





## Challenges in realising the potential of wastewater-based epidemiology to quantitatively monitor and predict the spread of disease

Julian Faraway <sup>a,\*</sup>, James Boxall-Clasby<sup>b</sup>, Edward J. Feil<sup>c</sup>, Marjorie J. Gibbon<sup>c</sup>, Oliver Hatfield<sup>d</sup>, Barbara Kasprzyk-Hordern  <sup>b</sup> and Theresa Smith <sup>a</sup>

<sup>a</sup> Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK

<sup>b</sup> Department of Chemistry, University of Bath, Bath BA2 7AY, UK

<sup>c</sup> Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK

<sup>d</sup> Institute for Mathematical Innovation, University of Bath, Bath BA2 7AY, UK

\*Corresponding author. E-mail: jjf23@bath.ac.uk

 JF, 0000-0002-7308-3084; BK-H, 0000-0002-6809-2875; TS, 0000-0002-7085-3864

### ABSTRACT

Researchers around the world have demonstrated correlations between measurements of SARS-CoV-2 RNA in wastewater (WW) and case rates of COVID-19 derived from direct testing of individuals. This has raised concerns that wastewater-based epidemiology (WBE) methods might be used to quantify the spread of this and other diseases, perhaps faster than direct testing, and with less expense and intrusion. We illustrate, using data from Scotland and the USA, the issues regarding the construction of effective predictive models for disease case rates. We discuss the effects of variation in, and the problem of aligning, public health (PH) reporting and WW measurements. We investigate time-varying effects in PH-reported case rates and their relationship to WW measurements. We show the lack of proportionality of WW measurements to case rates with associated spatial heterogeneity. We illustrate how the precision of predictions is affected by the level of aggregation chosen. We determine whether PH or WW measurements are the leading indicators of disease and how they may be used in conjunction to produce predictive models. The prospects of using WW-based predictive models with or without ongoing PH data are discussed.

**Key words:** COVID-19, prediction models, public health surveillance, SARS-CoV-2, sewage

### HIGHLIGHTS

- This study is a worked example of predicting the spread of disease using wastewater-based epidemiology (WBE).
- It demonstrates non-linear relationships between wastewater and public health measures of the disease.
- It demonstrates time-varying relationships in WBE.
- It investigates time lags in the prediction of WBE.
- It makes recommendations for building predictive models in WBE.

### INTRODUCTION

The utility of wastewater-based epidemiology (WBE) as an efficient approach to monitoring the health of populations and the environment has been increasingly recognised over the last decade. WBE has been used to estimate illicit drug use trends (Thomas *et al.* 2012; Ort *et al.* 2014) but has also been applied to estimate public exposure to alcohol (Reid *et al.* 2011; Baz-Lomba *et al.* 2016), tobacco (Castiglioni *et al.* 2015), counterfeit medicines (Venhuis *et al.* 2014; Causanilles *et al.* 2018), as well as levels of stress biomarkers, such as isoprostanes (Ryu *et al.* 2016; O'Brien *et al.* 2019), and usage of pharmaceuticals (Escolà Casas *et al.* 2021; Kasprzyk-Hordern *et al.* 2021). Interest in this approach dramatically increased because of the COVID-19 pandemic, resulting in rapid conceptual and methodological advances. In our examples, we examine the application of WBE to monitoring SARS-CoV-2 and highlight some of the uncertainties and challenges that remain in modelling the data. These considerations will apply to many other pathogens of ongoing interest and are relevant for planning responses to future pandemics.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

If it were possible to obtain real-time reliable estimates of disease prevalence in a community, there would be little need for WBE. In practice, it is expensive, time-consuming, and intrusive to obtain good public health (PH) data because this requires direct sampling or recording of individuals. In contrast, wastewater (WW) can be sampled at a lower cost, processed more quickly, and carried out with much lower privacy concerns. WBE is already being used around the world to monitor SARS-CoV-2 RNA in WW in various ways. We identify three increasingly challenging modes of use:

1. *Presence or absence of RNA in WW.* In countries where a so-called zero-COVID policy has been attempted, this can identify the presence of infection in the community that might prompt lockdown measures.
2. *Qualitative level of RNA in WW.* For example, this might be reported as absent, low, medium, or high. This gives an indication of the general level of COVID-19 infection in the community without attempting a numeric estimation. It also provides an indication of change over time in levels of infection. Public dashboards for WW RNA reports often use this mode (see Scotland example below).
3. Numeric estimates of the numbers of individuals currently infected with COVID-19 coupled with a prediction of near-future numbers of infected along with realistic expression of uncertainty in the estimates and predictions.

WBE has largely succeeded in the first two of these modes, but the third goal of numerical estimation and prediction remains a challenge. A discussion of this challenge is the topic of this article. The first two modes of operation do not require detailed PH data, while the third mode requires at least some quality PH data (in our example, we wish laboratory-confirmed COVID-19 cases) and a model for the relationship between PH and WW data. For WW data to provide an added value beyond PH data, it needs to produce satisfactory predictions in the absence of PH data after some initial (or training) PH data have been used to produce a predictive model. Once a WW-based model has been established, it might compensate for future shortcomings in PH data. For example, there have been failures in PH testing laboratories and the exceedance of test capacity. In the longer term, one might allow that the WW model would suffice without any PH data.

Many recent papers (Ai *et al.* 2021; Feng *et al.* 2021; Greenwald *et al.* 2021; Melvin *et al.* 2021; Zulli *et al.* 2021; Wurtzer *et al.* 2022) show a correlation between the daily or weekly amount of COVID-19 RNA in WW and the numbers of COVID-19 infections reported to the PH authorities, but some do not go beyond showing a correlation. Certainly, the presence of a correlation indicates the potential for a prediction but substantial additional work is necessary before an operational model that provides numerical predictions can be produced. There is some debate about how much WW data can anticipate PH data, which relates to the goal of WW models providing an ‘early warning system’.

The quality of predictions depends on the quality of WW measurements. Improving these measurements is a major focus of WBE research. There are many sources of variation that affect these measurements that are discussed in Alhama *et al.* (2021) and Wade *et al.* (2022). We accept that these measurements are subject to some variability. Our interest lies in the consequences of this variability.

We start with a description of the WW and PH data used in our running Scotland example. We discuss how WW and PH data are aligned and the issues with the use and quality of PH data. We look at the form of the relationship between PH and WW data and how this can vary over time. We consider the relative advantages of predicting with WW and PH data. We conclude with a discussion of how predictive models might continue to perform into the future.

## DATA

The data for the main example come from Scotland from the beginning of September 2020 to the end of August 2021. Public Health Scotland has publicly provided daily reports of laboratory-confirmed COVID-19 cases broken down by the 14 Health Area Boards that cover all of Scotland – see Public Health Scotland (2021). The Scottish Environmental Protection Agency has collated and publicly provided data – see Scottish Environmental Protection Agency (2021) – from up to 110 WW treatment plants (WWTPs) in our data. Composite samples were drawn over 24-h periods. Samples were processed and analysed using RT-qPCR methods to produce a measure of SARS-CoV-2 RNA in gene copies per litre. Some WWTPs were sampled frequently but other WWTPs less densely and not for the whole period of this study. Further details will be provided in the analysis to follow.

Our purpose here is not to conclude anything about Scotland specifically, but to use this data as a representative of the type of data that might be collected anywhere. Data of this nature are subject to many uncertainties that will be present in any location. Our interest lies in the modelling process and the general lessons that may be learned.

In practice, one might also allow additional data such as flow and water quality data from WWTPs. Such data might improve the quality of the predictions but will not always be available in all locations. In this study, we restrict attention to only the case and RNA count data to focus on the modelling aspects.

We also consider data from the USA provided publicly by Biobot (2021) (see Duvallet *et al.* (2021) for an analysis partially based on this data). These data differ in that only weekly COVID-19 case data are available, and the WWTPs within each of the 30 counties available have already been combined. We use this data to confirm that the general conclusions drawn from the Scotland data apply elsewhere.

### Registering WW and PH data

In our example data from Scotland, there are 14 PH areas reporting daily cases. These areas cover Scotland completely. We also have WW measures from 110 WWTPs. These WWTPs serve catchments that are usually (but not always) within a single PH area. Not all WW flows to a WWTP, particularly in rural areas, and not all WWTPs (particularly smaller ones) are sampled. Three sparsely populated areas have only one WWTP sampled, while one area has 17 WWTPs. In Glasgow, the largest city, almost all the population is served by a sampled WWTP, while as little as a third of the population in the more sparsely populated areas is served by a sampled WWTP. Overall, about 70% of the population is served by a WWTP included in the data. A WWTP serving a large population might be sampled several times a week across the whole period of the study, while smaller WWTPs might be sampled less frequently and only in the later part of the study. Although reporting and measurement practices vary across the world, the challenges seen in our example data are likely to arise in many locations. In less wealthy countries, less information is likely to be available.

We have taken a simple approach to aligning the WW to the PH data.

1. For each WWTP, during the period for which observations are available, we use LOCF (last observation carried forward, i.e., the estimate does not change until a new observation becomes available) to create a daily estimated WW measure. We avoid interpolation because, when making predictions, we will know only past values and not the future value.
2. For each day within each area, we computed a weighted average of the WWTPs where the weights are equal to the population served by the WWTP.
3. Reporting of cases is sensitive to the day of the week. We use a rolling weekly mean to avoid day-of-the-week effects. The rolling mean uses only the past seven values and no future values.
4. In processing both the WW and PH data, we have taken care to avoid using data from the future, but this also results in some small lag to both estimates.

One might reasonably suggest improvements to this procedure. We might use Geographic Information Systems data to better allocate WWTPs across PH areas. We might use other methods of integrating data from the recent past. But the point is that uncertainty will unavoidably be added, and the nature of the data will limit the localisation and confidence in our predictions.

There are time lags in both the WW and PH data processes. The amount of RNA deposited in WW varies over time within an individual as determined by the faecal shedding profile and between individuals – see Hoffmann & Alsing (2021). When considering data from a narrow locality such as a school or hospital, it is necessary to model these effects if we wish to estimate the number of infected individuals. For the larger areas we are considering, these effects are averaged over many individuals, and it is not feasible to deconvolve the individual profiles. We do not consider the faecal shedding profile in the models considered later for this reason, but there will be a time lag between infection and measurable RNA appearing at the WWTP. Similarly, PH reporting will experience time lags due to the delay in the appearance of symptoms and the processing and reporting of PCR test results. We have not modelled either of these lags in the models to follow.

### Issues with PH data

WW and PH data both measure the quantity of disease present in the population. We have acknowledged the variability in WW measurements, but PH measures have their own set of problems.

Although WW measures prevalence, many studies report the correlation between RNA measures and incidence. Incidence relates to new cases, whereas prevalence relates the number of currently infected individuals. RNA concentrations measure prevalence but imperfectly. Governments tend to be more interested in incidence, particularly large increases that might motivate interventions, so there is impetus to make incidence the target of predictive modelling. One can attempt to estimate the prevalence by integrating the incidence with an appropriate weighting function that takes into account the length of infection

and other factors. By the same reasoning, incidence can be regarded as a form of derivative of the prevalence function and is inherently more difficult to predict. For the purposes of this study, we use the 7-day rolling mean of the daily incidence. This has the advantage of resolving the day-of-the-week variation and the use of the mean results in a measure that is somewhere between incidence and prevalence. This will be the target of our predictive model and we will call it the *case rate*.

One major goal of WBE is to replace more expensive and intrusive PH measures of SARS-CoV-2 (or other diseases) with less-expensive WW measures. To build a predictive model for case rates, we need a training set with reliable PH measures. This may be difficult to obtain, particularly when our goal is to avoid the necessity of collected PH measures in the longer term.

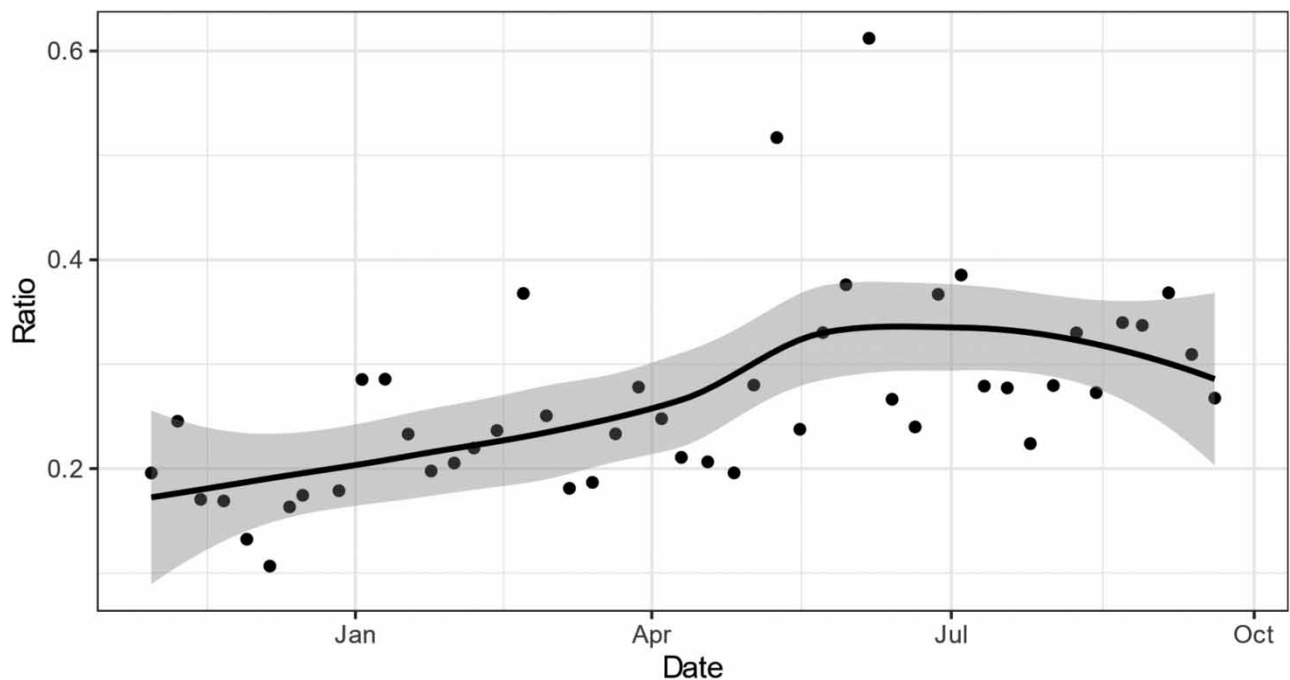
In the UK, symptomatic individuals and, sometimes, close contacts are strongly encouraged to take a PCR test and report the results to the PH authorities. Unfortunately, from a modelling perspective, some individuals will be asymptomatic and others will not be compliant. Furthermore, there have been problems with some commercial PCR testing laboratories. Independently, the UK's Office of National Statistics (ONS) takes weekly random samples from the population to estimate the proportion who are currently infected. This survey is small in scale and can only produce reliable estimates over large areas – in this case, all of Scotland. It cannot be used alone to provide the localised information we require for our prediction model. Even so, we can use it to check the PH data.

We compute the ratio of the weekly number of cases across all Scotland to the weekly estimated number from the ONS infection survey. This ratio is shown in Figure 1 as it varies over time. We see that the ratio reaches a maximum value of about 0.35. We expect that the ratio will be less than one because infected individuals will only report at most once in the PH data but will test positive over more than 1 week for infection survey purposes. Furthermore, the infection survey will also capture asymptomatic cases. We could develop a method to estimate the prevalence from the PH data, but there is a more serious problem. The ratio is not constant over time. A plausible explanation is that PH reporting improved over the time period, but other explanations are possible.

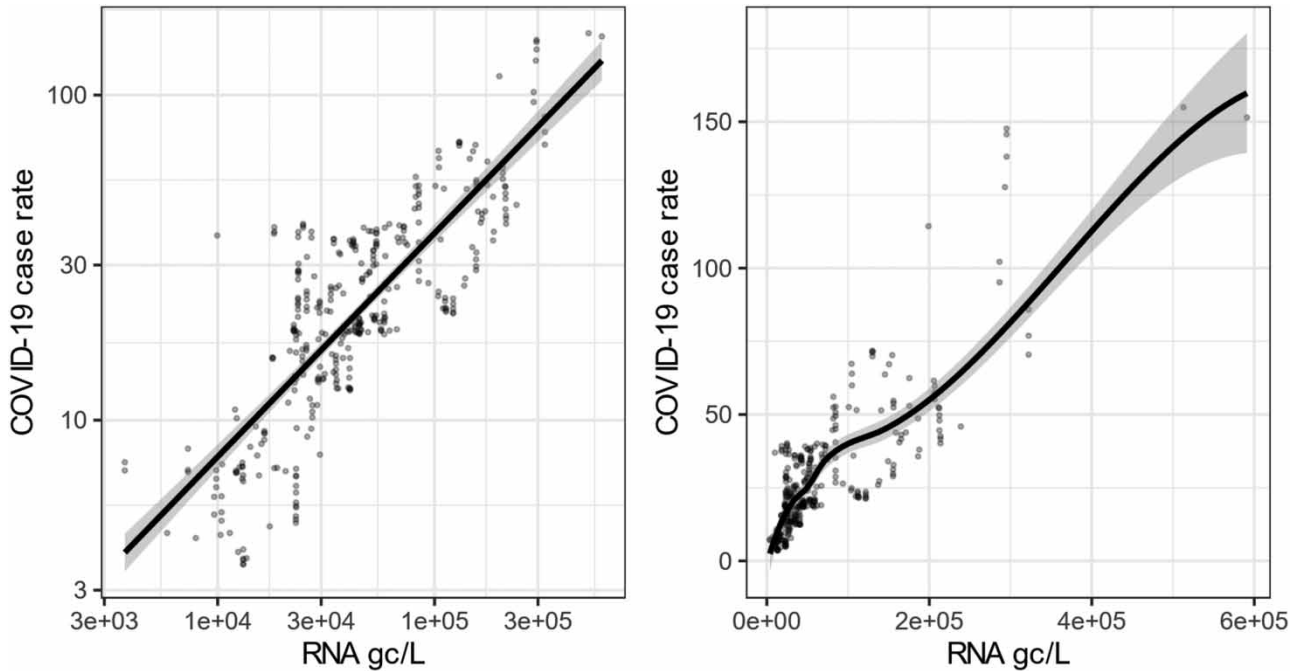
Any proposed prediction model will need to address this variable bias in the PH measure. In this location, we have the benefit of two independent measures of the case rate, but this will not be available in general (it is not present in our secondary USA example).

### Case rates are not proportional to RNA measures

We start with a focus on the data from the largest city in Scotland, Glasgow, which has the densest data across the available areas as seen in Figure 2. Both the PH and the WW measures are caused by the true number of infected individuals. The PH



**Figure 1** | Ratio of reported COVID-19 cases to cases estimated by the ONS infection survey.



**Figure 2** | SARS-CoV2 RNA and COVID-19 case rates in Glasgow. The case rate is the number infected per 100,000 population (as defined earlier). A log-scale is used on the left with a linear fit and 95% confidence band. On the right, the same data are shown untransformed and with smoothed LOESS fit.

measure is less variable but more expensive to obtain, while the WW measure is more variable but less expensive. The interest lies in predicting the PH measure from the WW measure because we envision a future use where only the WW measure is available. Furthermore, the PH measure is in the units we need, while the WW measure is an RNA concentration, which is not directly useful. For these reasons, our models treat the RNA measure as the predictor and the case rate as the response.

There are strong biological reasons to expect that cases should be proportional to measured RNA. Let  $c$  be cases (per 100,000 population) and  $g$  be gene copies per litre. By the proportionality assumption, these are linked by:

$$c = pg \quad (1)$$

Ideally, one would determine the constant  $p$  by some calibration process that would depend on the locality. We can then measure  $g$  and directly estimate  $c$ . Naturally, we expect some measurement errors. We expect that other variables, such as flow rates or water quality indicators, will add variation. If we wish to use WBE to estimate disease prevalence when PH data are not available, it is critical that this underlying linear relationship holds.

Because  $c$  and  $g$  often vary over orders of magnitude, it is understandably common to plot them on a log-scale (the base of the log is not important) as seen in the first plot of [Figure 2](#). An approximate linear relationship is seen, as is typical in such plots, like so:

$$\log c = a + b \log g \quad (2)$$

This linearity on the log-scale becomes  $c = pg^b$  on the untransformed scale where  $a = \log p$ . Unless  $b=1$  (which we would prefer), the relationship is non-linear. In the Glasgow data, we can estimate  $b$  from the slope as 0.69 with a standard error (SE) of 0.026. We see that  $b < 1$ , so that a convex curvilinear relationship is fitted. Of all the 14 areas in this data, this is the highest estimate of  $b$ . Furthermore, when we fit a smoothed curve to the untransformed data as seen in the second plot of the figure, we do not even see the expected power curve relationship. We will see later that this is typical and not unique to Glasgow.

How can we address this situation? As we shall see later, there is a time-varying relationship that may explain at least part of the problem. Furthermore, if we had additional auxiliary data, we might be able to normalise (or correct) the measurements to restore a  $c = pg$  relationship.

An additional concern is the classic errors-in-variables or measurement error problem as described in Fuller (1987). Both  $c$  and  $g$  are subject to a substantial measurement error, which leads to some complications with the estimation of  $p$ . There are well-established methods for working with this, which we do not pursue further here.

There remains a possibility that even after these corrections and adjustments, the proportional relationship is not restored. This would make predictive modelling substantially more difficult. We return to this issue in the discussion.

For very low or zero counts, there are difficulties in using the log transformation. There are also complications when the RNA measure approaches the limits of quantification and detection. Some modification to the modelling approach is necessary, but these complications do not arise with the Glasgow data due to the high population and prevalence of COVID-19.

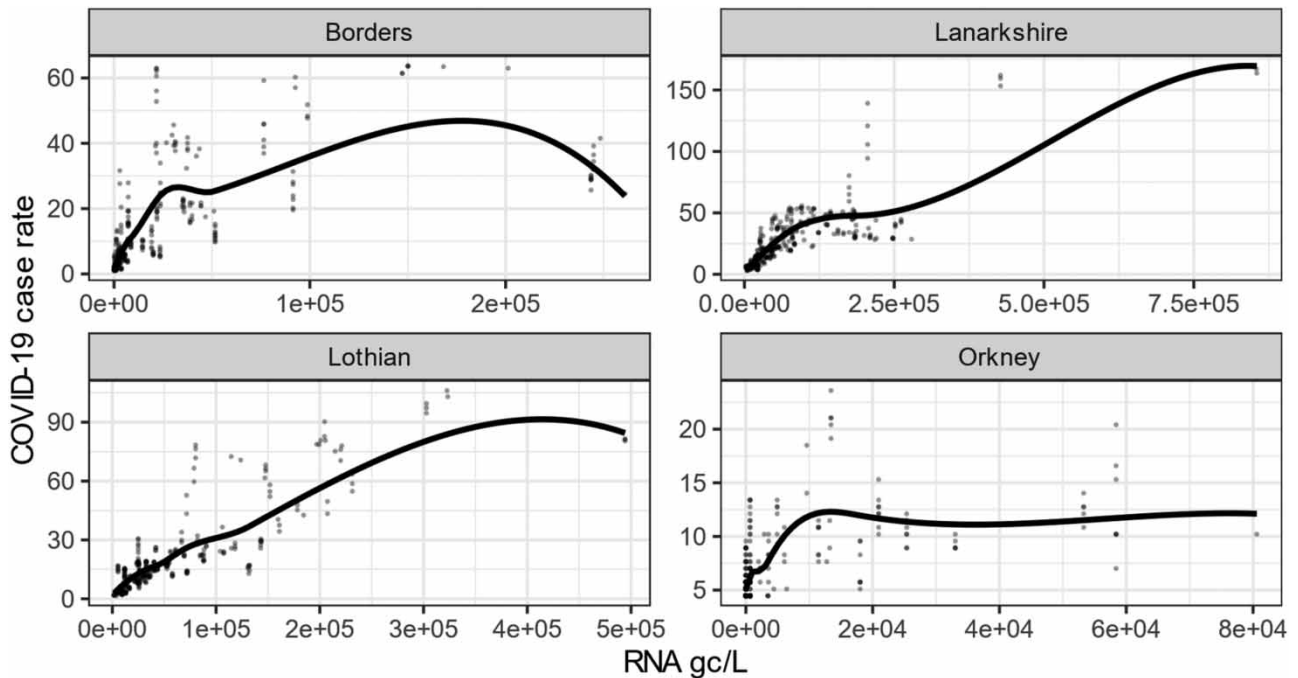
**Locality of the relationship between case rates and RNA measures**

Figure 3 shows the same information as in the right panel of Figure 2 but for four representative areas. We see that the shape of the relationship between case rates and RNA measures varies by locality. The weakest relationship is seen in Orkney, which is the least populated area shown. The heterogeneity in the relationship carries over to the areas not shown. The relationship varies in ways that would be difficult to parameterise. We can make a prediction for these locations, but we would find this difficult in new locations without substantial PH data.

Let us suppose that, with appropriate normalisation and improvements in measurement, we can get closer to linearity on a log-scale. As is, the data already show a roughly approximate linear relationship on the log-scale. Consider a linear model on this log-scale, where the parameters vary by area  $i$ :

$$\log c = a_i + b_i \log g \tag{3}$$

In fitting this model, we find that the values of  $a$  and  $b$  vary significantly by  $i$ . The estimated values of the slope  $b$  are always well less than 1, with more populated areas having  $b \approx 0.65$ , while the less populated areas show lower values indicating a weaker relationship in those areas. In these areas, the WWTP coverage is lower, fewer samples are taken, and the infection rates tend to be lower due to the sparser population. The intercept  $a_i$  represents the amount RNA measured per infected



**Figure 3** | Relationship between SARS-CoV2 RNA and COVID-19 case rates in four areas of representative areas of Scotland.

individual, but these show substantial variation. There are several plausible reasons for this variation. The amount of time RNA spends in the sewerage will vary according to the locality and other environmental factors that will affect the rate of decay.

Provided these parameters remain constant in time, we might use training data to build a model to predict future case rates based on future WW measures. A more problematic situation would arise where we introduce WBE to a new location where we lack the training data we have in this example. Can these parameters be determined without data? How would this work for a new virus or other target? We might attempt some physical modelling of the amount of RNA (or other output) to be expected per infected person that would be sampled at a given WWTP, given knowable characteristics of the catchment and necessary environmental information. This was attempted in [Morvan \*et al.\* \(2021\)](#), but without much success. Progress in this area is necessary to cope with situations where little or no PH information is available, or the target is new.

### Aggregation and prediction uncertainty

Suppose we fit the linear model to the logged Glasgow-only data as described previously. We use this model to predict the case rate based on the average RNA measure observed across all of Scotland for the entire period, which is 56,300 gc/L. Predicting the case rate for this central value of RNA is an easier task than a prediction for a very low or very high value of RNA. The predicted number of cases is 25.3 per 100,000 with a 95% prediction interval of (10.6, 60.4). The width of the interval reflects the uncertainty. For decision-making purposes, we would prefer this to be narrower.

We can reduce the prediction uncertainty by aggregating over all of Scotland and use this aggregate data to fit the linear model. We compute population-weighted averages of both PH and WW data over all the areas and fit the same model. Under this model, the predicted case rate for the same RNA measurement of 56,300 gc/L is 21.3, with a 95% prediction interval of (11.6, 39.1). Thus, fitting a model based on aggregated data reduces the variation, allowing us to obtain a much narrower prediction interval, relatively speaking.

In contrast, should we wish to predict cases within the catchment of a single WWTP and even if we were able to specify suitable parameters for the catchment area of that WWTP, we would obtain a prediction interval considerably wider than the Glasgow interval.

We have used a simple but transparent modelling approach. We might improve the prediction uncertainty with some more sophisticated modelling, but the natural variation in the data, as seen in the figures, limits the precision of the predictions. The use of auxiliary data not available to us here might also improve the predictions. The acceptable level of uncertainty would depend on the intended use of these predictions.

### Time-varying relationships

We have already seen that the PH reporting of cases varies over time. Now we look at how the relationship between PH and WW measures changes over time. Consider the varying coefficient model:

$$c(t) = a_i(t) + b(t)g(t) \quad (4)$$

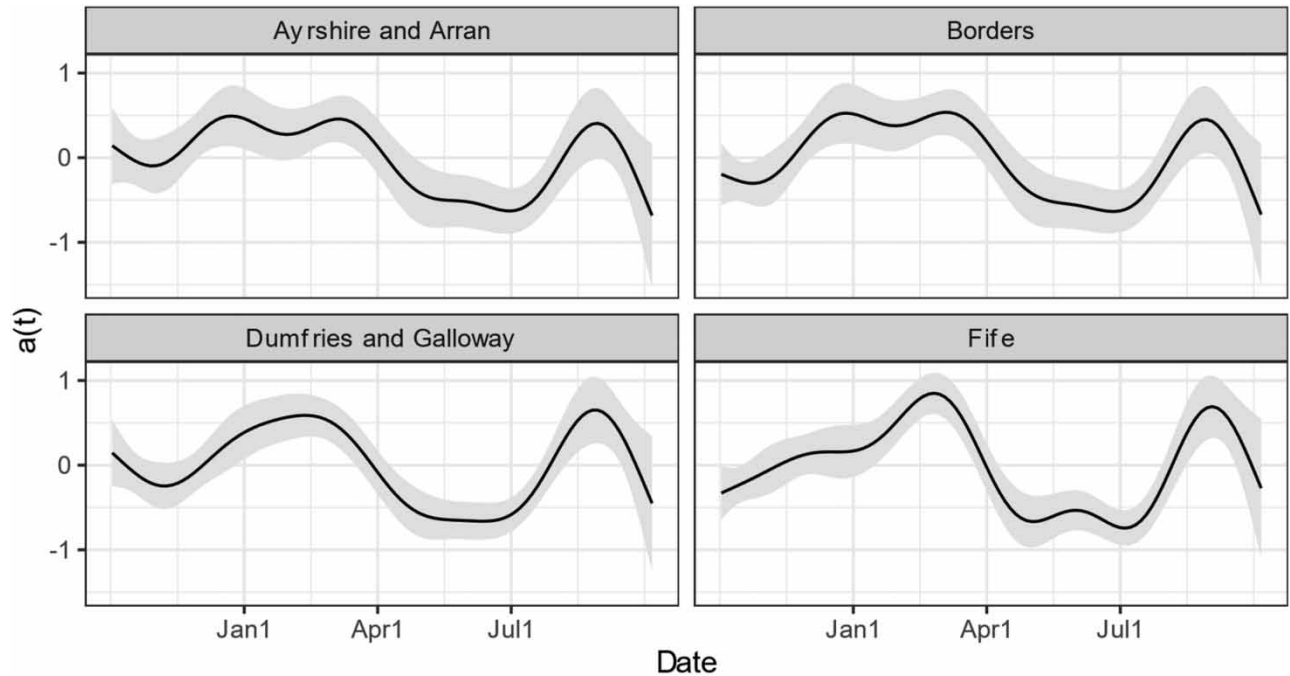
where the intercept coefficient functions  $a_i(t)$  vary with area  $i$ , but the slope coefficient  $b(t)$  is the same across areas. Variations on this choice are reasonable, but the essential conclusion would be the same. These models can be fit using the `mgcv` package in R ([Wood 2017](#)).

In [Figure 4](#), we can see that the  $a(t)$  parameters change over time. We show only four of the areas in this plot, but the other 10 are qualitatively similar. The larger the value, the less RNA is measured per infected individual and so smaller values are preferable. We can see that there were large declines in the early part of 2021 indicating a greater recovery of RNA, but this improvement was reversed later in the year.

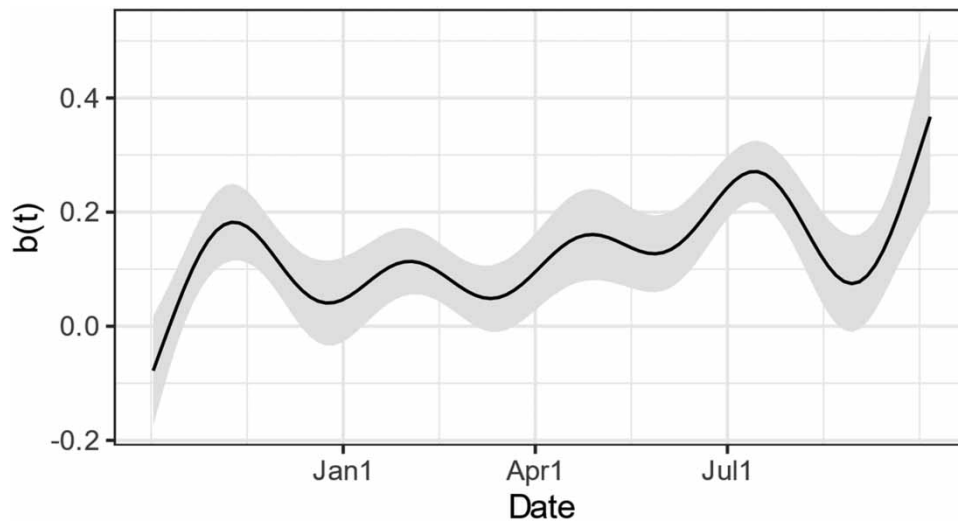
The estimated slope coefficient  $b(t)$  is seen in [Figure 5](#). We see that this varies over time and does not approach the desired value of one. This indicates that the non-proportionality continues throughout the study period while varying in its form.

Why do we see this variability in the relationship? We saw earlier that PH reporting improved during the period of the study, but this works in opposition to the trends seen in the  $a(t)$ . Several explanations are plausible.

In [Figure 6](#), we see that the number of samples drawn per month increased sharply in early 2021 and fell off again later in the year. This is one plausible explanation for why coefficient function varies in the manner seen previously. More samples



**Figure 4** | Estimated  $a(t)$  for four selected areas.



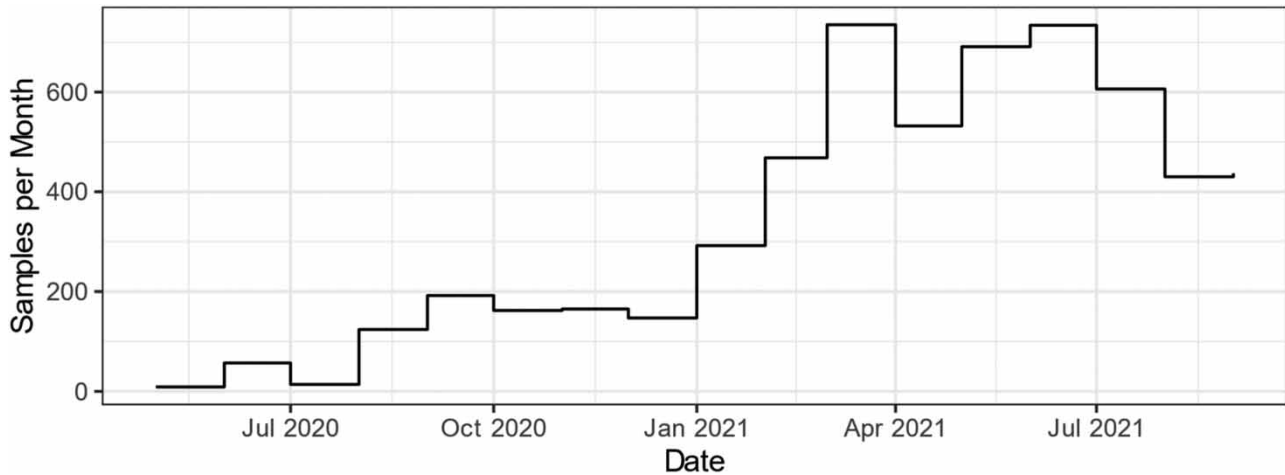
**Figure 5** | Estimated  $b(t)$  for the time-varying model.

would have resulted in more accurate and less variable estimates of the area of RNA concentrations, which would have strengthened the correlation with observed cases. However, there are several other reasons why these functions might vary:

- Variations in PH reporting (as seen in this and the USA data);
- Variations in laboratory methods (these have improved over the period);
- Variations in faecal shedding possibly due to increasing vaccination or SARS-CoV-2 variants;
- Environmental variations such as agricultural runoff or weather, for example.

The time-varying relationship means that it will be difficult to build predictive WBE systems for the long term, and continuing updates, requiring the collection of further PH data, may be necessary.





**Figure 6** | Number of WW samples taken each month.

### Predictive ability of WW measurements in comparison with PH measures

How well does WW data predict PH case rates? First, we consider how well WW data can predict future PH data. Consider lagged models of the form:

$$\log c(t) = a_i + b_j \log g(t - j) \quad \text{for } j = 0, 1, \dots, 7 \quad (5)$$

We report  $R^2$  and SE of the models in the table below. These are in-sample estimates, but the model class is quite simple. We would expect the out-of-sample performance to be a little worse.

In [Table 1](#), we see that the fit of the model is best at lag=0, concurrent prediction and that the greater the lag, the worse the fit of the model. This indicates that although we can use WW to predict cases further ahead in time, the quality of these predictions is worse than the concurrent prediction. There is some variation in the findings in published studies. Some claim improved prediction ahead of time by as much as 2 weeks ([Melvin \*et al.\* 2021](#)), while others make more modest claims ([Morvan \*et al.\* 2021](#)), and some (like us) do not find a look-ahead ([Zulli \*et al.\* 2021](#)). New infections take varying amounts of time to appear in the PH and WW systems, and as mentioned earlier, the data analysis induces some lag on both variables. With adequate resourcing, there is no strong evidence that either system can anticipate the other by any significant number of days.

Now consider using PH data to predict PH data in the future. If we fit a model of the form where we predict 1 week ahead:

$$\log c(t) = a_i + b_i \log c(t - 7) \quad (6)$$

**Table 1** | Performance of lagged prediction models

| Lag in days | $R^2$ | Standard error |
|-------------|-------|----------------|
| 0           | 0.715 | 0.526          |
| 1           | 0.702 | 0.536          |
| 2           | 0.687 | 0.548          |
| 3           | 0.671 | 0.561          |
| 4           | 0.653 | 0.575          |
| 5           | 0.634 | 0.589          |
| 6           | 0.615 | 0.603          |
| 7           | 0.596 | 0.617          |

we find an  $R^2$  of 0.845 and an SE of 0.383. This is better than even the concurrent-based WW prediction model. This model is very simplistic in that we are essentially using last week's cases to predict this week's cases. Certainly, we could use forecasting techniques to do better than this.

Do WW measurements improve this last model? We consider the model:

$$\log c(t) = a_i + b_i \log c(t - 7) + d_i \log g(t - 7) \quad (7)$$

This results in an  $R^2$  of 0.851 and an SE of 0.375. This is a negligible improvement in predictive ability.

These results suggest that if you have good PH data, you do not gain anything with additional WW data. If you do not have PH data, you can use WW data to predict PH data, but it works best concurrently.

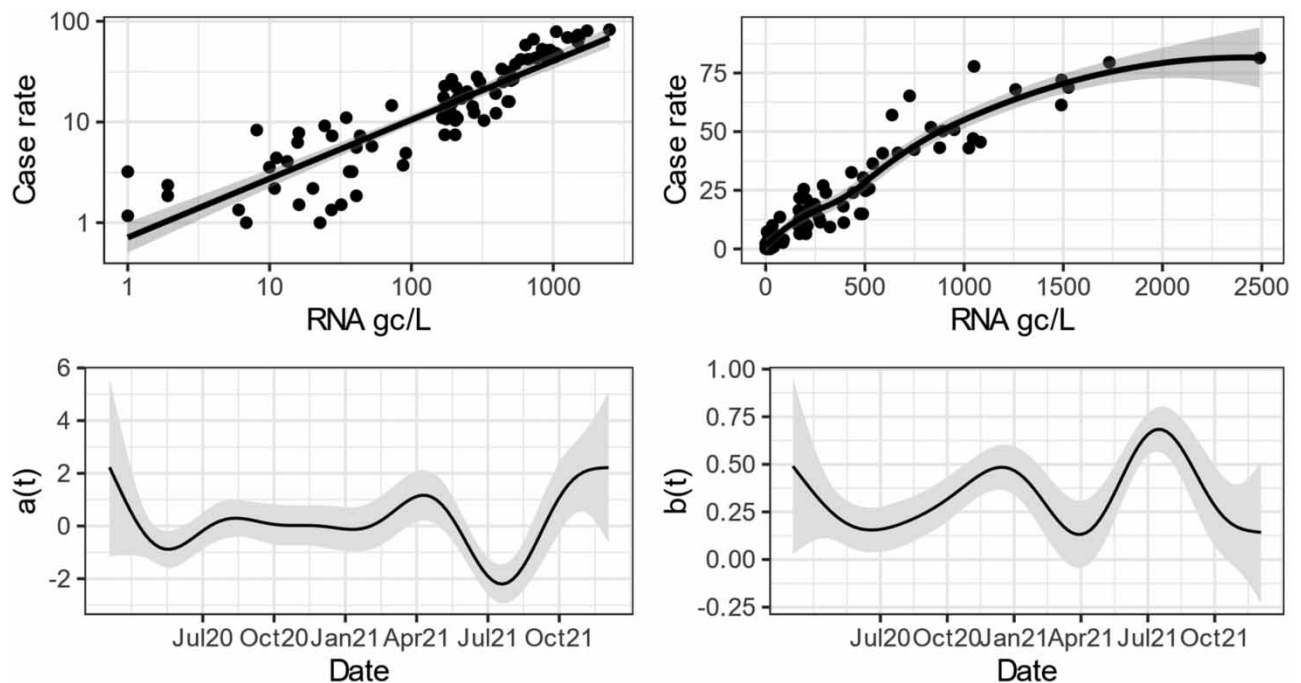
Our model framing with the PH measure as the response and the WW predictor carries the assumption that the PH measure is more reliable. If this assumption is correct and the PH data continue to be available, WW measures add little. But if PH measures are not reliable or do not continue to be available, the WW measures become valuable. Also, there is the possibility that PH measures are not consistently reliable due to testing failure (BBC 2021), in which case the WW measures provide an alternative estimate of disease prevalence.

### USA comparison

To assess the generalisability of our findings to other contexts, we repeated the analysis data from the USA provided publicly by Biobot (2021). Similar issues were seen in the USA data.

As seen in the top two plots of Figure 7, the relationship between cases and RNA appears approximately linear on logged scales, but the non-linearity is apparent on the untransformed scale. We show only Indiana county, Pennsylvania, here, but this is typical of the 30 counties in the data. There is greater heterogeneity in the amount of RNA observed per case in the USA data compared to Scotland, which reflects the greater geographical spread in the USA, differences in WW infrastructure, and variation in PH testing and reporting.

Time-varying relationships were also seen when the same model was fit to the USA as can be seen in the lower two plots of Figure 7. Again, greater heterogeneity was seen in the USA with no consistent patterns seen in the  $a(t)$ .



**Figure 7** | Top two plots show relationship between cases and RNA for Indiana county, PA, first on a logged scale and then on the untransformed scale. The bottom two plots show estimates from the time-varying model for the USA data. The lower left plot is the estimate  $a(t)$  for Indiana county, while the lower right plots show the estimated  $b(t)$ .

There was less scope to try lagged models with the USA data because of the weekly data. For a concurrent model predicting cases using RNA as in Equation (5), the value of  $R^2$  was 0.681 and the SE was 0.587, while for the week-lagged model, the value of  $R^2$  was 0.470 and the SE was 0.755. As with the Scotland data, concurrent prediction was best. The fit was just slightly worse for the USA data. Again, the WW data did not help a PH-only forecasting model.

## DISCUSSION

When building a predictive model, it is not sufficient to fit the present data well. It is necessary to consider how well the model will perform in the future. We consider three scenarios.

1. Large amounts of PH data continue to be available. We will be able to update our models continuously. We will be free to use complex models that require substantial PH data to estimate. This will be particularly valuable when a time-varying relationship persists. Some recommendations for the updating of prediction models may be found in [Jenkins \*et al.\* \(2021\)](#). With both PH- and WW-based estimates of disease prevalence, we will have some insurance against anomalies that may arise in either reporting system. This is the most expensive scenario since we must resource two systems of measurement.
2. A small amount of occasional PH data is available. Suppose we can normalise or adjust our WW measurements, such that our RNA concentration  $g$  follows the proportional relationship  $c=pg$ , where  $c$  is the case rate as discussed in Equation (1). A single observation of  $c$  obtained from a brief infection survey would suffice to estimate  $p$ . This scenario does not require that we maintain extensive PH testing infrastructure.
3. No PH data are available. This scenario would arise for a disease or location for which we have not been collecting data or is new. We would need to match this disease with a similar disease for which we know  $p$ . For example, if we have SARS-CoV-3, we can use our knowledge of SARS-CoV-2.

How can we improve the process? We can improve the WW measurement process in how the samples are drawn and analysed. With current levels of research, we can expect reductions in variation, which will result in more precise predictions. We can also increase sampling which will also reduce variation through smoothing and averaging, but this will also increase the cost. Beyond these desirable measurement process improvements, there is a strategic modelling choice.

We can build complex statistical or machine learning models that require large amounts of WW and PH data to setup. Along with helpful auxiliary environmental variables, we may achieve strongly predictive models. The drawback of this approach is that it requires substantial PH data. An alternative is to build a stronger understanding of the sewerage system by which individuals deposit quantities of RNA (or other substances), while lesser quantities of RNA are sampled at the WWTP. This requires physical models that describe the effects of environmental variables such as water quality or flow rates. Better methods of normalisation using biochemical markers would be valuable. Although promising, crAssphage has not been a clear success – see [Ai \*et al.\* \(2021\)](#) and [Greenwald \*et al.\* \(2021\)](#). The USA data discussed here used pepper mild mottle virus for normalisation, but the results were not better than the unnormalised Scottish data. If we could obtain reliable measures of  $g$  that follow the proportional  $c=pg$  relationship, calibrating the relationship to case rates will not require much PH data.

We have focussed on SARS-CoV-2, which is a relatively difficult target but it has generated much useful data. Some diseases are endemic and vary slowly – these targets are easily monitored using WBE. Other diseases, such as influenza, are mostly present at low levels but have occasional larger outbreaks. WBE can be useful here, particularly in the absence of much PH reporting. Reductions in the variation in WW measurement will help quickly distinguish real large changes in case rates from noise. The monitoring of novel disease outbreaks using established WBE infrastructure, before PH reporting can be widely deployed, requires some progress as discussed above.

In this article, we have focussed on the quantitative monitoring or prediction of disease. WBE has already been successful for other uses and has achieved good progress with the quantification problem. Further improvements in the measurement process and advanced understanding of how RNA and other substances progress through the sewerage system will enable more precise and less-expensive estimates and predictions of the disease.

## ACKNOWLEDGEMENTS

We are grateful to the Scottish government and Biobot for making the data available. Funding from the UKRI (EP/V028499/1) is greatly appreciated.

## DATA AVAILABILITY STATEMENT

Code and links to data are available from <https://github.com/julianfaraway/wbequant>.

## CONFLICT OF INTEREST STATEMENT

The authors declare there is no conflict.

## REFERENCES

- Ai, Y., Davis, A., Jones, D., Lemeshow, S., Tu, H., He, F., Ru, P., Pan, X., Bohrerova, Z. & Lee, J. 2021 Wastewater-based epidemiology for tracking COVID-19 trend and variants of concern in Ohio, United States. medRxiv: 2021.2006.2008.21258421. Preprint.
- Alhama, J., Maestre, J. P., Martín, M. Á. & Michán, C. 2021 Monitoring COVID-19 through SARS-CoV-2 quantification in wastewater: progress, challenges and prospects. *Microbial Biotechnology* **15** (6), 1719–1728.
- Baz-Lomba, J. A., Salvatore, S., Gracia-Lor, E., Bade, R., Castiglioni, S., Castrignanò, E., Causanilles, A., Hernandez, F., Kasprzyk-Hordern, B. & Kinyua, J. 2016 Comparison of pharmaceutical, illicit drug, alcohol, nicotine and caffeine levels in wastewater with sale, seizure and consumption data for 8 European cities. *BMC Public Health* **16** (1), 1–11.
- BBC 2021 Covid Test Lab in Wolverhampton Suspended Over Wrong Results. Available from: <https://www.bbc.co.uk/news/health-54923641>. (Accessed 13 June 2022).
- Biobot 2021. Available from: <https://biobot.io/data/>. (Accessed 13 June 2022).
- Castiglioni, S., Senta, I., Borsotti, A., Davoli, E. & Zuccate, E. 2015 A novel approach for monitoring tobacco use in local communities by wastewater analysis. *Tobacco Control* **24** (1), 38–42.
- Causanilles, A., Rojas Cantillano, D., Emke, E., Bade, R., Baz-Lomba, J. A., Castiglioni, S., Castrignanò, E., Gracia-Lor, E., Hernández, F., Kasprzyk-Hordern, B., Kinyua, J., McCall, A.-K., van Nuijs, A. L. N., Plósz, B. G., Ramin, P., Rousis, N. I., Ryu, Y., Thomas, K. V. & de Voogt, P. 2018 Comparison of phosphodiesterase type V inhibitors use in eight European cities through analysis of urban wastewater. *Environment International* **115**, 279–284.
- Duvallet, C., Wu, F., McElroy, K. A., Imakaev, M., Endo, N., Xiao, A., Zhang, J., Floyd-O'Sullivan, R., Powell, M. M., Mendola, S., Wilson, S. T., Cruz, F., Melman, T., Sathyanarayana, C. L., Olesen, S. W., Erickson, T. B., Ghaeli, N., Chai, P., Alm, E. & Matus, M. 2021 Nationwide trends in COVID-19 cases and SARS-CoV-2 wastewater concentrations in the United States. medRxiv: 2021.2009.2008.21263283. Preprint.
- Escolà Casas, M., Schröter, N. S., Zammit, I., Castaño-Trias, M., Rodríguez-Mozaz, S., Gago-Ferrero, P. & Corominas, L. 2021 Showcasing the potential of wastewater-based epidemiology to track pharmaceuticals consumption in cities: comparison against prescription data collected at fine spatial resolution. *Environment International* **150**, 106404.
- Feng, S., Roguet, A., McClary-Gutierrez, J. S., Newton, R. J., Kloczko, N., Meiman, J. G. & McLellan, S. L. 2021 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* **1** (8), 1955–1965.
- Fuller, W. A. 1987 *Measurement Error Models*. Wiley, New York.
- 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., Kantor, R. & 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**, 100111.
- Hoffmann, T. & Alsing, J. 2021 Faecal shedding models for SARS-CoV-2 RNA amongst hospitalised patients and implications for wastewater-based epidemiology. medRxiv: 2021.2003.2016.21253603. Preprint.
- Jenkins, D. A., Martin, G. P., Sperrin, M., Riley, R. D., Debray, T. P. A., Collins, G. S. & Peek, N. 2021 Continual updating and monitoring of clinical prediction models: time for dynamic prediction systems? *Diagnostic and Prognostic Research* **5** (1), 1.
- Kasprzyk-Hordern, B., Proctor, K., Jagadeesan, K., Lopardo, L., O'Daly, K. J., Standerwick, R. & Barden, R. 2021 Estimation of community-wide multi-chemical exposure via water-based chemical mining: key research gaps drawn from a comprehensive multi-biomarker multi-city dataset. *Environment International* **147**, 106331.
- Melvin, R. G., Chaudhry, N., Georgewill, O., Freese, R. & Simmons, G. E. 2021 Predictive power of SARS-CoV-2 wastewater surveillance for diverse populations across a large geographical range. medRxiv: 2021.2001.2023.21250376. Preprint.
- Morvan, M., Lo Jacomo, A., Souque, C., Wade, M. J., Hoffmann, T., Pouwels, K., Singer, A., Bunce, J. T., Engeli, A., Grimsley, J. M. S., 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. Preprint.
- O'Brien, J. W., Choi, P. M., Li, J., Thai, P. K., Jiang, G., Tschärke, B. J., Mueller, J. F. & Thomas, K. V. 2019 Evaluating the stability of three oxidative stress biomarkers under sewer conditions and potential impact for use in wastewater-based epidemiology. *Water Research* **166**, 115068.
- Ort, C., Banta-Green, C. J., Bijlsma, L., Castiglioni, S., Emke, E., Gartner, C., Kasprzyk-Hordern, B., Reid, M. J., Rieckermann, J. & Van Nuijs, A. L. 2014 Sewage-based epidemiology requires a truly transdisciplinary approach. *GAIA – Ecological Perspectives for Science and Society* **23** (3), 266–268.

- Public Health Scotland 2021. Available from: <https://www.opendata.nhs.scot/dataset/weekly-covid-19-statistical-data-in-scotland>. (Accessed 13 June 2022).
- Reid, M. J., Langford, K. H., Mørland, J. & Thomas, K. V. 2011 Analysis and interpretation of specific ethanol metabolites, ethyl sulfate, and ethyl glucuronide in sewage effluent for the quantitative measurement of regional alcohol consumption. *Alcoholism: Clinical and Experimental Research* **35** (9), 1593–1599.
- Ryu, Y., Gracia-Lor, E., Bade, R., Baz-Lomba, J. A., Bramness, J. G., Castiglioni, S., Castrignanò, E., Causanilles, A., Covaci, A., de Voogt, P., Hernandez, F., Kasprzyk-Hordern, B., Kinyua, J., McCall, A.-K., Ort, C., Plósz, B. G., Ramin, P., Rousis, N. I., Reid, M. J. & Thomas, K. V. 2016 Increased levels of the oxidative stress biomarker 8-iso-prostaglandin F<sub>2α</sub> in wastewater associated with tobacco use. *Scientific Reports* **6** (1), 39055.
- Scottish Environmental Protection Agency 2021. Available from: <https://informatics.sepa.org.uk/RNAmonitoring/>. (Accessed 13 June 2022).
- Thomas, K. V., Bijlsma, L., Castiglioni, S., Covaci, A., Emke, E., Grabic, R., Hernandez, F., Karolak, S., Kasprzyk-Hordern, B., Lindberg, R. H., Lopez de Alda, M., Meierjohann, A., Ort, C., Pico, Y., Quintana, J. B., Reid, M., Rieckermann, J., Terzic, S., van Nuijs, A. L. & de Voogt, P. 2012 Comparing illicit drug use in 19 European cities through sewage analysis. *Science of the Total Environment* **432**, 432–439.
- Venhuis, B. J., Zwaagstra, M. E., Keizers, P. H. & de Kaste, D. 2014 Dose-to-dose variations with single packages of counterfeit medicines and adulterated dietary supplements as a potential source of false negatives and inaccurate health risk assessments. *Journal of Pharmaceutical and Biomedical Analysis* **89**, 158–165.
- 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., Nejad, B. F., Morvan, M., Quintela-Balujá, M., Roberts, A. M. I., Singer, A. C., Souque, C., Speight, V. L., Sweetapple, C., Walkers, D., Watts, G., Weightman, A. & 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**, 127456.
- Wood, S. N. 2017 *Generalized Additive Models: An Introduction with R*. CRC Press/Taylor & Francis Group, Boca Raton.
- Wurtzer, S., Waldman, P., Levert, M., Cluzel, N., Almayrac, J. L., Charpentier, C., Masnada, S., Gillon-Ritz, M., Mouchel, J. M., Maday, Y. & Boni, M. 2022 SARS-CoV-2 genome quantification in wastewaters at regional and city scale allows precise monitoring of the whole outbreaks dynamics and variants spreading in the population. *Science of the Total Environment* **810**, 152213.
- Zulli, A., Pan, A., Bart, S. M., Crawford, F. W., Kaplan, E. H., Cartter, M., Ko, A. I., Cozens, D., Sanchez, M., Brackney, D. E. & Peccia, J. 2021 Predicting daily COVID-19 case rates from SARS-CoV-2 RNA concentrations across a diversity of wastewater catchments. *FEMS Microbes* **2**, xtab022.

First received 14 January 2022; accepted in revised form 27 May 2022. Available online 11 June 2022