Faecal shedding models for SARS-CoV-2 RNA amongst hospitalised patients and implications for wastewater-based epidemiology

Till Hoffmann*

Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom

Justin Alsing

Oskar Klein Centre, Department of Physics, Stockholm University, Stockholm SE-106 91, Sweden

The concentration of SARS-CoV-2 RNA in faeces is not well established, posing challenges for wastewater-based surveillance of COVID-19 and risk assessments of environmental transmission. We develop versatile hierarchical models for faecal RNA shedding and apply them to data collected in six studies. We find that the mean number of gene copies per mL of faeces is 1.9×10^6 (2.3×10^5—2.0×10^8 95% credible interval) amongst hospitalised patients. We find no evidence for a subpopulation of patients who do not shed RNA: limits of quantification can account for negative stool samples. Our models indicate that hospitalised patients represent the tail of the shedding profile with a half-life of 34 hours (28—43 95% credible interval), suggesting that wastewater-based surveillance signals are more indicative of incidence than prevalence and can be a leading indicator of clinical presentation. Shedding amongst inpatients cannot explain high RNA concentrations observed in wastewater, consistent with more abundant shedding during the early infection course.

*Corresponding author (t.hoffmann@imperial.ac.uk).

1. Introduction

The novel virus SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) infected over 80 million people in 2020, and more than 2.6 million people have succumbed to the resultant disease, COVID-19 (coronavirus disease 2019) [1]. Individuals infected by the virus primarily suffer from respiratory symptoms [2], but gastrointestinal manifestations of the disease have also been observed [3]. The presence of viral RNA in faeces allows for the surveillance of COVID-19 by quantifying gene copies in sewage [4]. So-called wastewater-based epidemiology (WBE) provides data that can complement traditional testing schemes and can be used to monitor the disease relatively cheaply by pooling wastewater from thousands of people [5]. Correlations between case numbers from individual testing schemes and RNA concentrations in wastewater have been observed [4, 6, 7]. However, associative studies cannot easily be used to calibrate WBE approaches: each sewerage system is different [8], and case numbers may not be a good indicator of prevalence [9]—especially when testing capacity is limited. The World Health Organisation considers "quantitative information on viral shedding" an imminent need to reap the potential benefits of WBE [10].

Faecal shedding of RNA suggests that the virus could be transmissible via the faecal-oral route [11]. While presence of RNA does not imply presence of infective virus, the likelihood increases with higher RNA loads [12]. Sewer overflows could cause spillover events, leading to new viral reservoirs [13]. For example, mink are susceptible to SARS-CoV-2 [14] and can be exposed to untreated sewage [13]. Furthermore, wastewater workers are at risk of contracting sewage-borne pathogens [15], and wastewater is a possible infection mode in densely populated communities [16, 17]. Quantifying these risks is essential for making informed policy decisions.

We developed a family of random-effect models [18, ch. 5] for SARS-CoV-2 RNA concentrations in faecal samples and applied them to data from six clinical studies to study three aspects of faecal RNA shedding. First, we studied the shedding profile, i.e. the temporal variability of shedding over the infection course, which affects the interpretation of WBE results [6]. We find that the profile decays quickly with a half-life of 34 hours. Second, the proportion of patients with one or more positive faecal samples has been extensively studied [3, 19, 20]. We determined that the limit of quantification of assays can account for patients without positive samples. Bayesian model comparison revealed no evidence for a subpopulation of patients who do not shed RNA faecally. Third, we obtained estimates of the mean faecal RNA concentration: a quantity important for inferring disease prevalence or incidence from wastewater data [21]. We show that the models are able to predict summary statistics of held-out studies accurately and consider the implications of our results for wastewater-based epidemiology.

2. Results

To study faecal RNA shedding quantitatively, we developed a suite of hierarchical models for RNA concentrations in faecal samples, as described in section 4.1. In contrast to

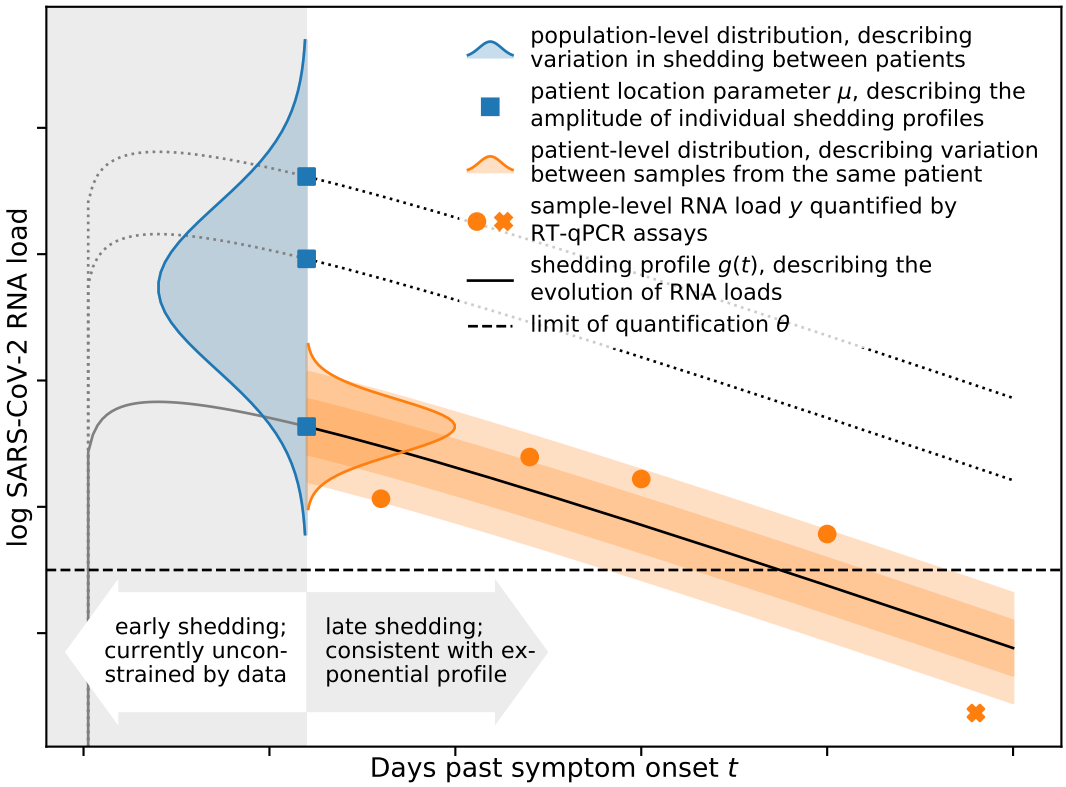

Figure 1: *A flexible hierarchical model can capture a wide range of aspects of real-world data.* The population-level distribution (shown in blue) captures variation in shedding behaviour between patients, giving rise to location parameters μ_i for each patient i (see section 4.1 for details). The location parameters (blue squares) describe the amplitude of individual shedding profiles as illustrated for three patients. The patient-level distribution (shown in orange for one patient) describes the variation between samples from the same patient, and the shedding profile (solid black line) modulates typical RNA concentrations in faecal samples over time. All samples are analysed using an RT-qPCR assay with a given limit of quantification (LOQ) θ (dashed black line), and concentrations above the LOQ (orange circles) can be quantified. Concentrations below the LOQ (orange cross) cannot be quantified. Currently available data do not allow us to constrain the early shedding profile, and late-stage shedding is consistent with an exponential decay profile.

	patients			samples			micro-	longi-	LOQ
study	n	+	−	m	+	−	data	tudinal	($\log_{10}\,\mathrm{mL}^{-1}$)
Wang et al. [11]	14	6	8	14	6	8	yes	no	1.4
Wölfel et al. [12]	9	8	1	82	68	14	yes	yes	2.0
Lui et al. [25]	11	11	0	43	23	20	yes	yes	2.8
Han et al. [26]	2	2	0	9	8	1	yes	yes	3.8
Kim et al. [27]	38	8	30	129	13	116	no	no	?
Ng et al. [28]	21	21	0	81	?	?	no	no	2.5

Table 1: *Six studies providing quantitative data on faecal RNA loads were analysed.* Microdata, i.e. sample-level RNA concentrations, provided by the first four studies were used to fit random effects models. The next two studies did not provide microdata, and we validated our models by comparing reported summary statistics with model-based predictions, as discussed in section 4.4. Samples with viral loads below the limit of quantification (LOQ) are considered negative for SARS-CoV-2. The table also reports the total number of patients n and number of samples m, including a breakdown by positivity.

existing quantitative approaches [22, 23], our models can account for a variable number of samples per patient, incorporate data from studies with different levels of quantification, and capture variability between patients as well as variability between samples from the same patient. The baseline model assumes that all infected patients shed viral RNA faecally and that typical RNA concentrations in samples vary over the course of the infection. We considered three different shedding profiles: an exponential decay profile, a gamma profile, and the exponential rise-and-decay profile proposed by Teunis et al. [24] for norovirus RNA shedding (see section 4.1.1 for details). We call this the *temporal standard* model, and it is illustrated in fig. 1.

We considered two modifications to the baseline model. First, we considered a *constant* model, where the time-variability of shedding was removed. Second, to assess whether there exists a subpopulation of patients who never shed viral RNA faecally, we introduced a shedding prevalence parameter ρ such that all samples of an infected patient are negative with probability $1 - \rho$. We call this the *subpopulation* model (as opposed to the *standard* model in which all patients shed RNA faecally). Combining the two modifications (with and without a subpopulation of non-shedders, and with and without time-variability) gives rise to four models in total.

We fitted each model to longitudinal data extracted from three studies [26, 25, 12] listed in table 1. All studies used RT-qPCR assays to quantify SARS-CoV-2 RNA copies in faecal samples collected from hospitalised patients. The two *constant* models were also fitted to an additional dataset collected by Wang et al. [11] that does not provide temporal information. As shown in table 2, we compared the models using the Bayesian model evidence [29] (i.e. the marginal likelihood of the data under each model), allowing us to draw two conclusions.

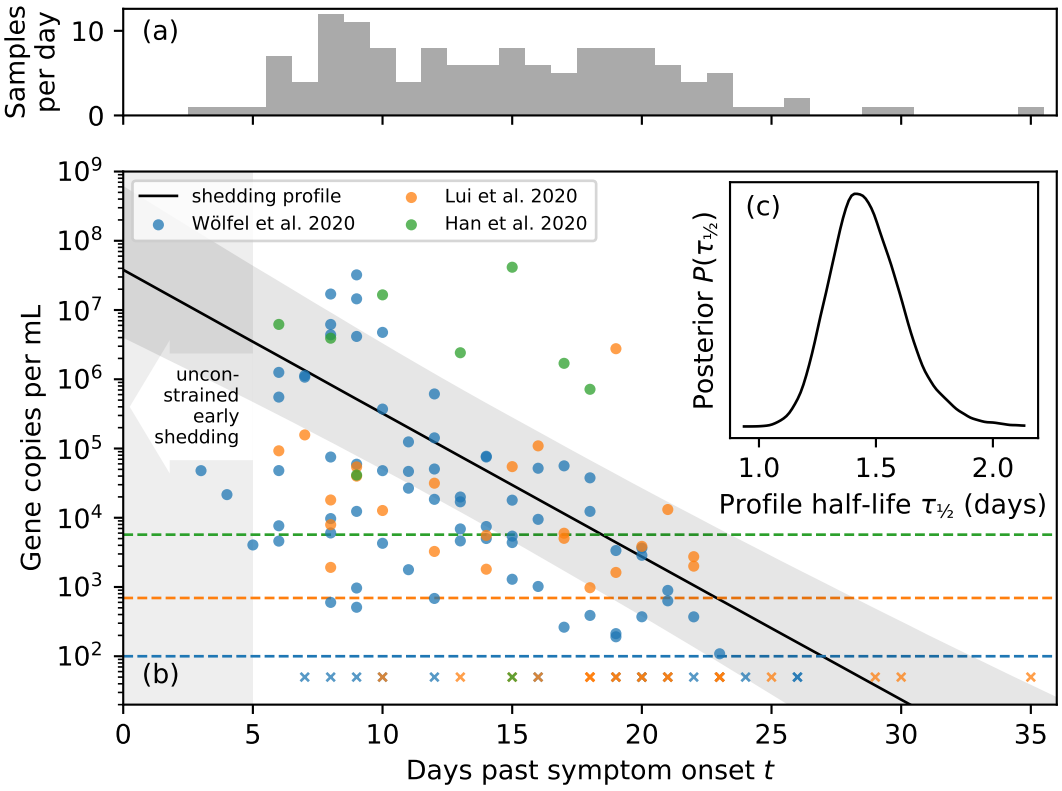

Figure 2: *Faecal RNA concentrations decay rapidly over the infection course.* Panel (a) shows a histogram of the number of samples collected on each day post symptom onset. The small number of samples collected within the first six days after symptom onset and the absence of samples collected prior to symptom onset make it difficult to constrain shedding during the early infection course. Panel (b) shows longitudinal faecal RNA concentration data from three studies together with the level of quantification for each study as dashed lines. The time-dependent exponential shedding profile of the *temporal standard* model is shown in black, and the shaded region represents the 95% credible interval of the profile (see section 4.1 for details). Panel (c) shows the posterior distribution for the half-life $\tau_{1/2}$ of the shedding profile.

temporal	subpopulation	log evidence	mean RNA conc. ($\log_{10}\mathrm{mL}^{-1}$)	half-life τ (hours)	shedding prevalence ρ
no	no	-1303.1 ± 0.2	6.28 [5.36, 8.30]	constant	1 (by defn.)
no	yes	-1305.5 ± 0.2	6.28 [5.40, 8.04]	constant	0.95 [0.74, 1.00]
yes	no	-1269.3 ± 0.2	variable	34 [28, 43]	1 (by defn.)
yes	yes	-1271.6 ± 0.2	variable	34 [28, 43]	0.97 [0.82, 1.00]

Table 2: *Accounting for temporal variability of the shedding profile is essential, and there is no evidence for a subpopulation of patients who do not shed RNA faecally.* Model evidences evaluated on three common datasets prefer *temporal* models (evidence shown for an exponential decay profile) over *constant* ones, and *standard* models are preferred over models with a *subpopulation* of patients who do not shed SARS-CoV-2 RNA faecally. Parameter estimates are consistent with conclusions based on model evidences. They are reported as the posterior mode together with the 95% credible interval in brackets. All reported credible intervals are highest posterior density intervals. Estimates for *constant* models include a fourth dataset without temporal information.

First, accounting for the time dependence of the shedding profile is essential. The three shedding profiles we considered are indistinguishable where data are available to constrain them (see section 4.1.1 for details). However, there are only few samples obtained prior to day six past symptom onset, as shown in fig. 2 (a), and shedding behaviour during the early infection course cannot be constrained given the available data. For simplicity, we use the exponential decay profile unless otherwise specified. Typical faecal RNA concentrations decay with a maximum a posteriori half-life of 34 hours (28—43 hours 95% credible interval) amongst hospitalised patients, as shown in fig. 2 (b) and (c).

Patients may not recall the number of days since symptom onset accurately, or they may present with atypical symptoms that are not easily identified as the onset of COVID-19 [30]. To assess the sensitivity of the inferred half-life to inaccurate reports, we repeated the inference after adding up to three days of reporting noise to the number of days since symptom onset. No sensitivity of the half-life inference to inaccurate reports was observed.

The second result is that there is no evidence for a subpopulation of patients who do not shed viral RNA faecally; *standard* models (without a non-shedding subpopulation) are preferred for both the *constant* (log odds of 3.1 ± 0.3) and *temporal* models (log odds of 1.8 ± 0.3). Consistent with the model comparison results, the inferred shedding prevalence is large and the 95% credible interval includes $\rho = 1$ for both *subpopulation* models. Consequently, as discussed in more depth in section 4.4, the level of quantification of the assays used in the three studies can explain the number of negative patients and samples because "the level of viral RNA present in stool can fluctuate around the margin of laboratory detection" [28].

To assess the out-of-sample predictive utility of the models, we considered predictions

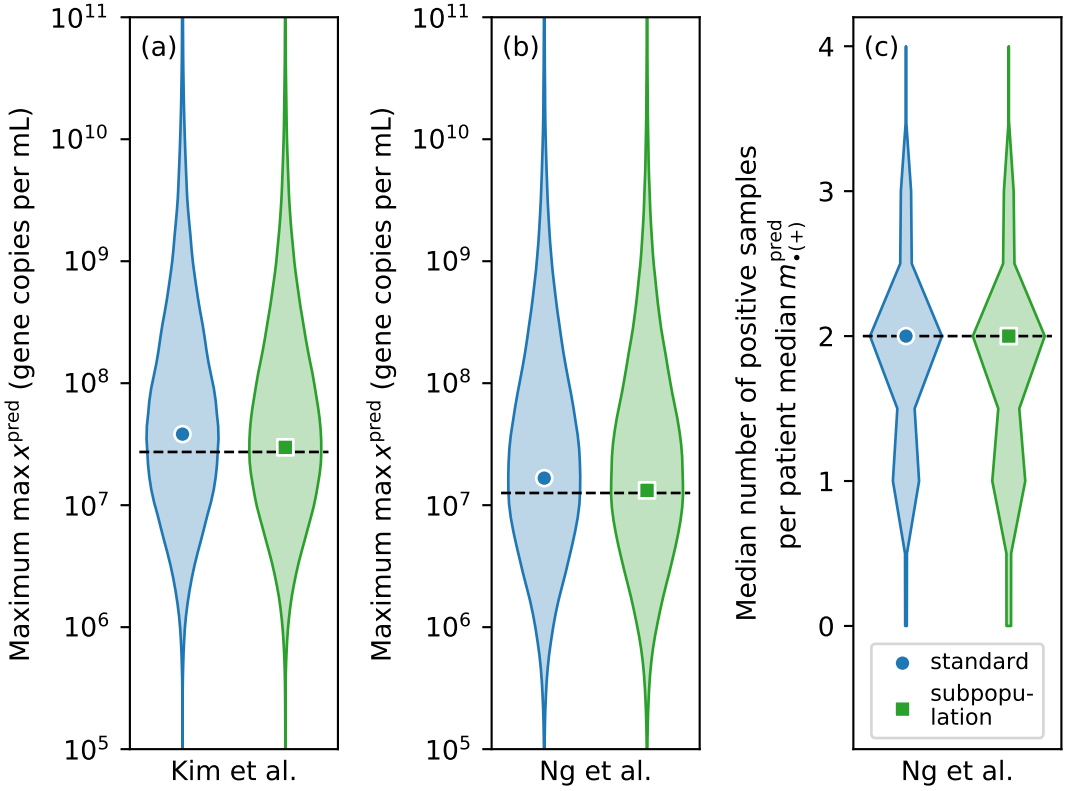

Figure 3: *The models can make accurate out-of-sample predictions.* Predictions of the maximum concentration observed by Kim et al. [27] and Ng et al. [28] are shown in panels (a) and (b) as violin plots, respectively. The value reported in the studies is shown as black dashed lines. Panel (c) shows predictions of the median number of positive samples per patient reported by Ng et al. [28]. Only results for the *constant* models are shown because the two studies did not provide temporal information.

for two held-out datasets that do not provide microdata, as listed in table 1. The studies conducted by Kim et al. [27] and Ng et al. [28] provided sufficiently detailed descriptions of their protocols to simulate the studies and make predictions by sampling from the posterior predictive distribution (see section 4.4 for details). Since our models have not been fit to these data, their ability to predict summary statistics of those data is an indication of how well the models generalise. *Temporal* models are preferred by the data, but we do not have any temporal information about the studies conducted by Kim et al. [27] and Ng et al. [28]. Nonetheless, the *constant* models can make accurate out-of-sample predictions because all studies consider the same population: hospitalised patients.

Kim et al. [27] collected 129 samples from 38 hospitalised patients, and they reported the largest observed concentration $\max x = 2.7 \times 10^7$ gene copies per mL of faeces. Ng et al. [28] collected 81 samples from 21 patients, and the largest observed RNA concentration was $\max x = 1.3 \times 10^7$ copies per mL. As shown in fig. 3 (a) and (b), predictions from our models are consistent with the reported values. Predictions of the maximum are smaller for the study by Ng et al. [28] than Kim et al. [27], which is expected owing to a smaller number of samples (so the tails of the distribution are less well sampled). Ng et al. [28] also reported the median number of positive samples per patient median $m_{\bullet(+)}$. Because they report the limit of detection of their assay, we can make predictions about the median number of positive samples per patient which agree with the reported value, as shown in panel (c) of fig. 3.

3. Discussion

We have inferred properties of the faecal SARS-CoV-2 RNA shedding distribution by fitting a suite of Bayesian hierarchical models to clinical data from four studies. The models account for the limits of quantification of RT-qPCR assays and a variable number of samples per patient. They are able to capture salient properties of the data and generalise well to two held-out datasets. There is no evidence of patients who do not shed viral RNA faecally.

The inferred temporal shedding profile is robust to inaccurate reports of the number of days since symptom onset, and it suggests that hospitalised patients are in the tail of the shedding profile because faecal RNA concentrations decay by an order of magnitude over the course of four to five days. While extrapolation should be treated with caution, wastewater-based surveillance of COVID-19 lends additional credibility to the hypothesis that SARS-CoV-2 RNA concentrations in wastewater are higher than expected based on faecal shedding inferred from hospitalised patients [31]. Assuming mean faecal RNA concentrations Λ are not larger in mild cases in the community than amongst hospitalised patients, a daily per capita wastewater volume of $V = 300\,\mathrm{L}$ [32], and faecal mass of $m = 128\,\mathrm{g}$ per person per day [33], we would expect wastewater RNA concentrations on the order of $m\Lambda/V \sim 10^3\,\mathrm{mL}^{-1}$ *if every person was infected.* In practice, concentrations in excess of $10^3\,\mathrm{mL}^{-1}$ have been observed [4] at times when seroprevalence of SARS-CoV-2 antibodies was less than ten percent [9]. Substantial shedding during the early stages of

the infection likely explains these observations. The rapid decay of the shedding profile also implies that signals from wastewater-based surveillance of SARS-CoV-2 are likely more indicative of incidence, rather than prevalence. Wastewater-based surveillance is thus a promising approach for early detection of cases in the community. A better understanding of the shape of the shedding profile, including prior to symptom onset, is essential for interpreting signals from WBE correctly during critical phases of rapid changes in levels of infection.

While one of the largest quantitative studies revealed no association between disease severity and faecal RNA concentration [34], results obtained from hospitalised patients are unlikely to apply to the general population, e.g. the former tend to be older and have more comorbidities. Faecal samples should be collected from a representative sample of patients over the entire infection course to refine quantitative estimates of faecal shedding of SARS-CoV-2 RNA. These data should include faecal volumes to estimate the total RNA load in faeces in addition to concentrations. The effect of vaccinations and emerging variants on faecal RNA shedding should also be investigated to make wastewater-based surveillance an effective *quantitative* monitoring tool.

4. Methods

4.1. Models

We developed a suite of hierarchical models to characterise faecal shedding of SARS-CoV-2 RNA quantitatively. Starting with the *standard constant* model described in section 2, we discuss the generative model in three steps.

First, the mean faecal RNA concentration λ_i for each patient $i \in \{1, \ldots, n\}$ follows a distribution with probability density function f, where n is the number of patients. Log-normal, Weibull, and gamma distributions are common choices for distributions that model positive, continuous data [35]. However, conclusions based on these distributions can differ substantially due to their tail behaviour [36]. We thus employed a generalised gamma distribution (GGD) with shape Q, location M, and scale S. The GGD is a flexible distribution which encompasses the log-normal distribution ($Q = 0$), Weibull distribution ($Q = 1$), and gamma distribution ($Q = S$) as special cases [37]. This choice comes at the cost of wider credible intervals, commensurate with our lack of prior knowledge about the shape of the shedding distribution.

Second, m_i samples are collected from each patient i. The RNA concentration y_{ij} in sample j from patient i follows a GGD with shape q, location μ_i, and scale σ. For the *constant* model (without time-varying shedding), the location parameter μ for the patient-level distribution is chosen such that $\langle y_{ij} \rangle = \lambda_i$ for each patient. Because of the properties of the generalised gamma distribution, we can express the location parameter in terms of the mean as

$$\mu_i = \log \lambda_i + -\frac{2\sigma}{q} \log q + \log \Gamma\left(\frac{1}{q^2}\right) - \log \Gamma\left(\frac{1 + q\sigma}{q^2}\right), \tag{1}$$

where Γ denotes the gamma function (see appendix A for details).

Third, the RNA concentration y_{ij} is quantified using an assay with level of quantification (LOQ) θ_{ij} (see Kitajima et al. [38] for an overview of different assays). If the concentration is below the LOQ, the sample is considered negative for the purpose of this study, and we denote the output of the assay as $x_{ij} = \circ$. Otherwise, the result of the assay faithfully captures the RNA concentration in the sample, i.e. $x_{ij} = y_{ij}$. We do not explicitly model measurement error or variability between technical replicates in this study because the relevant data are not available and assays tend to yield reproducible results. For example, the CDC N1 assay [39] has a coefficient of variation $< 3.7\%$—much smaller than typical variability between samples from the same patient. The RNA quantification is censored by the LOQ, and the likelihood of observing a particular assay result x_{ij} is thus

$$P\left(x_{ij}|q, \mu_i, \sigma, \theta_{ij}\right) = \begin{cases} F\left(\theta_{ij}; q, \mu_i, \sigma\right) & \text{if } x_{ij} = \circ \\ f\left(x_{ij}; q, \mu_i, \sigma\right) & \text{otherwise,} \end{cases} \tag{2}$$

where f denotes the probability density function of the generalised gamma distribution and F denotes the corresponding cumulative distribution function [37].

4.1.1. Time-variation

RNA shedding varies over the course of the infection [40], and the *constant* model cannot capture these changes. We incorporate temporal variability by introducing a shedding profile such that the expected RNA concentration λ_i for each patient i varies over time. In particular, we let $\lambda_i(t) = \lambda_i(t = 0)g(t)$, where t is the number of days since symptom onset, $g(t)$ is the shedding profile that modulates the expected RNA concentration, and $\lambda_i(t = 0)$ is sampled from the population-level distribution. The shape and scale parameters q and σ of the patient-level distribution are kept constant.

Because all available data were collected from hospitalised patients several days after the initial onset of symptoms, we can only constrain the later part of the shedding profile. We used an exponential profile $g_{\exp}(t) = \exp\left(-\alpha t\right)$, where α is the decay rate of the profile, because it provides an adequate fit for late-stage faecal shedding of other viruses [24]. Substituting into eq. (1), the exponential shedding profile gives rise to a location parameter that varies linearly with the number of days after symptom onset for each patient i such that $\mu_i(t) = \mu_i(t = 0) - \alpha t$.

Any reporting error associated with the number of days since symptom onset t can be compensated for by a corresponding change in $\mu_i(t = 0)$. This explains why the inferred shedding profile is robust to reporting inaccuracies, as discussed in section 2. The profile shown in fig. 2 (b) is $\exp\left(M - \alpha t\right)$, and it captures the decay of the effective population-level location parameter M over time. The profile half life is $\tau_{1/2} = \frac{\log 2}{\alpha}$.

We also considered two further shedding profiles to assess the sensitivity of our inferences to the choice of $g(t)$. First, we used a gamma shedding profile

$$g_\gamma(t) = (t - t_0)^\alpha \exp\left(-\beta(t - t_0)\right),$$

where $\alpha > 0$ and $\beta > 0$ control the shape of the profile and t_0 is the time at which shedding can first occur, i.e. $g(t < t_0) = 0$. Second, we considered the exponential

rise-and-decay profile proposed by Teunis et al. [24] for norovirus shedding

$$g_{\text{Teunis}}(t) = [1 - \exp(-\alpha(t - t_0))] \exp(-\beta(t - t_0)),$$

where α and β control the rise and decay of the profile, respectively.

4.1.2. Non-shedding subpopulation

To investigate whether there is a subpopulation of patients who do not shed SARS-CoV-2 RNA faecally, we extended the model by introducing a binary indicator $z_i \in \{0, 1\}$ for each patient i. If $z_i = 1$, the shedding behaviour of the patient is unchanged. If $z_i = 0$, patient i does not shed RNA, and $x_{ij} = \circ$ for all samples j. The indicator variables follow a Bernoulli distribution with probability ρ, i.e. the prevalence of shedding amongst the population of patients. This extension gives rise to what we refer to as the *subpopulation* models.

We know that $z_i = 1$ for any patient with one or more positive samples, and the likelihood follows eq. (2). However, for any patient i whose m_i samples $x_{i\bullet} = \{x_{i1}, \ldots, x_{im_i}\}$ are all below the LOQ, the likelihood is

$$P\left(x_{i\bullet} | q, \mu_i, \sigma, \theta_{i\bullet}\right) = (1 - \rho) + \rho \prod_{j=1}^{m_i} F\left(\theta_{ij}; q, \mu_i, \sigma\right), \tag{3}$$

where the first term accounts for a patient who cannot shed RNA ($z_i = 0$) and the second accounts for a patient whose samples are all below the LOQ ($z_i = 1$).

4.2. Data acquisition

We searched the literature for primary research that reported quantitative information on SARS-CoV-2 RNA loads in faeces, and we identified eleven studies of interest [3, 11, 12, 26, 27, 28, 34, 41, 42, 43]. Semi-quantitative studies reporting only cycle-threshold values were excluded.

Data from four studies were used to fit the hierarchical models described in section 4.1. First, Wölfel et al. [12] reported longitudinal faecal RNA concentrations for nine patients in Munich who had been in close contact with a common index case. Second, Lui et al. [25] followed the first eleven patients hospitalised due to COVID-19 in Hong Kong. Third, Han et al. [26] studied the viral dynamics of a neonate and her mother in Seoul. Fourth, Wang et al. [11] reported cycle threshold (C_t) values of RT-qPCR assays for faecal samples from a subset of patients admitted to three hospitals in China, and we transformed C_t values to RNA concentrations using conversion constants provided in the publication. Longitudinal information was not available. Table 1 lists information on the number of patients and samples for each study.

Two studies provided summary statistics of viral RNA concentrations in faecal samples, such as the highest concentration observed, together with sufficiently detailed descriptions of the study design to replicate the studies *in silico*, as described in section 4.4. In particular, Kim et al. [27] quantified RNA concentrations in faecal samples collected

from hospitalised patients, and Ng et al. [28] studied RNA concentrations in faecal samples to assess transmission risks associated with faecal microbiota transplants. These data are not used to fit the hierarchical models but instead to assess the out-of-sample predictive utility of our fitted models.

Five studies were excluded from the analysis despite providing quantitative information. First, using an assay originally developed for samples from ferrets [44], Jeong et al. [41] reported RNA concentrations of various specimen types, including faecal samples. However, the reported concentrations are orders of magnitude lower than reported in any of the other quantitative studies for any specimen type. Indeed, most measurements are below the limit of quantification of other studies. Second, Zhang et al. [42] reported detailed data in a figure, but it is not possible to extract the relevant information because of the low resolution of the figure, and the data could not be obtained by other means. Third, Zheng et al. [34] conducted a large study, comprising 85 patients and 842 faecal samples. Unfortunately, viral loads are only reported as the median for each patient (private communication). Finally, Pan et al. [43] and Cheung et al. [3] collected faecal RNA concentration data from 17 and 59 confirmed cases, respectively, but neither study provides information on the number of samples collected.

4.3. Parameter inference and model comparison

We used the nested sampler *polychord* [45] which draws samples from the posterior distribution and evaluates the model evidence, i.e. the marginal likelihood of the data under the model, as listed in table 2. The study by Wang et al. [11] was not used to estimate parameters for *temporal* models because longitudinal data were not available. Similarly, to facilitate a direct comparison between models, that study was also omitted from the evaluation of evidences. We used unit-scale half-Cauchy distributions for the shape parameters Q and q as well as the scale parameters S and σ. A flat prior on the interval 6—23 was used for the population location M which is sufficiently wide to avoid boundary effects. A flat prior on the unit interval was used for the shedding prevalence ρ. For models with temporal variability, a unit-scale Cauchy prior was used for the profile parameters α and β. We used uniform prior on the interval -14—7 for t_0 such that the onset of shedding can differ from the onset of symptoms. All inferences were run in triplicate, and no sensitivity to different pseudo-random number generator seeds was observed (Gelman-Rubin diagnostic $\hat{R} < 1.05$ for all models [46]).

In addition to the results discussed in section 2, we consider four technical implications of the inference here.

First, the gamma and Teunis shedding profiles are both consistent with the exponential shedding profile for late-stage shedding where data are available to constrain them. However, the available data cannot constrain early shedding behaviour, as illustrated by the wide range of shedding profiles consistent with the data shown in fig. 4. Miura et al. [22] reported tight posteriors for early shedding behaviour using the shedding profile proposed by Teunis et al. [24], but their inferences rely on the assumption that the time of symptom onset and shedding onset coincide (i.e. $t_0 = 0$), an assumption that is not yet supported by evidence.

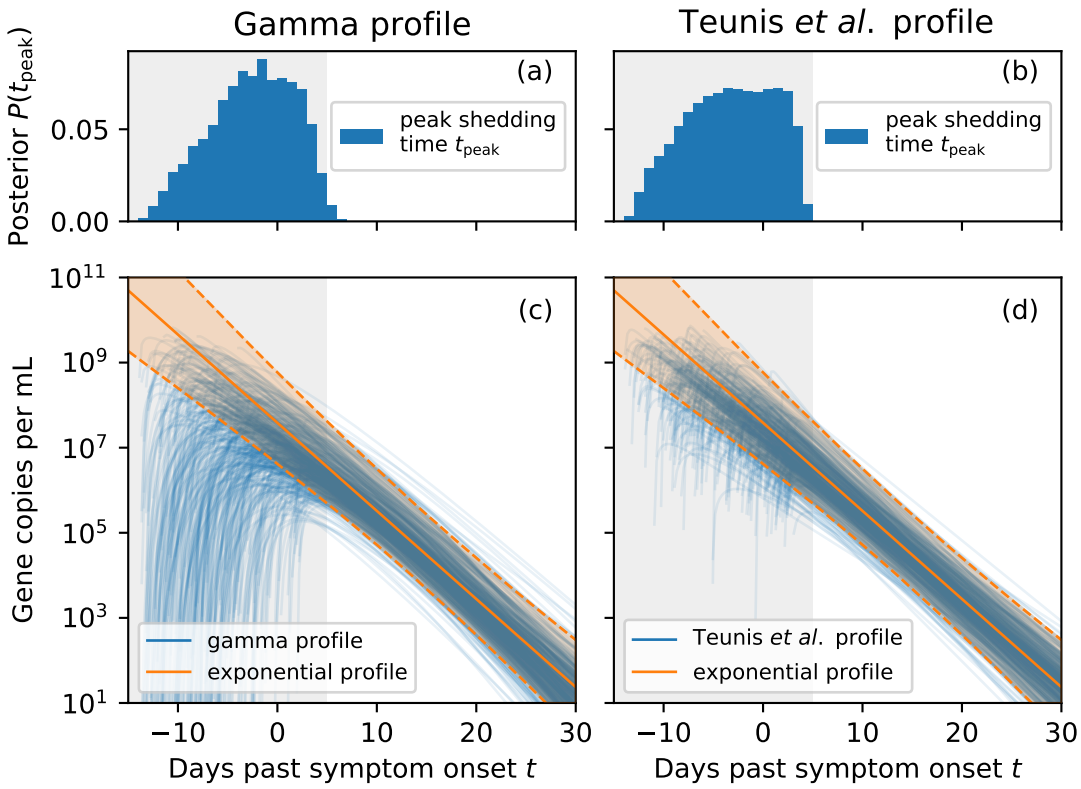

Figure 4: *Different shedding profiles yield consistent results for late-stage shedding, but early shedding behaviour (prior to day five after symptom onset) cannot be constrained given the available data.* Panels (a) and (b) show the the posterior distributions for the time t_{peak} at which the inferred gamma and Teunis shedding profiles peak, respectively. Because the data cannot constrain early shedding, the posterior for t_{peak} has broad support. Panels (c) and (d) show posterior samples of the gamma and Teunis profiles, respectively, in blue. The exponential profile discussed in section 2 is shown as an orange line, and the shaded region corresponds to the 95% credible interval.

Second, the shape Q controls the tails of the population-level distribution: the larger Q, the lighter the tails (see Prentice [37] for details). The population-level mean RNA concentration, i.e. the expected RNA concentration in a random faecal sample from a previously unobserved patient, depends on the tail behaviour because the mean is sensitive to outliers [36]. As shown in fig. 5 (a), the inferred mean under the *constant standard* model is larger for smaller Q because of the heavier tails. The corresponding credible intervals are also wider because constraining the mean requires more data for heavier-tailed distributions. Employing log-normal, Weibull, or gamma distributions would have entailed a poorly-motivated implicit choice about the shape of the distribution. Using generalised gamma distributions instead allows us to explicitly account for our prior uncertainty about the shape of the tails.

Third, the 95% credible region for the population-level shape Q and scale S is consistent with the log-normal distribution ($Q = 0$) or Weibull distribution ($Q = 1$) for both the *constant* and *temporal* models, as shown in fig. 5 (b). However, the special case of the gamma distribution ($Q = S$) can be confidently excluded because its tails are too light. The same conclusions apply to the patient-level distribution.

Fourth, the joint posterior distribution for the population-level shape and scale parameters are similar under both the *constant* and *temporal* models (exponential shedding profile) without a subpopulation of non-shedders, as shown in fig. 5 (b): modelling the temporal shedding profile does not have a significant effect on the inferred variability in shedding behaviour between patients. However, as shown in fig. 5 (c), the patient-level scale σ is significantly larger for the *constant* than the *temporal* model. The *constant* model can only account for large (time-integrated) sample-to-sample variability with a broad distribution. In contrast, the *temporal* model can explain variability between samples from the same patient using the shedding profile and a narrower distribution that accounts for residual idiosyncratic noise.

4.4. Posterior predictive model assessment

We assess the goodness-of-fit of the model to the data using posterior predictive replication of various summary statistics, and validate the ability of the fitted model to predict summary statistics of held-out datasets using posterior predictive validation.

Posterior predictive replication is a useful tool for assessing the fit of a model to data [18, ch. 6]. In short, synthetic replicates of the data generated by sampling from the posterior predictive distribution should not be easily distinguishable from the data the model was fit to. For example, 104 of 148 samples were positive in the composite dataset used to fit the *constant standard* model. If posterior predictive replicates of the data have a similar number of positive samples, the model is able to describe this particular aspect of the data. In contrast, the model would evidently not be suitable if it confidently predicted that all samples should be positive. We assess the goodness-of-fit to the data via posterior predictive replication of two key summary statistics.

The first summary statistics we consider are the number of positive samples $m_{(+)}$ and positive patients $n_{(+)}$ (i.e. patients with at least one positive sample) because these statistics have been extensively discussed in meta-analyses [3, 19, 20]. As shown in

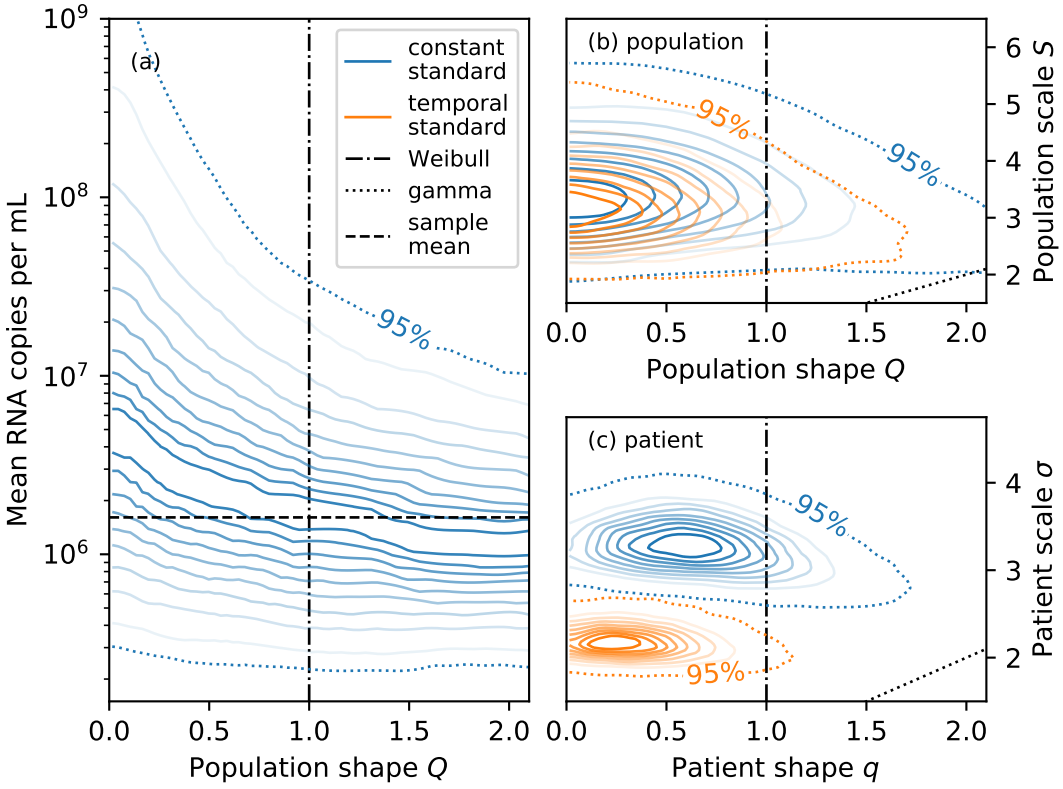

Figure 5: *The mean number of RNA copies per mL of faeces is sensitive to the shape of the shedding distribution.* Panel (a) shows the conditional distribution of the mean number of RNA copies per mL of faeces given the population shape parameter Q for the *constant standard* model. The mean tends to be larger for smaller Q due to heavier tails. Panels (b) and (c) show the joint distributions of the shape and scale parameters for the population-level distribution and patient-level distribution, respectively. Parametrisations corresponding to the Weibull distribution ($Q = q = 1$) and gamma distribution ($Q = S$ and $q = \sigma$) are shown as black dot-dashed and dotted lines, respectively. The patient-level scale parameter σ is larger for the *constant* than the *temporal* model (both without a non-shedding subpopulation) because the patient-level distribution needs to account for both temporal variability and idiosyncratic sample-to-sample variability.

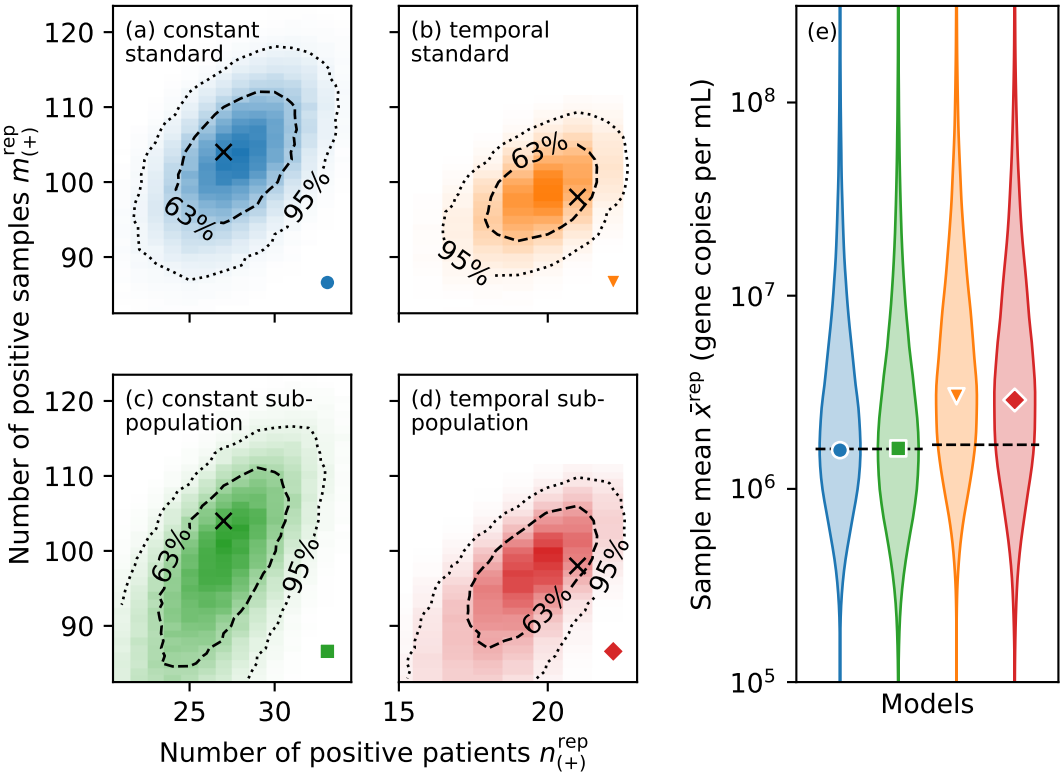

Figure 6: *Sampling from the posterior predictive distribution can replicate summary statistics of the data well.* Heat maps in panels (a) to (d) represent the joint distribution of the number of positive patients $n^{rep}_{(+)}$ and samples $m^{rep}_{(+)}$ under posterior predictive replication. The observed numbers (shown as black crosses) are consistent with the 63% credible region of the replications. Note that the values differ between the *temporal* and *constant* models because inference for the latter included an additional dataset that did not provide longitudinal information. Panel (e) show posterior predictive replications of the sample mean (excluding negative samples) for all models as violin plots. Markers represent the mode of the distribution, and the observed values are shown as black dashed lines.

fig. 6 (a) to (d), replicates from all models are consistent with the observed number of positive patients $n_{(+)}$ and number of positive samples $m_{(+)}$, corroborating our result that the limit of quantification of assays is sufficient to account for negative samples. Replicates from the two *subpopulation* models exhibit larger variability because assay results $x_{i\bullet}$ for samples from the same patient i are correlated due to their mutual dependence on the shedding indicator z_i.

Second, we evaluate the sample mean (i.e. the mean of all positive samples) across posterior predictive replicates and compare it with the sample mean of the data. The models can indeed replicate the statistic well, and the observed sample mean is consistent with the predicted sample means, as shown in fig. 6 (e).

Posterior predictive replication can only assess the models' ability to explain the data they were fit to. In other words, a model that satisfies such posterior predictive checks may nevertheless fail to generalise to other datasets. To assess the out-of-sample predictive utility of the models, we consider predictions for two held-out datasets.

Kim et al. [27] collected 129 samples from 38 hospitalised patients, and we assumed that at least one sample was collected from each patient. The remaining 91 samples were assumed to be collected from patients with equal probability. This information is sufficient to generate model-based predictions in two steps by sampling from the posterior predictive distribution. First, we sampled the patient-level means λ from the population-level distribution. Second, we sampled the assay results according to eqs. (2) and (3) for the *standard* and *subpopulation* models, respectively. While Kim et al. [27] also report the number of positive and negative samples, as shown in table 1, we could not replicate these summary statistics because they do not provide the limit of detection of their assay. Ng et al. [28] collected 81 samples from 21 patients, and we used the same allocation of samples to patients to generate predictions as for the study by Kim et al. [27].

Data and code availability

The data analysed during the current study and custom computer code are available at https://github.com/tillahoffmann/shedding.

Acknowledgements

This work was supported by the Natural Environment Research Council grant number NE/V010387/1. We thank Will Handley for discussions about the *polychord* sampler as well as Nick Jones, Kathleen O'Reilly, Vincent Savolainen, Emma Ransome, Mary Burkitt-Gray, and Christopher Coleman for comments on the manuscript.

References

[1] E. Dong, H. Du, and L. Gardner. "An interactive web-based dashboard to track COVID-19 in real time". In: *The Lancet Infectious Diseases* 20.5 (2020), pp. 533–534. DOI: 10.1016/s1473-3099(20)30120-1.

[2] C. Huang et al. "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China". In: *The Lancet* 395.10223 (2020), pp. 497–506. DOI: 10.1016/s0140-6736(20)30183-5.

[3] K. S. Cheung et al. "Gastrointestinal Manifestations of SARS-CoV-2 Infection and Virus Load in Fecal Samples from the Hong Kong Cohort and Systematic Review and Meta-analysis". In: *Gastroenterology* 159.1 (2020), pp. 81–95. DOI: 10.1053/j.gastro.2020.03.065.

[4] G. Medema, L. Heijnen, G. Elsinga, R. Italiaander, and A. Brouwer. "Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands". In: *Environmental Science & Technology Letters* 7.7 (2020), pp. 511–516. DOI: 10.1021/acs.estlett.0c00357.

[5] D. D. Polo et al. "Making waves: Wastewater-based epidemiology for SARS-CoV-2 – Developing robust approaches for surveillance and prediction is harder than it looks". In: *Water Research* (2020), p. 116404. DOI: 10.1016/j.watres.2020.116404.

[6] F. Wu et al. "SARS-CoV-2 titers in wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases". In: *medRxiv* (2020), p. 2020.06.15.20117747. DOI: 10.1101/2020.06.15.20117747.

[7] M. Wade et al. *Wastewater COVID-19 monitoring in the UK: summary for SAGE*. Tech. rep. Joint Biosecurity Centre, Department for Health and Social Care, 2020.

[8] A. P. Banks et al. "Potential impact of the sewer system on the applicability of alcohol and tobacco biomarkers in wastewater-based epidemiology". In: *Drug Testing and Analysis* 10.3 (2018), pp. 530–538. DOI: 10.1002/dta.2246.

[9] E. Slot et al. "Low SARS-CoV-2 seroprevalence in blood donors in the early COVID-19 epidemic in the Netherlands". In: *Nature Communications* 11.1 (2020), p. 5744. DOI: 10.1038/s41467-020-19481-7.

[10] Global Infectious Hazard Preparedness Team. "Status of environmental surveillance for SARS-CoV-2 virus". In: *WHO COVID-19 Scientific Briefs* (2020), WHO/2019–nCoV/Sci_Brief/EnvironmentalSampling/2020.1.

[11] W. Wang et al. "Detection of SARS-CoV-2 in Different Types of Clinical Specimens". In: *JAMA* 323.18 (2020), pp. 1843–1844. DOI: 10.1001/jama.2020.3786.

[12] R. Wölfel et al. "Virological assessment of hospitalized patients with COVID-2019". In: *Nature* 581.7809 (2020), pp. 1–10. DOI: 10.1038/s41586-020-2196-x.

[13] A. B. Franklin and S. N. Bevins. "Spillover of SARS-CoV-2 into novel wild hosts in North America: A conceptual model for perpetuation of the pathogen". In: *Science of The Total Environment* 733 (2020), p. 139358. DOI: 10.1016/j.scitotenv.2020.139358.

[14] N. Oreshkova et al. "SARS-CoV-2 infection in farmed minks, the Netherlands, April and May 2020". In: *Eurosurveillance* 25.23 (2020), p. 2001005. DOI: 10.2807/1560-7917.es.2020.25.23.2001005.

[15] J. W. Zabinski, K. J. Pieper, and J. M. Gibson. "A Bayesian Belief Network Model Assessing the Risk to Wastewater Workers of Contracting Ebola Virus Disease During an Outbreak". In: *Risk Analysis* 38.2 (2018), pp. 376–391. DOI: 10.1111/risa.12827.

[16] M. Kang et al. "Probable Evidence of Fecal Aerosol Transmission of SARS-CoV-2 in a High-Rise Building". In: *Annals of Internal Medicine* 173.12 (2020), pp. 974–980. DOI: 10.7326/m20-0928.

[17] J. Yuan et al. "Sewage as a Possible Transmission Vehicle During a Coronavirus Disease 2019 Outbreak in a Densely populated Community: Guangzhou, China, April 2020". In: *Clinical Infectious Diseases* (2020), ciaa1494. DOI: 10.1093/cid/ciaa1494.

[18] A. Gelman et al. *Bayesian Data Analysis*. Boca Raton: Chapman & Hall/CRC, 2014.

[19] A. S. van Doorn, B. Meijer, C. M. A. Frampton, M. L. Barclay, and N. K. H. Boer. "Systematic review with meta-analysis: SARS-CoV-2 stool testing and the potential for faecal-oral transmission". In: *Alimentary Pharmacology & Therapeutics* (2020). DOI: 10.1111/apt.16036.

[20] M. C. Wong et al. "Detection of SARS-CoV-2 RNA in fecal specimens of patients with confirmed COVID-19: A meta-analysis". In: *Journal of Infection* 81.2 (2020), e31–e38. DOI: 10.1016/j.jinf.2020.06.012.

[21] W. Ahmed et al. "First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community". In: *Science of The Total Environment* 728 (2020), p. 138764. DOI: 10.1016/j.scitotenv.2020.138764.

[22] F. Miura, M. Kitajima, and R. Omori. "Duration of SARS-CoV-2 viral shedding in faeces as a parameter for wastewater-based epidemiology: Re-analysis of patient data using a shedding dynamics model". In: *Science of The Total Environment* (2021), p. 144549. DOI: 10.1016/j.scitotenv.2020.144549.

[23] A. E. Benefield et al. "SARS-CoV-2 viral load peaks prior to symptom onset: a systematic review and individual-pooled analysis of coronavirus viral load from 66 studies". In: *medRxiv* (2020). DOI: 10.1101/2020.09.28.20202028.

[24] P. F. M. Teunis et al. "Shedding of norovirus in symptomatic and asymptomatic infections". In: *Epidemiology and Infection* 143.8 (2015), pp. 1710–1717. DOI: 10.1017/s095026881400274x.

[25] G. Lui et al. "Viral dynamics of SARS-CoV-2 across a spectrum of disease severity in COVID-19". In: *Journal of Infection* 81.2 (2020), pp. 318–356. DOI: 10.1016/j.jinf.2020.04.014.

[26] M. S. Han et al. "Viral RNA Load in Mildly Symptomatic and Asymptomatic Children with COVID-19, Seoul, South Korea - Volume 26, Number 10—October 2020 - Emerging Infectious Diseases journal - CDC". In: *Emerging Infectious Diseases* 26.10 (2020), pp. 2497–2499. DOI: 10.3201/eid2610.202449.

[27] J.-M. Kim et al. "Detection and Isolation of SARS-CoV-2 in Serum, Urine, and Stool Specimens of COVID-19 Patients from the Republic of Korea". In: *Osong Public Health and Research Perspectives* 11.3 (2020), pp. 112–117. DOI: 10.24171/j.phrp.2020.11.3.02.

[28] S. C. Ng, F. K. L. Chan, and P. K. S. Chan. "Screening FMT donors during the COVID-19 pandemic: a protocol for stool SARS-CoV-2 viral quantification". In: *The Lancet Gastroenterology & Hepatology* 5.7 (2020), pp. 642–643. DOI: 10.1016/s2468-1253(20)30124-2.

[29] A. Gelman, J. Hwang, and A. Vehtari. "Understanding predictive information criteria for Bayesian models". In: *Statistics and Computing* 24.6 (2014), pp. 997–1016. DOI: 10.1007/s11222-013-9416-2.

[30] J. M. Gan et al. "Atypical presentation of COVID-19 in hospitalised older adults". In: *Irish Journal of Medical Science* (2020), pp. 1–6. DOI: 10.1007/s11845-020-02372-7.

[31] F. Wu et al. "SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases". In: *mSystems* 5.4 (2020). DOI: 10.1128/msystems.00614-20.

[32] B. J. Tscharke et al. "Harnessing the Power of the Census: Characterizing Wastewater Treatment Plant Catchment Populations for Wastewater-Based Epidemiology". In: *Environmental Science & Technology* 53.17 (2019), pp. 10303–10311. DOI: 10.1021/acs.est.9b03447.

[33] C. Rose, A. Parker, B. Jefferson, and E. Cartmell. "The characterisation of faeces and urine; a review of the literature to inform advanced treatment technology". In: *Critical Reviews in Environmental Science and Technology* 45.17 (2015), pp. 1827–1879. DOI: 10.1080/10643389.2014.1000761.

[34] S. Zheng et al. "Viral load dynamics and disease severity in patients infected with SARS-CoV-2 in Zhejiang province, China, January-March 2020: retrospective cohort study". In: *BMJ* 369 (2020), p. m1443. DOI: 10.1136/bmj.m1443.

[35] R. F. Kappenman. "Estimation for the three-parameter Weibull, lognormal, and gamma distributions". In: *Computational Statistics & Data Analysis* 3 (1985), pp. 11–23. DOI: 10.1016/0167-9473(85)90054-4.

[36] D. B. Rubin. "Distinguishing between the Scale of the Estimand and the Transformation to Normality". In: *Journal of the American Statistical Association* 79.386 (1984), pp. 309–312. DOI: 10.1080/01621459.1984.10478046.

[37] R. L. Prentice. "A log gamma model and its maximum likelihood estimation". In: *Biometrika* 61.3 (1974), pp. 539–544. DOI: 10.1093/biomet/61.3.539.

[38] M. Kitajima et al. "SARS-CoV-2 in wastewater: State of the knowledge and research needs". In: *Science of The Total Environment* 739 (2020), p. 139076. DOI: 10.1016/j.scitotenv.2020.139076.

[39] X. Lu et al. "US CDC Real-Time Reverse Transcription PCR Panel for Detection of Severe Acute Respiratory Syndrome Coronavirus 2 - Volume 26, Number 8—August 2020 - Emerging Infectious Diseases journal - CDC". In: *Emerging Infectious Diseases* 26.8 (2020), pp. 1654–1665. DOI: 10.3201/eid2608.201246.

[40] M. Cevik et al. "SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis". In: *The Lancet Microbe* (2020). DOI: 10.1016/s2666-5247(20)30172-5.

[41] H. W. Jeong et al. "Viable SARS-CoV-2 in various specimens from COVID-19 patients". In: *Clinical Microbiology and Infection* (2020). DOI: 10.1016/j.cmi.2020.07.020.

[42] N. Zhang et al. "Comparative study on virus shedding patterns in nasopharyngeal and fecal specimens of COVID-19 patients". In: *Science China Life Sciences* (2020), pp. 1–3. DOI: 10.1007/s11427-020-1783-9.

[43] Y. Pan, D. Zhang, P. Yang, L. L. M. Poon, and Q. Wang. "Viral load of SARS-CoV-2 in clinical samples". In: *The Lancet Infectious Diseases* 20.4 (2020), pp. 411–412. DOI: 10.1016/s1473-3099(20)30113-4.

[44] Y.-I. Kim et al. "Infection and Rapid Transmission of SARS-CoV-2 in Ferrets". In: *Cell Host & Microbe* 27.5 (2020), 704–709.e2. DOI: 10.1016/j.chom.2020.03.023.

[45] W. J. Handley, M. P. Hobson, and A. N. Lasenby. "polychord: next-generation nested sampling". In: *Monthly Notices of the Royal Astronomical Society* 453.4 (2015), pp. 4384–4398. DOI: 10.1093/mnras/stv1911.

[46] A. Gelman and D. B. Rubin. "Inference from Iterative Simulation Using Multiple Sequences". In: *Statistical Science* 7.4 (1992), pp. 457–472. DOI: 10.1214/ss/1177011136.

A. Properties of the generalised gamma distribution

We use the parametrisation of the generalised gamma distribution presented by Prentice [37], and the random variable x follows a generalised gamma distribution if

$$\log x = \mu + \frac{\sigma}{q} \log\left(q^2 s\right),$$

where s is a gamma random variable with shape q^{-2}. The probability density function used in eq. (2) is thus

$$f(x;q,\mu,\sigma) = \frac{q}{\sigma x \Gamma\left(q^{-2}\right)}[s(x)]^{q^{-2}} \exp\left(-s(x)\right),$$
$$\text{where } s(x) = q^{-2}\exp\left(\frac{q}{\sigma}\left[\log x - \mu\right]\right).$$

Similarly, the cumulative distribution function is

$$F(x;q,\mu,\sigma) = \frac{\gamma\left(q^{-2}, s(x)\right)}{\Gamma\left(q^{-2}\right)},$$

where γ is the lower incomplete gamma function.

The expected value is

$$\lambda = \exp(\mu) q^{2\sigma/q} \frac{\Gamma\left(q^{-2} + \sigma/q\right)}{\Gamma\left(q^{-2}\right)},$$

and expressing the location parameter μ in terms of the mean λ yields eq. (1). The coefficient of variation is

$$\sqrt{\frac{\Gamma\left(q^{-2}\right)\Gamma\left(q^{-2}+2\sigma/q\right)}{\left[\Gamma\left(q^{-1}+\sigma/q\right)\right]^2} - 1},$$

and the generalised gamma distribution has a constant coefficient of variation independent of the location parameter μ.