Supplemental Information

Modeling Infection from SARS-CoV-2 Wastewater Concentrations: Promise, Limitations, and Future Directions

Authors: Jeffrey Soller[1], Wiley Jennings[2], Mary Schoen[1], Alexandria Boehm[3], Krista Wigginton[4], Raul Gonzalez[5], Katherine E. Graham[3], Graham McBride[6], Amy Kirby[2], and Mia Mattioli[2]

[1] Soller Environmental, LLC, 3022 King St, Berkeley, CA 94703, USA
[2] Waterborne Disease Prevention Branch, Division of Foodborne, Waterborne, and Environmental Diseases, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA
[3] Stanford University Department of Civil and Environmental Engineering, Stanford, California, USA
[4] Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor 48109, Michigan, USA
[5] Hampton Roads Sanitation District, 1434 Air Rail Avenue, Virginia Beach, VA 23455, USA
[6] National Institute of Water & Atmospheric Research Ltd (NIWA), Hillcrest, Hamilton, New Zealand (retired)

Sections:

SI A. Mathematical Specifications of Model Formulations and Evaluation Metrics

In this study, we consider three mechanistic model formulations to map measurements of SARS-CoV-2 in wastewater to infection prevalence in the associated sewershed. Conceptually, all three formulations are similar. Wastewater samples can be thought of as a fecal sample pooled from the sewershed population, diluted in wastewater and subjected to decay and loss processes in the sewer system. Thus, theoretically, infection prevalence can be estimated from SARS-CoV-2 RNA measurements in wastewater after adjusting them for dilution, decay, and loss in the sewage system, and accounting for variation in clinical features of fecal shedding among SARS-CoV-2 infections.

The three model formulations below parameterize this conceptual model using 1) per-capita wastewater generation (estimated from per-capita domestic potable water consumption, adjusted for wastewater generation from non-domestic sources); 2) per-capita wastewater generation (estimated from in-situ sampling site flow measurements and sewershed population estimates); and 3) an endogenous human fecal control that serves as a proxy for dilution and loss in the sewer system and through lab processing. These three model formulations translate directly into the approaches commonly taken for representing SARS-COV-2 concentrations for wastewater surveillance: 1) unaltered SARS-CoV-2 concentrations; 2) SARS-CoV-2 concentrations normalized by flow and population; and 3) SARS-CoV-2 concentrations normalized by an endogenous control such as pepper mild mottle virus (PMMoV).

A.1 Flow generation formulation

The flow generation model formulation relies on estimates of per-capita wastewater input into the system, as well as the volume fraction of wastewater coming from a source that could potentially contain SARS-CoV-2. To estimate the fraction of the population with an active infection (F_{active}), the flow generation model can be parameterized using a time-invariant flow composition parameter (F_{cont}) for the fraction of human contribution to the waste stream (versus commercial, industrial, run-off, etc.) as follows:

$$F_{active} = \frac{D_{WW}V_{WW}}{(D_{stool})V_{stool}(e^{-k_{virus}T_{sewer}})F_{shed}(F_{cont})} \quad \text{Eq. A1.1}$$

Where:

D_{WW} = SARS-CoV-2 density in wastewater influent (viral copies/L)
D_{stool} = SARS-CoV-2 density in stool (viral copies/L)
V_{ww} = wastewater generated per person per day (L/day)
V_{stool} = stool generated per person per day (L/day)
F_{shed} = fraction of infected individuals that shed viral RNA in feces (unitless)
k_{virus} = pseudo-first order decay coefficient for SARS-CoV-2 RNA in sewage (1/day)
T_{sewer} = time wastewater spends in the sewer prior to reaching the wastewater treatment plant (days)
F_{cont} = volume fraction of wastewater coming from a source that could potentially contain SARS-CoV-2 (unitless)

Note that the pathogen stool density and stool volume can be broken into diarrheal and non-diarrheal terms as follows:

$$D_{stool}V_{stool} = D_{diar}V_{diar}F_{diar} + D_{nodiar}M_{nodiar}(1 - F_{diar}) \quad \text{Eq. A1.2}$$

Where:

D_{diar} = SARS-CoV-2 density in feces among infected individuals with diarrhea (viral copies/mL)
V_{diar} = volume of feces per person per day among infected individuals with diarrhea (mL/day)
F_{diar} = fraction of infected individuals with diarrhea (unitless)
D_{nodiar} = SARS-CoV-2 density in feces among infected individuals without diarrhea (viral copies/g)
M_{nodiar} = mass of feces per person per day among infected individuals without diarrhea (g/day)

The flow generation model could thus be parameterized as:

$$F_{active} = \frac{D_{WW}V_{WW}}{[D_{diar}V_{diar}F_{diar} + D_{nodiar}M_{nodiar}(1-F_{diar})]\, e^{-k_{virus}T_{sewer}}\, F_{shed}F_{cont}} \quad \text{Eq A1.3}$$

The same substitution applies to the following model formulations as well, but they will be written in the simpler parameterization to facilitate comparison across model formulations.

A.2 Flow receipt formulation

Like the flow generation formulation, the flow receipt formulation relies on flow rate estimates. However, these flow estimates are measured at the sampling site (which is often a treatment plant) (V_{WWTP}), which enables them to vary from sample-to-sample. To put flow on a per-capita basis, this formulation also requires an explicit sewershed population estimate (N). This formulation can be derived from a mass balance-type equation on the number of RNA copies in wastewater, written in words as:

Daily viral RNA copies in wastewater influent	=	Daily per capita viral RNA copies in stool	x	Wastewater decay factor	x	Number of people shedding virus in stool

$$D_{WW}V_{WWTP} = D_{stool}V_{stool}\, e^{-k_{virus}T_{sewer}}\, F_{shed}\, I_{active} \quad \text{Eq. A2.1a}$$

and

$$I_{active} = F_{active}N \quad \text{Eq. A2.1b}$$

Thus,

$$\frac{D_{WW}V_{WWTP}}{N} = D_{stool}V_{stool}\, e^{-k_{virus}T_{sewer}}\, F_{shed}\, F_{active} \quad \text{Eq. A2.2}$$

or

$$F_{active} = \frac{D_{WW} V_{WWTP}}{D_{stool} V_{stool} e^{-k_{virus} T_{sewer}} F_{shed} N} \quad \text{Eq. A2.3}$$

A.3 Fecal strength formulation

Both the flow generation and flow receipt formulations account for the fecal strength of a wastewater by incorporating the per-capita volumetric flow of the wastewater system (in the flow generation formulation, this is V_{WW}/F_{cont}; in the flow receipt formulation, this is V_{WWTP}/N). Another approach to account for the fecal strength of a wastewater is by measuring an endogenous control – a wastewater constituent that is present in relatively stable, high, and known concentrations in human excreta. Fecal strength can then be written as a dimensionless parameter describing the fraction, theoretically bounded by [0, 1], of wastewater that is human feces. This parameterization of fecal strength (FS) is the ratio of the concentration (density) of the endogenous control measured in wastewater ($D_{WW-endog}$), to the expected wastewater endogenous control concentration if only domestic wastewater sources were contributing (e.g., sources of wastewater that could potentially contain SARS-CoV-2) and adjusted for decay in the sewer. A fecal strength of 1 means that domestic wastewater is undiluted by other sources.

$$FS = \frac{D_{WW-endog}}{e^{-k_{endog} T_{sewer}} (D_{stool-endog} M_{stool} / V_{WW})} \quad \text{Eq. A3.1}$$

Where

k_{endog} = pseudo-first order decay coefficient for endogenous control in sewage (1/day)
$D_{stool\text{ - }endog}$ = endogenous control density in stool (gc/g)
M_{stool} = stool mass generated per person per day (g/day)

The dimensional analysis of fecal strength is as follows:

$$[-] = \frac{\frac{gc}{L}}{\exp\left(\frac{hr}{hr}\right)\left(\frac{\frac{gc}{g}\frac{g}{day-person}}{\left(\frac{L}{day-person}\right)}\right)} = \frac{\left(\frac{L}{day-person}\right)\frac{gc}{L}}{\frac{gc}{g}\frac{g}{day-person}}$$

This dimensionless parameterization can be thought of as another way to estimate F_{cont}, the volume fraction of wastewater coming from a source that could potentially contain SARS-CoV-2, in the flow generation formulation. Substituting in FS for F_{cont} in the flow generation formulation, and adding 'SARS' subscripts to parameters to clearly distinguish them from 'endog' parameters:

$$F_{active-SARS} = \frac{D_{WW-SARS} V_{WW}}{D_{stool-SARS} V_{stool} e^{-k_{SARS} T_{sewer}} F_{shed-SARS} \frac{D_{WW-endog}}{e^{-k_{endog} T_{sewer}} (D_{stool-endog} M_{stool} / V_{WW})}} \quad \text{Eq. A3.2}$$

$$= \frac{D_{WW-SARS} V_{WW} e^{-k_{endog} T_{sewer}} D_{stool-endog} M_{stool}}{D_{stool-SARS} V_{stool} e^{-k_{SARS} T_{sewer}} F_{shed-SARS}\ D_{WW-endog} V_{WW}} \quad \text{Eq. A3.3}$$

Assuming volume of stool is given in mL and mass of stool is given in g, then M_{stool} / V_{stool} can be taken to equal 1 g per ml (Penn et al. 2018).

$$F_{active-SARS} = \frac{D_{stool-endog} D_{WW-SARS} e^{-k_{endog} T_{sewer}}}{D_{stool-SARS} D_{WW-endog} e^{-k_{SARS} T_{sewer}} F_{shed-SARS}} \quad \text{Eq. A3.4}$$

Assuming that the decay rates of the endogenous control and SARS-CoV-2 are negligible or the same, which is most plausible for an RNA viral control such as PMMoV:

$$F_{active-SARS} = \frac{D_{stool-endog} D_{WW-SARS}}{D_{stool-SARS} D_{WW-endog} F_{shed-SARS}} \quad \text{Eq. A3.5}$$

Alternatively, this endogenous control formulation can also be derived from the flow receipt formulation. To do this, the population, *N*, is estimated from endogenous control measurements by re-writing eq. A2.3 in terms of the endogenous control, and substituting back into eq. A2.

$$\frac{D_{WW-endog} V_{WWTP}}{D_{stool-endog} V_{stool} e^{-k_{endog} T_{sewer}} F_{shed-endog} N} = F_{active-endog} \quad \text{Eq. A3.6}$$

Rearranging to solve for N:

$$\frac{D_{WW-endog} V_{WWTP}}{D_{stool-endog} V_{stool} e^{-k_{endog} T_{sewer}} F_{shed-endog} F_{active-endog}} = N \quad \text{Eq. A3.7}$$

Assuming that all individuals shed the endogenous control: $F_{shed\text{-}endog} = F_{active\text{-}endog} = 1$. Therefore:

$$\frac{D_{WW-endog} V_{WWTP}}{D_{stool-endog} V_{stool} e^{-k_{endog} T_{sewer}}} = N \quad \text{Eq. A3.8}$$

Substituting eq. A3.8 for *N* into eq. A2.1:

$F_{active-SARS} =$

$$\frac{D_{WW-SARS}V_{WWTP}}{D_{stool-SARS}V_{stool}\,e^{-k_{SARS}T_{sewer}}\,F_{shed-SARS}\frac{D_{WW-endog}V_{WWTP}}{D_{stool-endog}V_{stool}\,e^{-k_{endog}T_{sewer}}}} \quad \text{Eq. A3.9}$$

$$= \frac{D_{stool-endog}V_{stool}e^{-k_{endog}T_{sewer}}D_{WW-SARS}V_{WWTP}}{D_{stool-SARS}V_{stool}\,e^{-k_{SARS}T_{sewer}}\,F_{shed-SARS}D_{WW-endog}V_{WWTP}} \quad \text{Eq. A3.10}$$

$$= \frac{D_{stool-endog}V_{stool}e^{-k_{endog}T_{sewer}}D_{WW-SARS}}{D_{stool-SARS}V_{stool}\,e^{-k_{SARS}T_{sewer}}\,F_{shed-SARS}D_{WW-endog}} \quad \text{Eq. A3.11}$$

Assuming either that decay of SARS and the endogenous control are negligible or the same eq. A3.11 simplifies to eq. A3.5:

$$F_{active-SARS} = \frac{D_{stool-endog}D_{WW-SARS}}{D_{stool-SARS}D_{WW-endog}F_{shed-SARS}} \quad \text{Eq. A3.5}$$

A.4 Ascertainment Ratio

A case ascertainment ratio (AR) can be defined as the ratio of infections to reported cases. Defining a case ascertainment ratio in terms of infections estimated from wastewater requires considering the time frame over which infections would be detected in wastewater, since fecal shedding may extend for weeks, albeit at low levels. Thus, a wastewater-based case AR can be defined as

$$AR = \frac{NF_{active-SARS}}{f(confirmed\ cases)} \quad \text{Eq. A4.1}$$

$$where: \quad f(confirmed\ cases) = sum\ of\ cases\ over\ T\ days = \sum_{t=1}^{T} Cases_t$$

Where the function of confirmed cases is a summation of cases over a defined time interval, *T*, defined as the average number of days the virus is shed in feces of infected individuals. An AR can be defined for each of the three model formulations.

AR 1: wastewater generation formulation:

$$AR1 = \frac{\frac{ND_{WW}V_{WW}}{D_{stool}V_{stool}e^{(-k_{SARS}T_{sewer})}F_{shed}F_{cont}}}{f(confirmed\ cases)} \quad \text{Eq. A4.2}$$

AR 2: wastewater receipt formulation:

$$AR2 = \frac{\frac{ND_{WW}V_{WWTP}}{D_{stool}V_{stool}e^{(-k_{SARS}T_{sewer})}F_{shed}N}}{f(confirmed\ cases)} \quad \textit{Eq. A4.3}$$

AR 3: endogenous control formulation:

$$AR3 = \frac{\frac{^{ND}_{stool-endog}{}^{D}_{WW-SARS}}{^{D}_{stool-SARS}{}^{F}_{shed-SARS}{}^{D}_{WW-endog}}}{f(confirmed\ cases)} \quad \text{Eq. A4.4}$$

A.5 Relative Ascertainment Ratio

Given the numerous uncertainties in estimating infection prevalence from wastewater measurements, a relative AR (RAR) measure can be developed to emphasize trends in AR over time, rather than absolute prevalence estimates. This is the ratio of AR estimated over time to the AR estimated on some reference date.

$$RAR = \frac{\frac{F_{active-SARS}}{f(confirmed\ cases)}}{\left(\frac{F_{active-SARS}}{f(confirmed\ cases)}\right)_{ref}} \quad \text{Eq. A5.1}$$

RAR can be written for each model formulation.

RAR 1: wastewater generation formulation:

$$RAR1 = \frac{\frac{\frac{D_{WW}V_{WW}}{D_{stool}V_{stool}e^{(-k_{SARS}T_{sewer})}F_{shed}F_{cont}}}{f(confirmed\ cases)}}{\left(\frac{\frac{D_{WW}V_{WW}}{D_{stool}V_{stool}e^{(-k_{SARS}T_{sewer})}F_{shed}F_{cont}}}{f(confirmed\ cases)}\right)_{ref}} \quad \text{Eq. A5.2}$$

$$= \frac{\frac{D_{WW}}{f(confirmed\ cases)}}{\left(\frac{D_{WW}}{f(confirmed\ cases)}\right)_{ref}} \frac{V_{WW}D_{stool}V_{stool}e^{(-k_{SARS}T_{sewer})}F_{shed}F_{cont}}{V_{WW}D_{stool}V_{stool}e^{(-k_{SARS}T_{sewer})}F_{shed}F_{cont}} \quad \text{Eq. A5.3}$$

$$RAR1 = \frac{\frac{D_{WW}}{f(confirmed\ cases)}}{\left(\frac{D_{WW}}{f(confirmed\ cases)}\right)_{ref}} \quad \text{Eq. A5.4}$$

RAR 2: wastewater receipt formulation:

$$RAR2 = \frac{\frac{\frac{D_{WW}V_{WWTP}}{D_{stool}V_{stool}e^{(-k_{SARS}T_{sewer})}F_{shed}N}}{f(confirmed\ cases)}}{\left(\frac{\frac{D_{WW}V_{WWTP}}{D_{stool}V_{stool}e^{(-k_{SARS}T_{sewer})}F_{shed}N}}{f(confirmed\ cases)}\right)_{ref}} \quad \text{Eq. A5.5}$$

$$= \frac{\frac{D_{WW} V_{WWTP}}{f(confirmed\ cases)}}{\left(\frac{D_{WW} V_{WWTP}}{f(confirmed\ cases)}\right)_{ref}} \frac{F_{shed} N e^{(-k_{SARS} T_{sewer})} D_{stool} V_{stool}}{F_{shed} N e^{(-k_{SARS} T_{sewer})} D_{stool} V_{stool}} \quad \text{Eq A5.6}$$

$$RAR2 = \frac{\frac{D_{WW} V_{WWTP}}{f(confirmed\ cases)}}{\left(\frac{D_{WW} V_{WWTP}}{f(confirmed\ cases)}\right)_{ref}} \quad \text{Eq A5.6}$$

RAR 3: endogenous control formulation:

$$RAR3 = \frac{\frac{D_{stool-endog} D_{WW-SARS}}{D_{stool-SARS} F_{shed-SARS} D_{WW-endog}}}{\left(\frac{D_{stool-endog} D_{WW-SARS}}{D_{stool-SARS} F_{shed-SARS} D_{WW-endog}}\right)_{ref}} \quad \text{Eq A5.7}$$

$$= \frac{\frac{\frac{D_{WW-SARS}}{D_{WW-endog}}}{f(confirmed\ cases)}}{\left(\frac{\frac{D_{WW-SARS}}{D_{WW-endog}}}{f(confirmed\ cases)}\right)_{ref}} \frac{\frac{D_{stool-endog}}{D_{stool-SARS} F_{shed-SARS}}}{\frac{D_{stool-endog}}{D_{stool-SARS} F_{shed-SARS}}} \quad \text{Eq A5.8}$$

$$RAR3 = \frac{\frac{\frac{D_{WW-SARS}}{D_{WW-endog}}}{f(confirmed\ cases)}}{\left(\frac{\frac{D_{WW-SARS}}{D_{WW-endog}}}{f(confirmed\ cases)}\right)_{ref}} \quad \text{Eq A5.9}$$

Any point on the scatter plot of SARS-CoV-2 levels (normalized or not, y-axis) versus cases (could be new cases or sum of recent cases, x-axis) can be thought of as proportional to an ascertainment ratio (where the proportionality constant is equal to the omitted model parameters) or relative ascertainment ratio (where the proportionality constant is the reference AR). The three model formulations and their associated RAR metrics correspond to the three most common approaches to representing SARS-CoV-2 concentrations: unaltered, normalized by flow and population, and normalized by an endogenous control. Each of these three approaches has potential benefits depending on the relative importance of different sources of SARS-CoV-2 concentration variation, as follows.

Three levels of variation might be expected to contribute an error in estimating relationships between wastewater levels and true prevalence: temporal variation in wastewater composition within a site;

variation in wastewater composition across sampling locations (sewersheds); and variation across laboratories and laboratory methods (this may include multiple methods within the same laboratory as well as the same method used by different laboratories). Flow-based normalization would be expected to help adjust for longitudinal variation and cross-sectional sampling location variation, though not laboratory variation, and not non-decay losses in the sewer system. Fecal-strength-based normalization would be expected to account for all forms of variation, assuming that shedding of the endogenous control (e.g., PMMoV) is, on average, identical across sewersheds and over time, and that losses and lab processing of SARS-CoV-2 and PMMoV are, on average, identical. In practice, systematic measurement bias in either flow or PMMoV could make one of these normalization approaches more accurate, and measurement uncertainty in flow or PMMoV measurements could make one of these approaches more precise. Data availability would also contribute to model selection.

A.6 Formulation Comparison

Table A1 Comparison of 3 model formulation parameters

	Population-level parameters				Sewershed-level parameters					Sample-level parameters		
Formulation	$D_{stool\text{-}SARS}$	$D_{stool\text{-}PMMoV}$	V_{stool}	$F_{shed\text{-}SARS}$	V_{WW}	K_{SARS}	T_{sewer}	F_{cont}	N	V_{WWTP}	$D_{WW\text{-}PMMoV}$	$D_{WW\text{-}SARS}$
A.1	X		X	X	X	X	X	X				X
A.2	X		X	X		X	X		X	X		X
A.3	X	X	X	X		X	X				X	X

SI B. SARS-CoV-2 Parameter References and Technical Justification

We conducted a review of available literature to parameterize the model formulations described in SI A. To use these data to populate the model formulations for SARS-CoV-2, preferentially we used SARS-CoV-2 specific data from the scientific literature, then, other environmental data from the literature as surrogate information for SARS-CoV-2, and finally, professional judgement as needed to address information gaps where SARS-CoV-2 or surrogate data were not available. Following is a summary of the data that were collected for this purpose.

B.1 Concentrations of SARS-CoV-2 in Feces (with and without diarrhea) (see SI Section C): (D_{diar})

1. Density of SARS-CoV-2 in feces from asymptomatic cases is likely similar to those from symptomatic cases (Lin et al. 2020).
2. Cases with diarrhea had reported mean density of $10^{5.1}$-with interquartile range [$10^{4.8}$–$10^{5.6}$] copies/mL (D_{diar}) (Cheung et al. 2020). We used these data and generated a Normal distribution with mean for log10 transformed density as reported and standard deviation derived from the reported IQR.
3. Cases without diarrhea are characterized in detail in SI C. The resultant distribution is a Normal distribution with log10 transformed density mean 4.067 and standard deviation 1.591 (D_{nodiar}) (Han et al. 2020; Wolfel et al. 2020).

B.2 Volume of diarrhea per day: (V_{diar})

Pan et al. (Pan et al. 2020) conducted a cross-sectional, multicenter study which enrolled confirmed patients with COVID-19 who presented to 3 hospitals from January 18, 2020, to February 28, 2020. All patients were confirmed by real-time polymerase chain reaction and were analyzed for clinical characteristics, laboratory data, and treatment. The study found that 103 patients (50.5%) reported a digestive symptom, including lack of appetite (81 [78.6%] cases), diarrhea (35 [34%] cases), vomiting (4 [3.9%] cases), and abdominal pain (2 [1.9%] cases). Patients with digestive symptoms had a significantly longer time from onset to admission than patients without digestive symptoms (9.0 days vs 7.3 days). In 6 cases, there were digestive symptoms, but no respiratory symptoms. As the severity of the disease increased, digestive symptoms became more pronounced. Cases of diarrhea were usually not high volume or clinically severe, but more commonly presented as nondehydrating loose stools, typically up to thrice daily. Based on these findings, we used best professional judgement and assumed the 1-liter/day varies according to a normal distribution with standard deviation of 100mL (BPJ).

B.3 Diarrhea prevalence in cases: (F_{diar})

1. A meta-analysis from China that includes diarrhea prevalence showed that diarrhea prevalence was higher in severe cases as compared to non-severe Covid-19 illness (Ji et al. 2020).
2. Guo et al. (2021) summarize reported findings on the GI manifestation of COVID-19. Diarrhea was the most common GI symptom from COVID-19 with 2–50% of patients with COVID-19 reported to have diarrhea. Most commonly reported range was 20–30%.
3. The above range from Guo et al. is consistent with SARS observations (Leung et al. 2003)

B.4 Feces generation per capita (non-diarrhea conditions) (M_{nodiar})

Rose (2015) report 126 g/person day median value for high income countries, but with a skewed distribution (Refer to Table 3). Using those data, we fit a lognormal distribution to those data. The resultant distribution has ln mean and ln std dev (4.84, 0.4)

B.5 Volume of water per capita (V_{ww})

The amount of water used per person per day was based on water utility billing data for a random sample of single-family homes across a wide range water utility agencies in the United States (DeOreo et al. 2016) (Summarized in Table B.2). We used the per capita water use data in Table B.2 to fit gamma, lognormal, normal, and Weibull distributions using R. The lognormal distribution resulted in the best fit across graphical comparisons and goodness-of-fit statistics and criteria. The best fit distribution is ln-normal (5.397, 0.1595) L/day.

Table B.1 Per Capita water use summary

Agency	Per capita water use	
	Gallon per day	Liter per day
Clayton	56	212
Denver	64	242
Fort	59	223
Peel	59	223
San	69	261
Scottsdale	164	621
Tacoma	59	223
Toho	60	227
Waterloo	43	163
Aurora	68	257
Austin	45	170
Cary	52	197
Chicago	71	269
EPCOR	54	204
Henderson	76	288
Miami-Dade	80	303
Mountain	50	189
Otay	59	223
Philadelphia	65	246
Portland	49	185
RWA	60	227
Santa Barbara	53	201
Santa Fe	49	185

*Scottsdale, AZ, at 164 gallons per capita per day (621 liters per capita per day), is an outlier in this data set and removed.

B.6 Persistence in sewage of SARS-CoV-2 (k_{virus}) and the endogenous control PMMoV (k_{PMMoV})

SARS-CoV-2:

1. Ahmed et al. (2020) report a 90% reduction of SARS-CoV-2 RNA in wastewater from 8 to 27 days.
2. Wang et al. (2005) report that SARS-CoV-1 can survive for 14 days in sewage at 4 degrees C, and 2 days at 20 degrees C. They also report that its RNA can be detected for 8 days though the virus had been inactivated.
3. Bivins et al. (2020) report 90% reduction of viable SARS-CoV-2 in wastewater 1.5 days. SARS-CoV-2 RNA was found to be significantly more persistent than infectious SARS-CoV-2.

Based on these data, we derived a decay coefficient for SARS-CoV-2, following the Chick's law derivation approach published by Boehm et al. (2019) = 0.29 day^{-1}. The shortest reported persistence times were used for conservative decay estimates.

PMMoV:

1. Recent research (Graham et al. 2021; Wolfe et al. 2021) indicate that specific data do not presently exist on PMMoV RNA decay in wastewater (Rosario et al. 2009; Symonds et al. 2018).
2. Evidence to date indicates that short-read SARS-CoV-2 RNA fragments are relatively stable in raw wastewater under temperature regimes common in the US. T_{90} for short read SARS-CoV-2 RNA targets have been measured to be between 8 days (at 37°C) and 28 days (at 4° C) (Ahmed et al. 2020). Decay rates for short-read PMMoV RNA targets in unaltered wastewater have not been reported but may be slower (more stable) based on experiments in other water matrices. For example, PMMoV RNA is generally more stable through wastewater treatment processes than human enteric viruses (Symonds et al. 2018), and has been found to have similar stability as SARS-CoV-2 RNA fragments in 4° C water and greater stability than SARS-CoV-2 RNA at 20C in fresh and seawater matrices (Sala-Comorera et al. 2021).

Without firm evidence on the decay coefficient for PMMoV in wastewater, we therefore assume that it is the same as for SARS-CoV-2[1].

B.7 Residence time of wastewater in sewage in USA: (T_{sewer})

Kapo et al. (2017) report detailed estimates of residence times of wastewater in sewage systems in the US. In that report, a summary of salient results was provided for treatment facilities of various sizes. We used the "All facilities" results and a lognormal distribution was fit numerically to the reported median and 90th percentile values. This results in a lognormal distribution with mean and std dev = (1.2, 0.85) (units of ln hr.)

B.8 Shedding prevalence in cases: (F_{shed})

1. Two independent laboratories from China successfully isolated live 2019-nCoV from the stool of patients (Gu et al. 2020).
2. Compared to pharyngeal swab specimens, nucleic acid detection of SARS-CoV-2 in fecal specimens was equally accurate (J Zhang et al. 2020).
3. Nine out of the 19 patients were detected 2019-nCoV infection using oropharyngeal swab samples, and viral nucleic acid was also detected in 8 of 9 patients using stool samples (Xie et al. 2020). No positive results were identified in blood or urine samples. In the nine confirmed patients, eight stool samples showed positive results for 2019-nCoV; interestingly, the virus could still be detected in stool samples from 89% of patients without diarrhea symptoms.
4. Zheng et al. (2020) reported that viral RNA was detected in the stool of 60% (55 of 93) patients.

[1] For wastewater transit, a reasonable estimate for transit time is 1 day. Therefore, the max expected factor this would contribute to our equations is ~ 0.114 per day. If k_{pmmov} at its extreme is 0 (no decay), then this would result in a factor (exp(-k_{sars})) of 0.89. So at its most conservative, would alter results by 10%, but it's likely not 0 and less than 10%.

5. Cheung et al. (2020) reported meta-analysis results indicating that the prevalence of stool samples that were positive for virus RNA was 48%.
6. Zhang et al. (2020) found the presence of 2019-nCoV in anal swabs and blood and more anal swab positives than oral swab positives in a later stage of infection, suggesting shedding and thereby transmitted through oral–fecal route
7. Regarding the shedding prevalence in asymptomatic cases: Tang et al. (Tang et al. 2020) reported that up to 75% of infections could be asymptomatic. Stool was positive 17 days after last virus exposure in asymptomatic child and positive for at least an additional 9 days.

Taken all together, we infer that shedding occurs in cases with and w/o diarrhea, in a range of approximately 60-80% of infections.

B.9 Volume fraction of wastewater coming from a potentially infectious source (F_{cont})

Our review of the literature review provided little evidence to support a widely applicable value for this parameter. Given that, we used best professional judgement to estimate a feasible range. This parameter could and should be modified for site specific purposes in the future.

B.10 Density of PMMoV in stool (D_{WW_endog})

Hamza (2011) tested 20 human stool samples and report concentration per mg. We used a uniform distribution based on data in Table 3 of this paper. This resulted in minimum and maximum values of 5.58 and 9.99 (Median = 8.28) log10 copies after conversion to mass (per g) basis.

SI C. Fecal Shedding Distribution Justification

We identified three studies that published quantitative concentrations of SARS-CoV-2 in stool (Han et al. 2020; Wolfel et al. 2020; Zheng et al. 2020). Wolfel and Zheng present quantitative concentrations for individuals on a per amount of stool basis (correspondence with Zheng). Results from Han could not be used to describe concentrations in stool because these measurements were made on a per volume of extract basis (correspondence with author).

Following is a summary of the data from those three studies:

1. The Wolfel et al. data come from 9 patients with mild disease courses in Germany, all of whom were young to middle-aged with no notable underlying conditions. 1 of 9 patients experienced intermittent diarrhea. Two non-detects are reported, but not included in these data. Data are digitized from Figure 1c.
2. The Zheng et al. data are digitized from Figure 2. Each data point represents one sample from an individual. It is not clear how many samples come from any given individual, or how many individuals are represented here. Although the paper reports detecting virus in stool of 56 patients, only 49 data points are visible in this figure. In summary, patients were in China, 22 with mild disease and 74 with severe disease. The median age was 55 years (interquartile range 44.3–64.8).
3. The Han et al. data are digitized from Figure 1B, representing 12 children. The methods are very sparsely reported in this study, so it was not possible to assess the quality of their measurements.

The inclusion criteria for this study are summarized in Table C.1 .

Table C2 Data inclusion criteria summary for SARS-CoV-2 fecal shedding

Characteristic	Paper	Wolfel	Zheng	Han
qPCR	Standard curve statistics	no	no	no
	Number qPCR replicates	no	no	no
	No template controls	no	no	no
	Inhibition testing indicated[1]	yes	no	no
	dilution or other used to reduce inhibition[1]	yes	n/a	n/a
	Assay name/reference is provided	yes	yes	yes
	Quantitative standard material described[1]	No; assumed same as European assay dev paper by Corman, who is 2nd author	No; assumed contained in government kit	yes, in-vitro transcribed RNA; kit was validated by manufacturer for quantitation using in vitro transcribed RNA; but FDA approval only for qualitative detection; kit provides positive control and internal extraction control
	Detection limit provided	no	yes	yes; figure caption
	Explanation/reference for detection limit	n/a	yes	no
	Quantification limit provided	yes; figure 2 caption	no	no
	Explanation/reference for quantification limit	no	n/a	n/a
RNA extraction	Extraction method indicated	Qiagen RNA mini kit	Roche MagNA Pure	SeeGene Allplex 2019-nCoV Assay kit
Stool collection/processing	Explanation of how measurement performed for concentration denominator (g or mL)[1]	no	no	no
	Description of stool sample acquisition/processing	yes	yes	no
	Description of stool testing frequency strategy[1]	no	yes	no
	Number of processing replicates	2	1	1
Results	Concentration in individual stool samples given[2]	yes	yes	yes
	Clear that concentrations are on a per stool basis rather than a per volume PCR template basis?[2]	Yes. "Viral loads were projected to RNA copies... per g (for stool samples)" while per ml for other samples, suggesting that concentration is per g stool. Furthermore, "stool samples were taken and shipped in native conditions" indicating that mass measurement of stool was possible in the lab.	Yes, per email correspondence described in Comments	No, per email correspondence described in comments
	Number of individuals represented in data is clear[1]	yes	described in text, although numbers slightly mismatch with digitized data; SI figures given further clarity on when which patients sampled	yes
	Figure number	2	2, top left	1B
Patient description	Explanation of patient selection/representativeness[1]	Patients selected as close contacts of initial index case to avoid bias by selecting patients based on symptoms	96 consecutively admitted hospital patients with confirmed SARS-CoV-2 infection	All positive <18 children admitted to hospital over date range
	Explanation of patient age[1]	yes	yes	yes
	Explanation of case severity[1]	yes	yes	yes
	Clear indication of days since onset corresponding to stool concentrations[1]	yes	no	yes
Other	Other description of quality control/method rigor	RT-qPCR analyses performed by 2 independent labs. Close congruence of results led to presenting data from both labs together - average presented in figure 2. "Experiments performed in duplicate" - based on caption of figure 1, this likely indicates 2 subsamples from same clinical sample. Second author, Corman, is the first author of the primary qPCR assay used in Europe.	Units and LOD for all sample types (sputum, saliva, stool, blood) are the same: log10 copies per ml. This suggests that these concentration results are likekly per mL of template in the reaction or extract, not per mL of sample, because otherwise it implies that the mass of each sample type exctracted was the same (mass stool = mass saliva, for example), and that the volume of saliva and sputum extracted was measured, which seems highly unlikely.	Units and LOD for all sample types (nasopharangeal, stool, saliva) are the same. This suggests that these concentration results may be per mL of template in the reaction, not per mL of sample, because otherwise it implies that the mass of each sample type exctracted was the same (mass stool = mass saliva, for example)
	Comments		Contacted author by email and learned that units are per ml of stool: "Dear Professor Jennings. I am the co-author of "Viral Load Dynamics and Disease Severity in patients infected with SARS-COV-2 in Zhejiang Province, China, January-March 2020: Retrospective Cohort Study ": Thank you for your recognition of our work. In this paper, for the stool measurements, are these per mL of stool. Best wishes! Shufa Zheng"	Contacted author by email and learned that concentrations are per ml extract not ml feces: "1.In Figure 1, the units of all three sample types (nasopharyngeal, feces, and saliva) are log10 copies/mL. For the feces measurements, are these per mL of feces, or per mL of RNA extract (or some other volume)? We used per mL of RNA extract 1.In Figure 1B, if these concentrations are per mL of feces, how was the volume of feces measured? We used per mL of RNA extract. We made feces suspension with stool-coated flocked swab and 3 mL of UTM media (Copan). RNA was extracted from the stool suspension. "
Conclusion	Can be included in quantitative assessment of SARS-CoV 2 RNA concentration per known amount of feces?	Yes	Yes	No

[1] Important for using data as quantitative measurement of viral RNA copies per amount feces

[2] Essential for using data as quantitative measurement of viral RNA copies per amount feces

Following are summaries of the reported fecal shedding studies and development of a shedding distribution based on evaluated data.

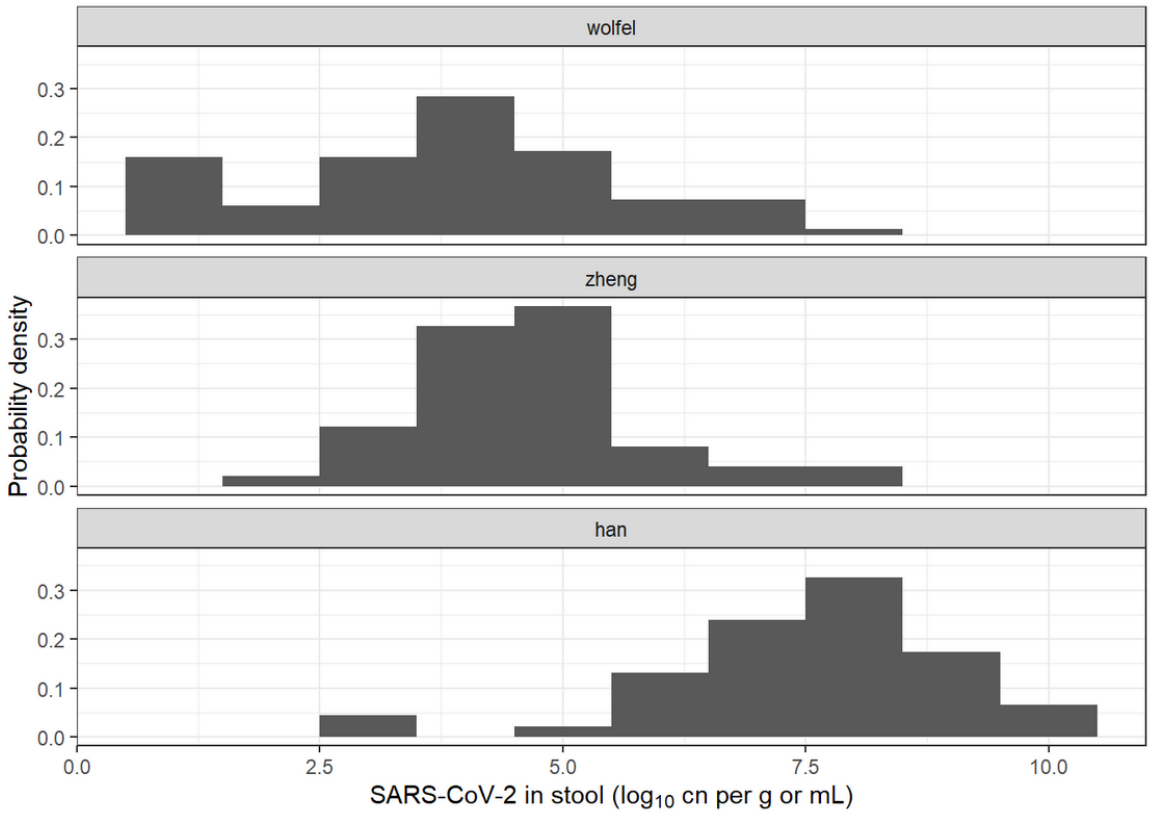

Figure C1 Histogram of SARS-CoV-2 fecal shedding study results demonstrating the distribution uncertainty

Wolfel data summary. Statistical data summaries for the Wolfel data are presented in Figure C2. The Lognormal distribution has the lowest AIC, so it may be an appropriate choice, although the normal and Weibull distributions fit better in the upper tail (see Q-Q plot).

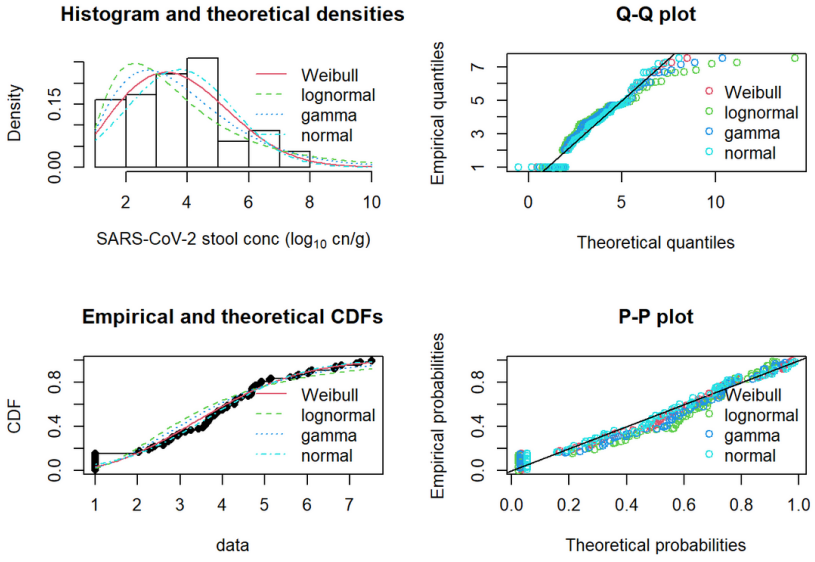

Figure C2. Wolfel data summary

Wolfel and Zheng aggregated data summary. Statistical data summaries for the combined Wolfel and Zheng dataset are presented in Figure C3. The Normal and Weibull distributions have the lowest AIC, and, as shown in the Q-Q plot, the normal distribution fits the data in the upper tail best. This combined dataset is used in the analysis and is characterized by the corresponding normal distribution.

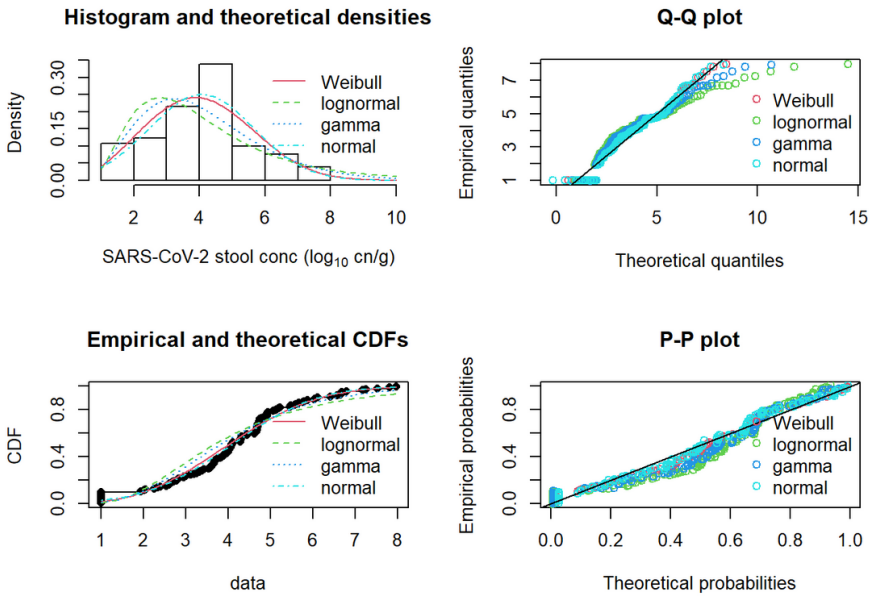

Figure C3. Wolfel and Zheng aggregated data summary

We also bootstrapped the fit of the aggregated Wolfel and Zheng data to the data using the normal distribution for the purpose of performing sensitivity analysis. To understand the uncertainty in the fit of the distribution, the Wolfel and Zheng stool data were resampled 999 times with replacement, and a normal distribution was fit to each resample. Then, the probability density of each distribution was computed for a stool concentration of 10^8 copies/g stool, and fits were ranked according to these densities. The 5th, 25th, 50th, 75th, and 95th percentile distributions are plotted below.

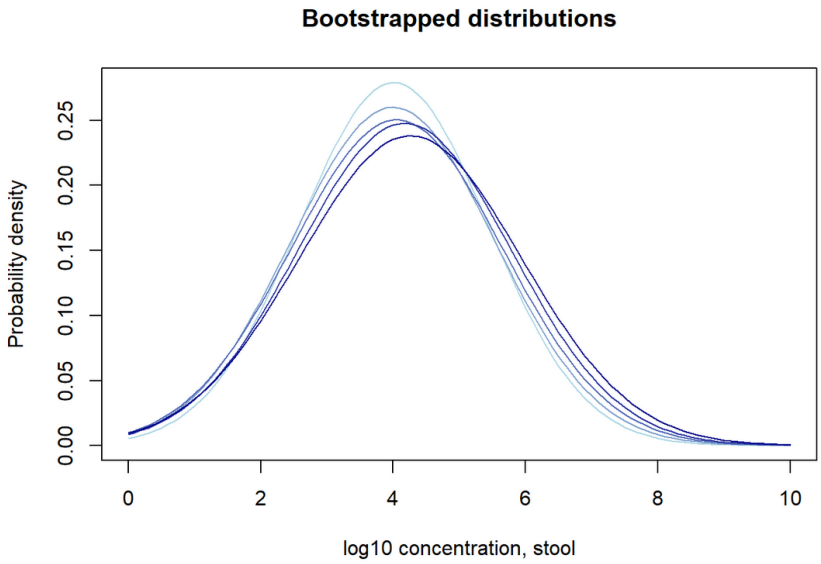

Figure C4. Wolfel and Zheng aggregated data bootstrapped distribution fits.

SI D. Evaluation of Model Output Variability/Uncertainty and Related Model Output Sensitivity to Partitioning Coefficients, SARS-CoV-2 Shedding Duration, and Model Formulations

Figure D1 shows the 10th percentile, median, and 90th percentile values for the modeling results of F_{active} for the fecal strength model formulation evaluating the Virginia Beach, VA case study.

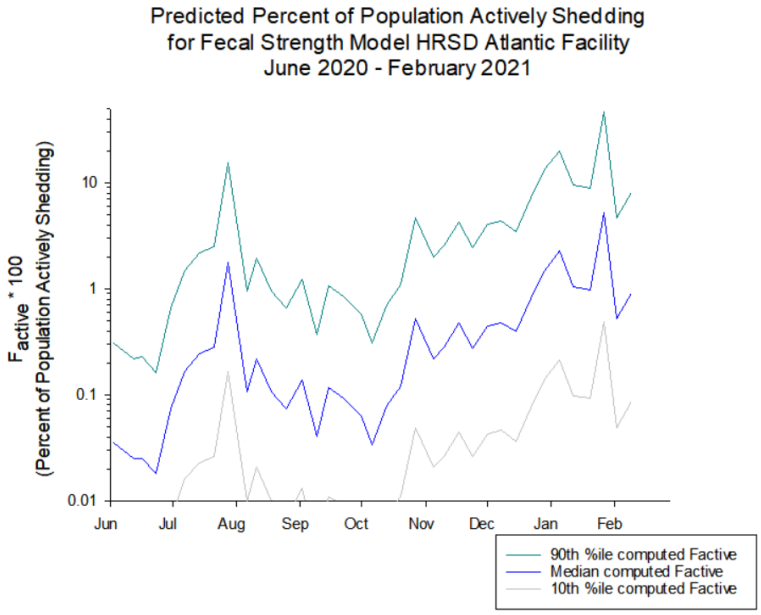

Figure D1. Percentile values for fecal strength formulation simulation results for Virginia Beach case study.

Figure D2 shows how the computed influent concentration for the San Jose, CA case study varies depending on the selection of dispersion coefficient for PMMoV and SARS-CoV-2.

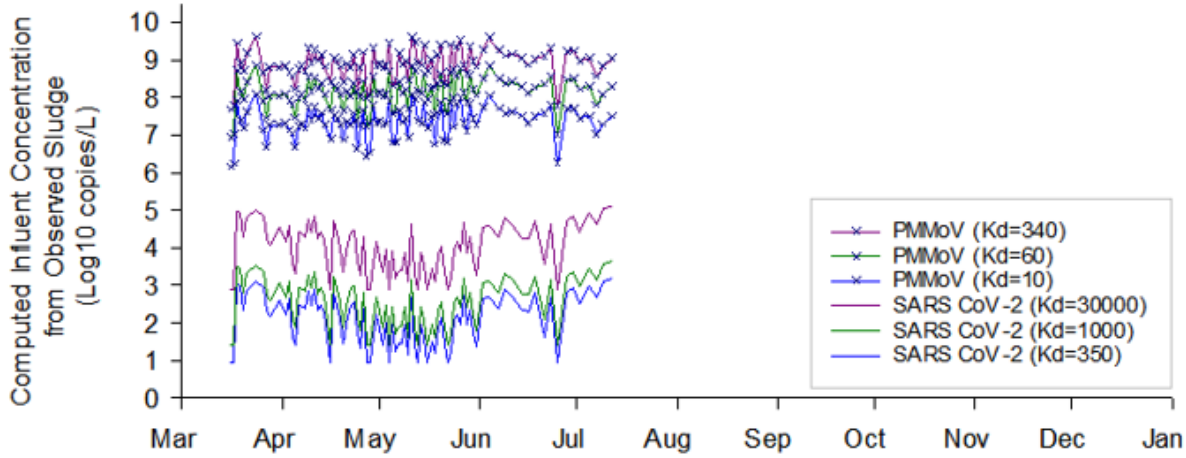

Figure D2. San Jose, CA estimated SARS-CoV-2 and PMMoV influent concentrations as a function of k_d values.

Figure D3 shows relative ascertainment ratio (RAR) values for 9 sewersheds in Hampton Roads Sanitation District (HRSD) calculated using the fecal strength model formulation. Here we evaluated RAR as a function of the number of days over which COVID-19 cases are summed. These figures show the sensitivity of RAR to three different case-summation time spans: 1, 7, and 21 days. For example, 7 days is the summation of cases over 7 total days prior to and including the wastewater sampling date. These figures indicate that RAR is generally not sensitive to the selection of the case-summation time span between 1 and 21 days.

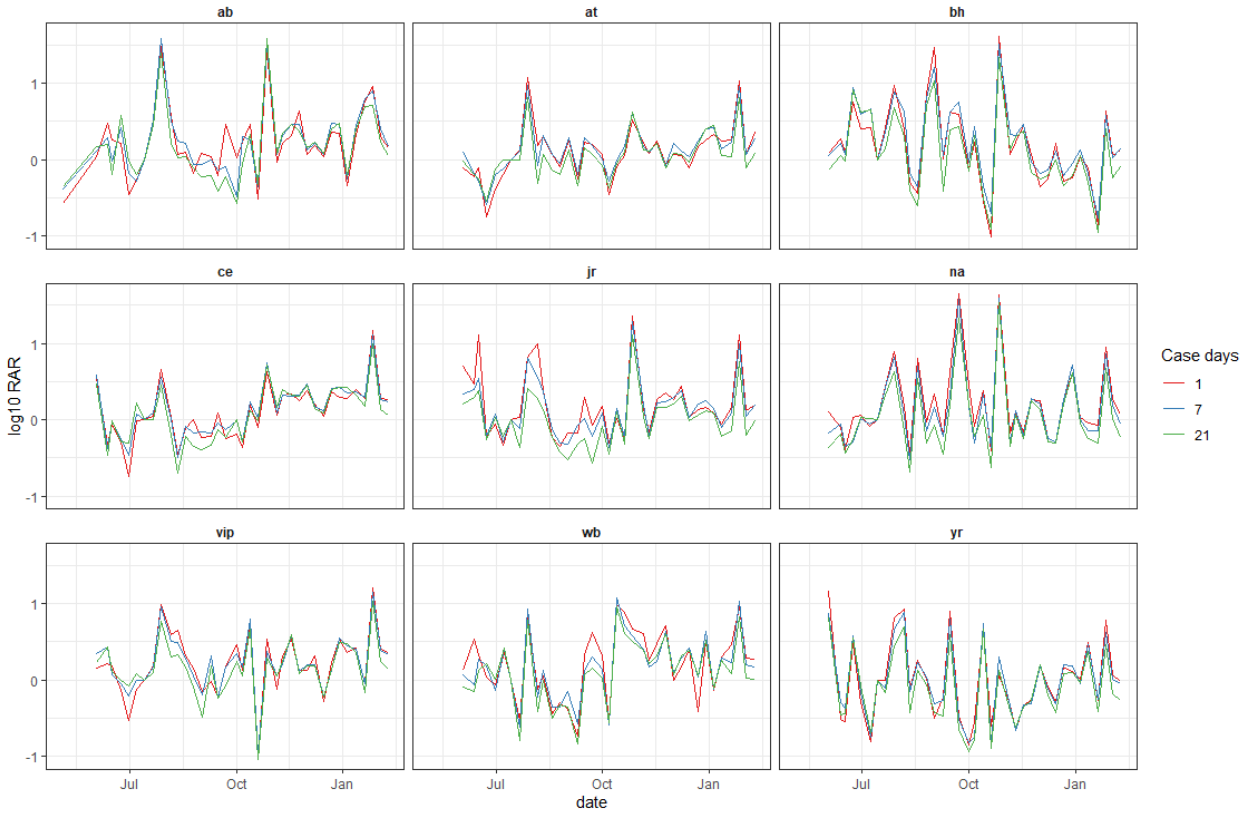

Figure D3. RAR as a function of the number of days over which COVID-19 cases are summed.

Figure D4 shows RAR values for 9 sewersheds in the HRSD service area calculated using the three model formulations described in the main body text. The case period evaluated in these figures is 21 days. These figures show the sensitivity of RAR to three different model formulations. These figures indicate that RAR is generally somewhat sensitive to the selection of model formulation selection. The model formulations are labels as: 1) flow receipt (red), 2) flow generation (blue), and 3) fecal strength (green).

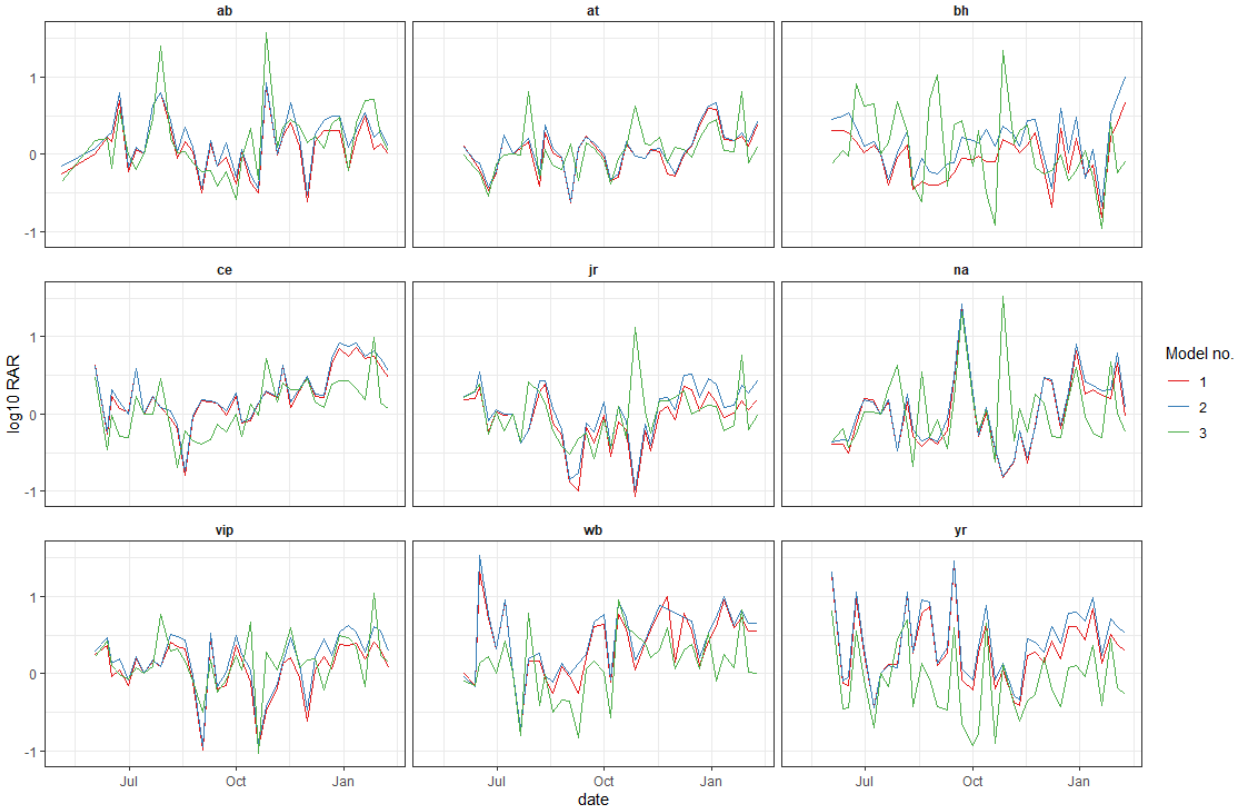

Figure D4. RAR as a function of model formulation.

Supplemental Information References

Ahmed W, Bertsch PM, Bibby K, Haramoto E, Hewitt J, Huygens F, Gyawali P, Korajkic A, Riddell S, Sherchan SP, Simpson SL, Sirikanchana K, Symonds EM, Verhagen R, Vasan SS, Kitajima M, Bivins A. 2020. Decay of sars-cov-2 and surrogate murine hepatitis virus rna in untreated wastewater to inform application in wastewater-based epidemiology. Environmental research 191:110092.

Bivins A, North D, Ahmad A, Ahmed W, Alm E, Been F, Bhattacharya P, Bijlsma L, Boehm AB, Brown J, Buttiglieri G, Calabro V, Carducci A, Castiglioni S, Cetecioglu Gurol Z, Chakraborty S, Costa F, Curcio S, de Los Reyes FL, 3rd, Delgado Vela J, Farkas K, Fernandez-Casi X, Gerba C, Gerrity D, Girones R, Gonzalez R, Haramoto E, Harris A, Holden PA, Islam MT, Jones DL, Kasprzyk-Hordern B, Kitajima M, Kotlarz N, Kumar M, Kuroda K, La Rosa G, Malpei F, Mautus M, McLellan SL, Medema G, Meschke JS, Mueller J, Newton RJ, Nilsson D, Noble RT, van Nuijs A, Peccia J, Perkins TA, Pickering AJ, Rose J, Sanchez G, Smith A, Stadler L, Stauber C, Thomas K, van der Voorn T, Wigginton K, Zhu K, Bibby K. 2020. Wastewater-based epidemiology: Global collaborative to maximize contributions in the fight against covid-19. Environ Sci Technol 54:7754-7757.

Boehm AB, Silverman AI, Schriewer A, Goodwin K. 2019. Systematic review and meta-analysis of decay rates of waterborne mammalian viruses and coliphages in surface waters. Water Res 164:114898.

Cheung KS, Hung IF, Chan PP, Lung KC, Tso E, Liu R, Ng YY, Chu MY, Chung TW, Tam AR, Yip CC, Leung KH, Yim-Fong Fung A, Zhang RR, Lin Y, Cheng HM, Zhang AJ, To KK, Chan KH, Yuen KY, Leung WK. 2020. Gastrointestinal manifestations of sars-cov-2 infection and virus load in fecal samples from the hong kong cohort and systematic review and meta-analysis. Gastroenterology.

DeOreo WB, Mayer PW, Dziegielewski B, Kiefer J. 2016. Residential end uses of water, version 2. Denver, CO:Water Research Foundation.

Graham KE, Loeb SK, Wolfe MK, Catoe D, Sinnott-Armstrong N, Kim S, Yamahara KM, Sassoubre LM, Mendoza Grijalva LM, Roldan-Hernandez L, Langenfeld K, Wigginton KR, Boehm AB. 2021. Sars-cov-2 rna in wastewater settled solids is associated with covid-19 cases in a large urban sewershed. Environ Sci Technol 55:488-498.

Gu J, Han B, Wang J. 2020. Covid-19: Gastrointestinal manifestations and potential fecal-oral transmission. Gastroenterology.

Guo M, Tao W, Flavell RA, Zhu S. 2021. Potential intestinal infection and faecal-oral transmission of sars-cov-2. Nature reviews Gastroenterology & hepatology 18:269-283.

Hamza IA, Jurzik L, Uberla K, Wilhelm M. 2011. Evaluation of pepper mild mottle virus, human picobirnavirus and torque teno virus as indicators of fecal contamination in river water. Water Res 45:1358-1368.

Han MS, Seong MW, Kim N, Shin S, Cho SI, Park H, Kim TS, Park SS, Choi EH. 2020. Viral rna load in mildly symptomatic and asymptomatic children with covid-19, seoul, south korea. Emerg Infect Dis 26:2497-2499.

Ji W, Zhang J, Bishnu G, Du X, Chen X, Xu H, Guo X, Cai Z, Shen X. 2020. Comparison of severe and non-severe covid-19 pneumonia: Review and meta-analysis. medRxiv:2020.2003.2004.20030965.

Kapo KE, Paschka M, Vamshi R, Sebasky M, McDonough K. 2017. Estimation of u.S. Sewer residence time distributions for national-scale risk assessment of down-the-drain chemicals. Sci Total Environ 603-604:445-452.

Leung WK, To KF, Chan PK, Chan HL, Wu AK, Lee N, Yuen KY, Sung JJ. 2003. Enteric involvement of severe acute respiratory syndrome-associated coronavirus infection. Gastroenterology 125:1011-1017.

Lin L, Jiang X, Zhang Z, Huang S, Zhang Z, Fang Z, Gu Z, Gao L, Shi H, Mai L, Liu Y, Lin X, Lai R, Yan Z, Li X, Shan H. 2020. Gastrointestinal symptoms of 95 cases with sars-cov-2 infection. Gut 69:997-1001.

Pan L, M Mu, HG Ren, P Yang, YSun, R Wang, J Yan, P Li, B Hu, Chaohu, Y Jin, X Niu, R Ping, Y Du, T Li, C Liu, G Xu, Q Hu, Tu L. 2020. Clinical characteristics of covid-19 patients with digestive symptoms in hubei, china: A descriptive, cross-sectional, multicenter study. Am J Gastroenterol.

Penn R, Ward BJ, Strande L, Maurer M. 2018. Review of synthetic human faeces and faecal sludge for sanitation and wastewater research. Water Res 132:222-240.

Rosario K, Symonds EM, Sinigalliano C, Stewart J, Breitbart M. 2009. Pepper mild mottle virus as an indicator of fecal pollution. Appl Environ Microbiol 75:7261-7267.

Rose C, Parker A, Jefferson B, Cartmell E. 2015. The characterization of feces and urine: A review of the literature to inform advanced treatment technology. Crit Rev Environ Sci Technol 45:1827-1879.

Sala-Comorera L, Reynolds LJ, Martin NA, O'Sullivan JJ, Meijer WG, Fletcher NF. 2021. Decay of infectious sars-cov-2 and surrogates in aquatic environments. Water Research:117090.

Symonds EM, Nguyen KH, Harwood VJ, Breitbart M. 2018. Pepper mild mottle virus: A plant pathogen with a greater purpose in (waste)water treatment development and public health management. Water Res 144:1-12.

Tang A, Tong ZD, Wang HL, Dai YX, Li KF, Liu JN, Wu WJ, Yuan C, Yu ML, Li P, Yan JB. 2020. Detection of novel coronavirus by rt-pcr in stool specimen from asymptomatic child, china. Emerg Infect Dis 26.

Wang XW, Li J, Guo T, Zhen B, Kong Q, Yi B, Li Z, Song N, Jin M, Xiao W, Zhu X, Gu C, Yin J, Wei W, Yao W, Liu C, Li J, Ou G, Wang M, Fang T, Wang G, Qiu Y, Wu H, Chao F, Li J. 2005. Concentration and detection of sars coronavirus in sewage from xiao tang shan hospital and the 309th hospital of the chinese people's liberation army. Water Sci Technol 52:213-221.

Wolfe MK, Archana A, Catoe D, Coffman MM, Dorevich S, Graham KE, Kim S, Grijalva LM, Roldan-Hernandez L, Silverman AI, Sinnott-Armstrong N, Vugia DJ, Yu AT, Zambrana W, Wigginton KR, Boehm AB. 2021. Scaling of sars-cov-2 rna in settled solids from multiple wastewater treatment plants to compare incidence rates of laboratory-confirmed covid-19 in their sewersheds. Environmental Science & Technology Letters 8:398-404.

Wolfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Muller MA, Niemeyer D, Jones TC, Vollmar P, Rothe C, Hoelscher M, Bleicker T, Brunink S, Schneider J, Ehmann R, Zwirglmaier K, Drosten C, Wendtner C. 2020. Virological assessment of hospitalized patients with covid-2019. Nature 581:465-469.

Xie C, Jiang L, Huang G, Pu H, Gong B, Lin H, Ma S, Chen X, Long B, Si G, Yu H, Jiang L, Yang X, Shi Y, Yang Z. 2020. Comparison of different samples for 2019 novel coronavirus detection by nucleic acid amplification tests. International Journal of Infectious Diseases 93:264-267.

Zhang J, Wang S, Xue Y. 2020. Fecal specimen diagnosis 2019 novel coronavirus–infected pneumonia. Journal of Medical Virology 92:680-682.

Zhang W, Du R-H, Li B, Zheng X-S, Yang X-L, Hu B, Wang Y-Y, Xiao G-F, Yan B, Shi Z-L, Zhou P. 2020. Molecular and serological investigation of 2019-ncov infected patients: Implication of multiple shedding routes. Emerging Microbes & Infections 9:386-389.

Zheng S, Fan J, Yu F, Feng B, Lou B, Zou Q, Xie G, Lin S, Wang R, Yang X, Chen W, Wang Q, Zhang D, Liu Y, Gong R, Ma Z, Lu S, Xiao Y, Gu Y, Zhang J, Yao H, Xu K, Lu X, Wei G, Zhou J, Fang Q, Cai H, Qiu Y, Sheng J, Chen Y, Liang T. 2020. Viral load dynamics and disease severity in patients infected with sars-cov-2 in zhejiang province, china, january-march 2020: Retrospective cohort study. BMJ 369:m1443.