewater-based epidemiology for improved understanding of SARS-CoV-2 prevalence: A multi-site study. MedRxiv, 2021.08.03.21261365. https://doi.org/10.1101/2021.08.03.21261365

Therrien, J.-D., Maere, T., Halle, S., Dallaire, P., & Vanrolleghem, P. A. (2022). Using the right wastewater characteristics for early COVID-19 pandemic warning and forecast using deep machine-learning. 12th Urban Drainage Modeling Conference, 10-12 January, Costa-Mesa, California, USA.

Wade, M. J., lo Jacomo, A., Armenise, E., Brown, M. R., Bunce, J. T., Cameron, G. J., Fang, Z., Farkas, K., Gilpin, D. F., Graham, D. W., Grimsley, J. M. S., Hart, A., Hoffmann, T., Jackson, K. J., Jones, D. L., Lilley, C. J., McGrath, J. W., McKinley, J. M., McSparron, C., … Kasprzyk-Hordern, B. (2022). Understanding and managing uncertainty and variability for wastewater monitoring beyond the pandemic: Lessons learned from the United Kingdom national COVID-19 surveillance programmes. Journal of Hazardous Materials, 424. https://doi.org/10.1016/j.jhazmat.2021.127456