Model Documentation for the R Library wem (**w**astewater-based **e**pidemic **m**odel)

David Champredon[1], Shokoofeh Nourbakhsh[1]

Affilitations:

[1] Public Health Risk Sciences Division, National Microbiology Laboratory, Public Health Agency of Canada, Guelph, ON, Canada

Correspondance:
david.champredon@canada.ca
shokoofeh.nourbakhsh@canada.ca

The R package wem is available for download at:
https://github.com/phac-nml-phrsd/wem.

Documentation date: 2021-12-02 15:57:47-05:00

1 Methods

The R library `wem` implements a mathematical model that mechanistically describes both the pathogen transmission at the population level ("above ground") and the pathogen concentration measured in wastewater as a result of faecal shedding/urinary from the infected individuals ("below ground"). This model was developed with SARS-CoV-2 in mind but could be applicable to many other pathogens where faecal/urinary shedding has been observed.

1.1 Transmission between individuals

To model patogen transmission in the population, we use a SEIR-type epidemiological model. The disease progression of individuals is captured through several compartments that reflect their epidemiological states and disease outcomes (Table 2). Individuals can be susceptible (S); exposed (infected but not yet infectious, E); symptomatically infected who will later become hospitalized (J) or recovered without hospitalization during active COVID-19 (I); asymptomatically infected (A); hospitalized (H); those recovered and no longer infectious but still shedding virus in faeces (Z); fully recovered and permanently immune but not shedding anymore (R) and deceased (D). We ignore any migration movements, so at any given time the total population is constant and equal to $N = S + E + J + I + A + H + Z + R + D$ without vaccination. Infection occurs at a time-dependent transmission rate β_t between infectious (states I, J or A) and susceptible individuals (S). Once infected, susceptible individuals enter the latent (non-infectious) state (E) for an average duration of $1/\epsilon$ days, where no faecal shedding occurs. A proportion α of all infections are asymptomatic. A fraction h of symptomatic individuals are hospitalized (H) for an average duration of $1/\ell$ days and for those, the pathogen-associatied mortality is δ. After their infectious period ends, patients enter the post-infection shedding state Z where faecal shedding still occurs for $1/\eta$ days on average. The exposed (E), infectious (A, I and J) and post-infection shedding (Z) states are modelled with a series of sub-compartments in order to have their respective sojourn time gamma-distributed. Note that in our model, we make the simplifying assumption that hospitalized patients–assumed mostly bedridden and a small fraction of the shedding population–do not contribute significantly to faecal shedding.

Vaccination against the pathogen is also modelled. We assume that only susceptible individuals can be vaccinated. Upon administration of the vaccine, individuals are not immediately protected against infection and are in the state V_{wa}, waiting (for a period of time on average equal to $1/d$) for their immunity to build up to a protective level. After this transitional state, individuals are considered fully vaccinated (V). If the vaccine is not fully protective against infection (*i.e.*, its efficacy against infection eff_{inf} is smaller than 1), individuals can be infected and move to the exposed-and-vaccinated state (E^V). Immunity waning is taken into account for both vaccinated and naturally infected individuals, with

a mean duration of immunity equal to $1/\tau$.

More details about vaccination modelling is provided in section **??**.

The transmission dynamics are represented by the system of differential equations 1a-1p and illustrated in Figure 1.

$$\dot{S} = \tau R + \tau V - r_t S - \beta_t S\left(A + I + J\right)/N \tag{1a}$$

$$\dot{V}_{\text{wa}} = r_t S - d V_{\text{wa}} \tag{1b}$$

$$\dot{V} = d V_{\text{wa}} - \tau V - (1 - \text{eff}_{\text{inf}})\,\beta_t V\left(A + I + J\right)/N \tag{1c}$$

$$\dot{E} = \beta_t S\left(A + I + J\right)/N - \varepsilon E \tag{1d}$$

$$\dot{E^V} = (1 - \text{eff}_{\text{inf}})\,\beta_t V\left(A + I + J\right)/N - \varepsilon E^V \tag{1e}$$

$$\dot{A}_1 = \alpha\varepsilon E + \alpha'\varepsilon E^V - n_{_A}\theta A_1 \tag{1f}$$

$$\dot{A}_k = n_{_A}\theta(A_{k-1} - A_k) \qquad 2 \leq k \leq n_{_A} \tag{1g}$$

$$\dot{I}_1 = (1-h)\,(1-\alpha)\varepsilon E + (1-h')\,(1-\alpha')\varepsilon E^V - n_{_I}\nu I_1 \tag{1h}$$

$$\dot{I}_k = n_{_I}\nu(I_{k-1} - I_k) \qquad 2 \leq k \leq n_{_I} \tag{1i}$$

$$\dot{J}_1 = h(1-\alpha)\varepsilon E + h'(1-\alpha')\varepsilon E^V - n_{_J}\mu J_1 \tag{1j}$$

$$\dot{J}_k = n_{_J}\mu(J_{k-1} - J_k) \qquad 2 \leq k \leq n_{_J} \tag{1k}$$

$$\dot{H} = \mu_{n_{_J}} J_{n_{_J}} - \ell H \tag{1l}$$

$$\dot{Z}_1 = n_{_I}\nu I_{n_{_I}} + n_{_A}\theta A_{n_{_A}} - n_{_Z}\eta Z_1 \tag{1m}$$

$$\dot{Z}_k = n_{_Z}\eta(Z_{k-1} - Z_k) \qquad 2 \leq k \leq n_{_Z} \tag{1n}$$

$$\dot{R} = n_{_Z}\eta Z_{n_{_Z}} + (1-\delta)\ell H - \tau R \tag{1o}$$

$$\dot{D} = \delta\,\ell H \tag{1p}$$

where $A = \xi\sum_{k=1}^{n_A}\phi_k A_k$, $I = \sum_{k=1}^{n_I}\psi_k I_k$ and $J = \sum_{k=1}^{n_J}\psi_k J_k$. We use the dot notation to symbolize derivation with respect to time (*e.g.*, $\dot{S} = dS/dt$).

The parameters ϕ and ψ are multiplicative adjustments to the baseline transmission rate β_t to represent the infectious profile during the course of infection. The values for ϕ_k and ψ_k were chosen to represent the best estimate of the temporal infectiousness profile given

the different results published (see Appendix Figure S2 for an example for SARS-CoV-2). The parameter ξ models the relative infectiousness of asymptomatic cases compared to symptomatic ones. The effective reproduction number of this model is:

$$\mathcal{R}_0 = m\beta_t \frac{S_0}{N} + m'\beta_t(1-\text{eff}_{\text{inf}})\frac{V_0}{N} \tag{2}$$

with

$$\begin{aligned} m &= \alpha\xi\frac{1}{\theta} + (1-h)(1-\alpha)\frac{1}{\nu} + h(1-\alpha)\frac{1}{\mu}\frac{\sum_{k=1}^{n_J}\psi_k}{n_I} \\ m' &= \alpha'\xi\frac{1}{\theta} + (1-h')(1-\alpha')\frac{1}{\nu} + h'(1-\alpha')\frac{1}{\mu}\frac{\sum_{k=1}^{n_J}\psi_k}{n_I} \end{aligned}$$

Moreover, the effective reproduction number $\mathcal{R}_t$ is:

$$\mathcal{R}_t = \mathcal{R}_0\frac{S_t + V_t}{N}. \tag{3}$$

1.1.1 Vaccination

For the vaccination part of the model, a daily proportion r of susceptible individuals receive a vaccine dose and wait in the V_{wa} compartment for $1/d$ days until their immunity builds up to full protection. For simplicity, we do not model multi-doses vaccines. Modellers need to input a long-enough waiting time ($1/d$) to build immunity to represent multiple doses. Vaccinated individuals in the V compartment have a $(1-\text{eff}_{\text{inf}})$ risk of infection to move to exposure compartment, where eff_{inf} is the vaccine effectiveness against infection. The remaining of protected individuals loose their immunity after $1/\tau$ days and become fully susceptible again. After spending their latency period in the E^V compartment, infected individuals become infectious and depending on the relative vaccine effectiveness against symptomatic infection and hospitalization, they move to either the I, J or A compartments.

The proportion of asymptomatic (α') and hospitalized (h') for vaccinated individuals depends on selected conditional vaccine effectiveness for these events.

$$\alpha' = \text{eff}_{\text{symp/inf}} \tag{4}$$
$$h' = 1 - \text{eff}_{\text{hosp/symp}} \tag{5}$$

In order to estimate these conditional vaccine effectiveness, we need input values for vaccine efficacies from clinical trials against infection, symptomatic and hospitalization given person is exposed. We provided an example in the table below:

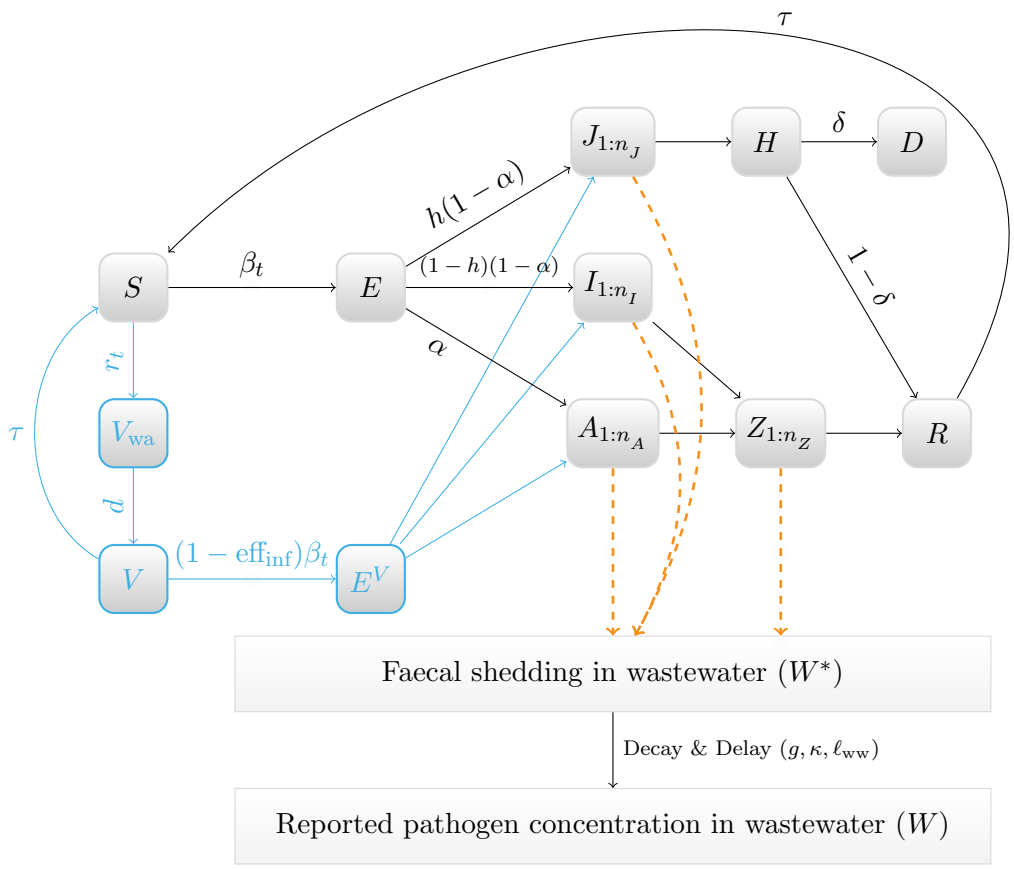

Figure 1: Diagram of compartmental model. See main text for a description of the epidemiological states. The notation $1 : n_\bullet$ indicates a modelling using $n_\bullet$ sub-compartments to obtain a gamma-distributed sojourn time in the associated epidemiological state. The blue pathway represents vaccinated individuals who get exposed and infected depending on the transmission rate and vaccine effectiveness.

Table 1: Conditional vaccine effectiveness. Example taking two different values for eff_{inf} (0.90 and 0.50) and how this is compensated by two values ($eff_{symp/inf} = 0.50$ and 0.90) for the vaccine effectiveness against symptomatic infection.

conditional vacc. effectiveness	vacc 1	vacc 2	source
$\text{eff}_{\text{inf/exp}}$	0.90	0.50	ref. clinical trials
$\text{eff}_{\text{symp/inf}}$	0.50	0.90	$1 - \frac{1-\text{eff}_{\text{symp/exp}}}{1-\text{eff}_{\text{inf/exp}}}$
$\text{eff}_{\text{symp/exp}}$	0.95	0.95	ref. clinical trials
$\text{eff}_{\text{hosp/symp}}$	0.40	0.40	$1 - \frac{1-\text{eff}_{\text{hosp/exp}}}{1-\text{eff}_{\text{symp/exp}}}$
$\text{eff}_{\text{hosp/exp}}$	0.97	0.97	ref. clinical trials

1.2 SARS-CoV-2 viral concentration in wastewater

1.2.1 Deposited Viral Concentration

The daily concentration of SARS-CoV-2 in wastewater is directly calculated from the total number of individuals that are actively shedding into the sewage system. Faecal shedding for the pathogen may vary according to the infected individual's clinical state and disease outcomes. Depending on the disease progression, infected individuals shed a variable amount of pathogen while they are in the shedding states (A, I, J and Z). The total pathogen concentration entering the wastewater at time t is given by

$$W^*(t) = \omega \times \left(\sum_{k=1}^{n_J} \lambda_k J_k(t) + \sum_{k=1}^{n_I} \lambda_k I_k(t) + \xi \sum_{k=1}^{n_A} \lambda_k^{\text{A}} A_k(t) + \sum_{k=1}^{n_Z} \lambda_k^{\text{Z}} Z_k(t) \right) \tag{6}$$

The parameters λ_k, λ_k^{A} and λ_k^{Z} represent the faecal shedding dynamics per capita when the infected individual is in the epidemiological states I, J, A and Z respectively. For simplicity (and likely lack of observational data), we used the same parameters λ_k for all epidemiological states (*i.e.*, $\lambda_k = \lambda_k^{\text{A}} = \lambda_k^{\text{Z}}$). Values for the parameters λ_k can be chosen to represent mid-range values of the different results published (see Appendix Figure S2 for an example on SARS-CoV-2). Note that we assume the same reduction in faecal shedding as in respiratory shedding for asymptomatic cases (parameter ξ). The parameter ω implies that our model can only determine up to a constant the concentration of SARS-CoV-2 in wastewater [**?**], even if the limit of detection of the assay is known. This reflects the inability (for most pathogens) to quantify the various complex processes that affect the concentration, from patients' shedding to the concentration measured in laboratories (*e.g.*, frequency and timing of sampling, RNA degradation in the sewer system, recovery efficiency of assays).

1.2.2 RNA transport and sampled viral concentration

We use a simple advection-dispersion-decay model to simulate the fate of thepathogen along its journey in wastewater from the shedding points to the sampling site. This model is a combination of an exponential viral decay [**?**] and a τ-day dispersed plug-flow function, $g(\tau)$, representing all possible hydrodynamic processes (*e.g.*, dilution, sedimentation and resuspension) that leads to RNA degradation as well as decrease and delay of signal at the time of sampling. The dispersed plug-flow $g(\tau)$ acts as a transformation function, which reshapes the initial deposited concentration, W^*, into a delayed viral distribution over τ days as a result of the transit of the pathogen in the sewer system. Hence, we defined the sampled viral concentration at time t as:

$$W_{\text{samp}}(t) = \int_0^t W^*(t-\tau)\, g(\tau)\, e^{-\kappa\tau} \mathrm{d}\tau, \tag{7}$$

where κ is the daily decay rate of the pathogen due to the harsh, complex and bioactive environment of wastewater [**?**]. The pathogen concentration entering the sewage system daily is modelled as a single hydrodynamic pulse per day and the plug-flow function, g, is obtained by the analytical solution of the axial dispersed plug flow differential equation [**?**]. We then re-parametrize the analytical solution with the mean delay time $\bar{\tau}$ and its standard deviation σ into a Gaussian distribution

$$g(\tau) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\tau-\bar{\tau})^2}{2\sigma^2}\right). \tag{8}$$

See the Appendix for more details on our advection-dispersion-decay model of the RNA transport.

1.2.3 Wastewater reported sample

The sample transportation to the analysis site, laboratory processing time and reporting lags, introduce reporting delays for the pathogen concentration in wastewater. Hence, we define the reported wastewater concentration as

$$W(t) = W_{\text{samp}}(t - \ell_{\text{ww}}), \tag{9}$$

where ℓ_{ww} is the reporting lag between wastewater sampling and concentration report after laboratory analysis. (Note that the reporting delay of the wastewater measurement is independent from the delay caused by the transport of pathogen particles in the sewer system as defined in Equation 7).

1.3 Clinical reported cases

We also model surveillance data derived from laboratory confirmed and clinically diagnosed cases, acknowledging that instantaneous identification and complete reporting after initial

infection is not possible. We assume that a fraction ρ of symptomatic incidence is reported with a lag of a ℓ_{clinical} days from the time of infection. If $i(t)$ is the total incidence at time t, we define the number of clinical cases reported at time t as:

$$C(t) = \rho\,(1-\alpha)\,i(t-\ell_{\text{clinical}}), \tag{10}$$

1.3.1 Fit to data

We use an Approximate Bayesian Computation (ABC) algorithm [**?**] to fit the unknown or unobserved model parameters to the available data. For each ABC prior iteration, the error function is defined as a weighted trajectory matching

$$e_i = w_C(C - C_{\text{obs}})^2 + w_H(H - H_{\text{obs}})^2 + w_W(W - W_{\text{obs}})^2 \tag{11}$$

Depending on the available data, the parameters fitted the can be the time-dependent transmission rate β_t, ω, the hospitalization rate h and the mean transit time $\bar{\tau}$. More details about the fitting procedure is given in Appendix.

As an example, one can define three types of fitting-to-data procedures. "Clinical" when $w_C = w_H = 1$ and $w_W = 0$, to use data from clinical sources only ; "wasterwater-only" when $w_W = 1$ and $w_H = w_C = 0$, to use wastewater data only; and finally "Combined" by choosing the weights w_C, w_H and w_W such that the contribution of each error term (in Equation 11) are, on average, approximately equal. The "Combined" fitting procedure aims to have approximately the same contribution from clinical and wastewater data sources (despite the different observation frequencies).

1.3.2 Inference of unobserved epidemiological quantities

For a given location, once the model is fitted data, we can infer unobserved quantities of epidemiological importance by generating epidemic trajectories from the posterior samples. The posterior prevalence distribution (at each time point) is defined by simply adding the populations from the compartments representing active infection, that is

$$\text{prev}(t) = E + \sum_{i=1}^{n_A} A_i + \sum_{i=1}^{n_I} I_i + \sum_{i=1}^{n_J} J_i + H \tag{12}$$

The posterior cumulative incidence is obtained by summing Equation 1a until time t

$$\text{cuminc}(t) = -\sum_{i=1}^{t} \dot{S}_i \tag{13}$$

The fitted model can also provide an estimate of the effective reproduction number from the different data sources (e.g., clinical and/or wastewater) using Equation 3.

References

[1] Andrew F. Brouwer, Joseph N. S. Eisenberg, Connor D. Pomeroy, Lester M. Shulman, Musa Hindiyeh, Yossi Manor, Itamar Grotto, James S. Koopman, and Marisa C. Eisenberg. Epidemiology of the silent polio outbreak in rahat, israel, based on modeling of environmental surveillance data. *Proceedings of the National Academy of Sciences*, 115(45):E10625–E10633, 2018.

[2] Warish Ahmed, Paul M. Bertsch, Kyle Bibby, Eiji Haramoto, Joanne Hewitt, Flavia Huygens, Pradip Gyawali, Asja Korajkic, Shane Riddell, Samendra P. Sherchan, Stuart L. Simpson, Kwanrawee Sirikanchana, Erin M. Symonds, Rory Verhagen, Seshadri S. Vasan, Masaaki Kitajima, and Aaron Bivins. 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, 2020.

[3] A. Kayode Coker. Chapter eight - residence time distributions in flow reactors. pages 663–761, 2001.

[4] Mark A Beaumont, Wenyang Zhang, and David J Balding. Approximate Bayesian Computation in Population Genetics. *Genetics*, 162(4):2025–2035, 12 2002.

[5] Muge Cevik, Matthew Tate, Ollie Lloyd, Alberto Enrico Maraolo, Jenna Schafers, and Antonia Ho. Sars-cov-2, sars-cov, and mers-cov viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *The Lancet Microbe*, 2(1):e13–e22, 2021.

[6] Rita Jaafar, Sarah Aherfi, Nathalie Wurtz, Clio Grimaldier, Thuan Van Hoang, Philippe Colson, Didier Raoult, and Bernard La Scola. Correlation Between 3790 Quantitative Polymerase Chain Reaction–Positives Samples and Positive Cell Cultures, Including 1941 Severe Acute Respiratory Syndrome Coronavirus 2 Isolates. *Clinical Infectious Diseases*, 09 2020. ciaa1491.

[7] Jared Bullard, Kerry Dust, Duane Funk, James E Strong, David Alexander, Lauren Garnett, Carl Boodman, Alexander Bello, Adam Hedley, Zachary Schiffman, Kaylie Doan, Nathalie Bastien, Yan Li, Paul G Van Caeseele, and Guillaume Poliquin. Predicting Infectious Severe Acute Respiratory Syndrome Coronavirus 2 From Diagnostic Samples. *Clinical Infectious Diseases*, 71(10):2663–2666, 05 2020.

[8] Daniel Owusu, Mary A Pomeroy, Nathaniel M Lewis, Ashutosh Wadhwa, Anna R Yousaf, Brett Whitaker, Elizabeth Dietrich, Aron J Hall, Victoria Chu, Natalie Thornburg, Kimberly Christensen, Tair Kiphibane, Sarah Willardson, Ryan Westergaard, Trivikram Dasu, Ian W Pray, Sanjib Bhattacharyya, Angela Dunn, Jacqueline E Tate, Hannah L Kirking, Almea Matanock, and Household Transmission Study Team. Persistent SARS-CoV-2 RNA Shedding without Evidence of Infectiousness: A Cohort Study of Individuals with COVID-19. *The Journal of Infectious Diseases*, 02 2021. jiab107.

[9] Dorothy Anderson and Ray Watson. On the spread of a disease with gamma distributed latent and infectious periods. *Biometrika*, 67(1):191–198, 1980.

[10] Sukbin Jang, Ji-Young Rhee, Yu Mi Wi, and Bo Kyeung Jung. Viral kinetics of sars-cov-2 over the preclinical, clinical, and postclinical period. *International Journal of Infectious Diseases*, 102:561–565, 2021.

[11] Nadège Néant, Guillaume Lingas, Quentin Le Hingrat, Jade Ghosn, Ilka Engelmann, Quentin Lepiller, Alexandre Gaymard, Virginie Ferré, Cédric Hartard, Jean-Christophe Plantier, Vincent Thibault, Julien Marlet, Brigitte Montes, Kevin Bouiller, François-Xavier Lescure, Jean-François Timsit, Emmanuel Faure, Julien Poissy, Christian Chidiac, François Raffi, Antoine Kimmoun, Manuel Etienne, Jean-Christophe Richard, Pierre Tattevin, Denis Garot, Vincent Le Moing, Delphine Bachelet, Coralie Tardivon, Xavier Duval, Yazdan Yazdanpanah, France Mentré, Cédric Laouénan, Benoit Visseaux, and Jérémie and Guedj. Modeling sars-cov-2 viral kinetics and association with mortality in hospitalized patients from the french covid cohort. *Proceedings of the National Academy of Sciences*, 118(8), 2021.

[12] Anne Weiss, Mads Jellingso, and Morten Otto Alexander Sommer. Spatial and temporal dynamics of sars-cov-2 in covid-19 patients: A systematic review and meta-analysis. *EBioMedicine*, 58:102916, 2020.

[13] Amy E Benefield, Laura A Skrip, Andrea Clement, Rachel A Althouse, Stewart Chang, and Benjamin Muir Althouse. Sars-cov-2 viral load peaks prior to symptom onset: a systematic review and individual-pooled analysis of coronavirus viral load from 66 studies. *medRxiv*, 2020.

[14] Mathilde Bellon, Stephanie Baggio, Frederique Jacquerioz Bausch, Hervé Spechbach, Julien Salamun, Camille Genecand, Aglae Tardin, Laurent Kaiser, Arnaud G L'Huillier, and Isabella Eckerle. SARS-CoV-2 viral load kinetics in symptomatic children, adolescents and adults. *Clinical Infectious Diseases*, 05 2021.

[15] Till Hoffmann and Justin Alsing. Faecal shedding models for sars-cov-2 rna amongst hospitalised patients and implications for wastewater-based epidemiology. *medRxiv*, 2021.

[16] Fuminari Miura, Masaaki Kitajima, and Ryosuke 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. *Science of The Total Environment*, 769:144549, 2021.

[17] Roy M Anderson and Robert M May. *Infectious Diseases of Humans - Dynamics and Control*. Oxford University Press, 1991.

[18] Gertjan Medema, Leo Heijnen, Goffe Elsinga, Ronald Italiaander, and Anke 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. *Environmental Science & Technology Letters*, 7(7):511–516, 2020.

[19] Paola Foladori, Francesca Cutrupi, Nicola Segata, Serena Manara, Federica Pinto, Francesca Malpei, Laura Bruni, and Giuseppina La Rosa. Sars-cov-2 from faeces to wastewater treatment: What do we know? a review. *Science of The Total Environment*, 743:140444, 2020.

[20] Wen-Cheng Liu and Wei-Cher Huang. Modeling the transport and distribution of fecal coliform in a tidal estuary. *Science of The Total Environment*, 431:1–8, 2012.

[21] Christopher Staley, Kenneth H. Reckhow, Jerzy Lukasik, and Valerie J. Harwood. Assessment of sources of human pathogens and fecal contamination in a florida freshwater lake. *Water Research*, 46(17):5799–5812, 2012.

[22] Rob Jamieson, Doug M. Joy, Hung Lee, Ray Kostaschuk, and Robert Gordon. Transport and deposition of sediment-associated escherichia coli in natural streams. *Water Research*, 39(12):2665–2675, 2005.

[23] Bhuban Ghimire and Zhiqiang Deng. Hydrograph-based approach to modeling bacterial fate and transport in rivers. *Water Research*, 47(3):1329–1343, 2013.

[24] Sen Bai and Wu-Seng Lung. Modeling sediment impact on the transport of fecal bacteria. *Water Research*, 39(20):5232–5240, 2005.

[25] Kyung Hwa Cho, Yakov A. Pachepsky, David M. Oliver, Richard W. Muirhead, Yongeun Park, Richard S. Quilliam, and Daniel R. Shelton. Modeling fate and transport of fecally-derived microorganisms at the watershed scale: State of the science and future opportunities. *Water Research*, 100:38–56, 2016.

[26] Chris R. Rehmann and Michelle L. Soupir. Importance of interactions between the water column and the sediment for microbial concentrations in streams. *Water Research*, 43(18):4579–4589, 2009.

[27] Octave Levenspiel. *The Dispersion Model*, pages 47–70. Springer New York, New York, NY, 2012.

[28] A McDonald, D Kay, and A Jenkins. Generation of fecal and total coliform surges by stream flow manipulation in the absence of normal hydrometeorological stimuli. *Applied and Environmental Microbiology*, 44(2):292–300, 1982.

[29] J. W. Nagels, R. J. Davies-Colley, A. M. Donnison, and R. W. Muirhead. Faecal contamination over flood events in a pastoral agricultural stream in New Zealand. *Water Sci Technol*, 45(12):45–52, 2002.

[30] P. M. Gundy, C. P. Gerba, and I. L. Pepper. Survival of Coronaviruses in Water and Wastewater. *Food Environ Virol*, 2009.

Appendix

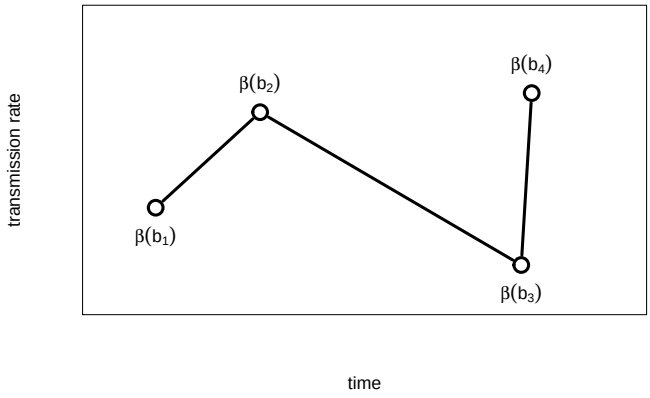

Figure S1: Illustration of the piecewise linear function to model the transmission rate β_t. The break dates $b_1, b_2, \ldots$ are chosen manually by visual inspection of the time series of interest, and the values $\beta(b_i)$ are fitted with an ABC algorithm.

A-1 Fitting procedure

To fit the model to the clinical and/or wastewater surveillance data, we model the transmission rate, β_t as a piecewise linear function. We parsimoniously choose the times $b_1, b_2, \ldots$ ("break times") defining each segment by visual inspection of the time series (*i.e.*, incidence of new COVID-19 cases and/or hospitalizations for clinical surveillance; SARS-CoV-2 concentration for wastewater surveillance). Those times should correspond to changes in the transmission dynamics. The value of the transmission rate at a break date, $\beta(b_i)$ is fitted using an Approximate Bayesian Computation (ABC). We chose not to fit the break times b_i because the fitting algorithm would require a computation time that would not be practical. See Figure S1.

The mean transit time $\bar{\tau}$ and the scaling factor ω are also fitted to data. For those parameters, the goal is primarily to allow for uncertainty rather than infer a posterior distribution as they are essentially not identifiable. Finally, the hospitalization rate is also fitted to the hospitalization data, when available.

Over the study period (from early March 2020 to June 1st, 2021), we have about a dozen break times b_i for each city to reflect the various interventions (*e.g.*, lockdowns) or behaviour (*e.g.*, change in contact rate when schools re-opened in September 2020). Hence, the parameter space to explore by the ABC algorithm is relatively large. To avoid unpractical computational times, we defined relative strong priors on all parameters, that is normal distributions with a mean close to the expected value (explored manually) and a relatively broad standard deviation (corresponding approximately to a coefficient of variation of

0.5). The normal distribution was censored to positive values. The outputs of the fitting procedure for all locations are shown in supplementary file File S2.

A-2 Relative infectiousness

In Eq. 1a, the force of infection from symptomatic cases is $\beta \sum_{k=1}^{n_I} \psi_k I_k$. For convenience, the ψ_k are chosen such that

$$\sum_{k=1}^{n_I} \psi_k = n_I. \tag{A.1}$$

Hence, when the infectiousness profile is constant ($\lambda_k = 1$, for $k = 1, \ldots, n_I$), the force of infection is $\beta \sum_{k=1}^{n_I} I_k$. Using this normalization allows to keep a single baseline parameter β. Only the *relative* values of the ψ_k affect the infectiousness profile. Similarly for asymptomatic infections, the parameters ϕ_k are chosen such that $\sum_{k=1}^{n_A} \phi_k = n_A$.

We assume the infectious period for symptomatic infections is 12 days on average [**?**, **?**, **?**, **?**] and divide this period into $n_I = 6$ sub-compartments $I_1, \ldots, I_6$ where infected individuals will stay, on average, 2 days in each of them. Note that with this representation, the duration of infectiousness has an Erlang distribution [**?**] with shape n_I (and mean 12 days). The parameter ψ_k represents the relative infectiousness of sub-compartment I_k. We assume infectiousness, that is the probability of transmission given contact, is proportional to the log viral load measured from respiratory samples in clinical studies [**?**, **?**] and choose $\tilde{\psi} = (3, 6, 5, 4, 3, 2)$ and then normalize according to Equation A.1 with $\psi_k = n_I \tilde{\psi}_k / \sum_k \tilde{\psi}_k$.

Similarly, we assume a shorter infectious period for asymptomatic infections of 10 days on average [**?**, **?**], divided into $n_A = 5$ sub-compartments with an average stay of 2 days each.

A-3 Respiratory and faecal Viral kinetics

Our model explicitly accounts for the temporal profile of respiratory shedding via the multiple sub-compartments for the infectious states (A, I and J) combined with the parameters ϕ and ψ. Similarly, it explicitly accounts for the faecal shedding kinetics via the shedding states (A, I, J and Z) and the parameters λ.

We parameterize our model such that the SARS-CoV-2 viral kinetics reflect with levels reported in the literature. The black line with points in Figure S2 shows the values used for respiratory (top panel) and faecal (bottom panel) shedding, and how they compare to observational studies.

A-4 Effective reproduction number

This section is not up to date! Check future versions...

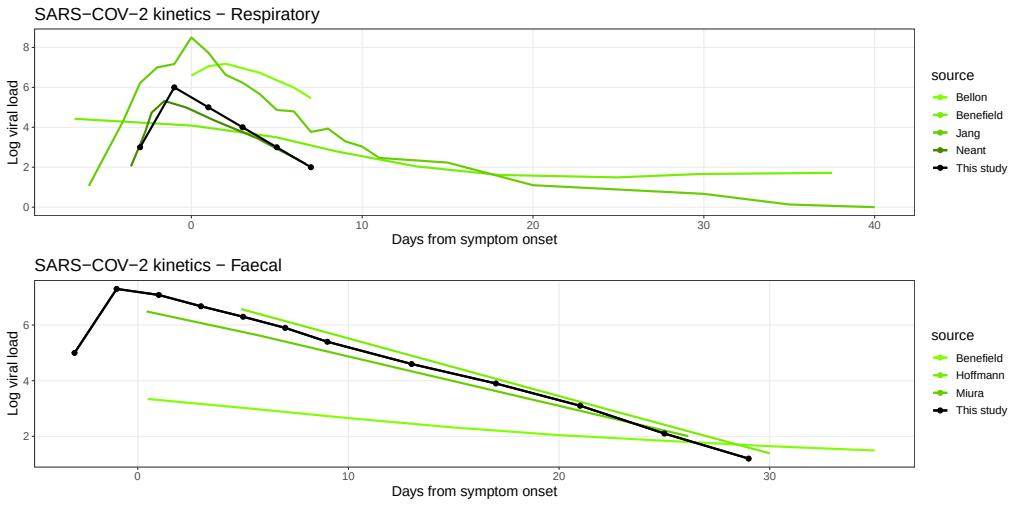

Figure S2: SARS-CoV-2 viral kinetics. The values used in our model are represented by the black curve and the green curves show estimates from the literature. The full reference of each study can be found in the bibliography: Benefield [?], Bellon [?], Jang [?], Neant [?], Hoffmann [?], Miura [?].

As a first step, we establish the *basic* reproduction number, $\mathcal{R}_0$, for the model defined by equations Eq. 1a - 1n. To derive $\mathcal{R}_0$, we follow the methodology presented in [?].

Asymptomatic individuals are infectious for an average duration of $1/\theta$, and the ratio of their infectiousness compared to all the other infectious states (I and J) is ξ. Asymptomatic incidence is a proportion α of the overall incidence. Hence, the contribution to the basic reproduction number from the asymptomatically infected individuals is

$$\mathcal{R}_0^A = \beta\alpha\xi/\theta \tag{A.2}$$

Infectious individuals that are symptomatic and that will recover without hospitalization (I) are infectious for an average duration of $1/\nu$. Their proportion of the overall incidence is calculated by simply stating they are not hospitalized $(1-h)$ and not asymptomatic $(1-\alpha)$, hence the proportion is $(1-h)(1-\alpha)$. The contribution to the basic reproduction number from the symptomatically infected individuals that will not require hospitalization is

$$\mathcal{R}_0^I = \beta(1-h)(1-\alpha)/\nu \tag{A.3}$$

For the symptomatically infected persons that will require hospitalization (J), the same logic applies to the I sub-group, except that we need to take into account their relative infectiousness compared to I, which is $\sum_{k=1}^{n_J}\psi_k / \sum_{k=1}^{n_I}\psi_k$. Note that the denominator is simply n_I thanks to the normalization defined in Equation A.1. The subgroup in state

J are hospitalized (h) and not asymptomatic ($1-\alpha$) so their contribution to the basic reproduction number is

$$\mathcal{R}_0^J = \frac{\sum_{k=1}^{n_J} \psi_k}{n_I} \beta h(1-\alpha)/\mu \tag{A.4}$$

We obtain the (overall) basic reproduction number by summing the respective contributions from all infectious epidemiological states

$$\mathcal{R}_0 = \mathcal{R}_0^A + \mathcal{R}_0^I + \mathcal{R}_0^J \tag{A.5}$$

Finally, the *effective* reproduction number $\mathcal{R}_t$ is simply defined as $\mathcal{R}_t = \frac{S_t}{N}\mathcal{R}_0$ which gives Eq. 2 in the main text.

A-5 Fate and RNA transport in wastewater

Upon fecal deposition into the wastewater, viral RNA undergoes various hydrodynamic processes and degrades during its journey from the shedding point to the sampling site [**?**, **?**]. This degradation mainly comes from RNA dilution in municipal wastewater constitutes (*e.g.*, hygiene products, household detergents, industrial wastewater and storm waters) and RNA decay resulting from harsh wastewater environment (*e.g.*, temperature, bioactive chemicals, solids, pH, etc.). As a common practice, complex hydraulic models simulate the in-fluid transportation of the water contaminants (endemic viruses and fecal microorganisms) via solving a set of physics-based differential equations describing the flow and transport mechanisms [**?**, **?**, **?**, **?**, **?**, **?**, **?**].

These hydrodynamical models are advection and dispersion mechanisms, which describe the microorganisms' transportation by the flow velocity along the longitudinal axis and its diluting process into the surrounding fluid [**?**]. Once a mass concentration enters the stream, it gradually disperses due to many physical factors such as dissolving process, velocity profile, turbulent mixing, molecular diffusion, etc. [**?**]. Moreover, sediment association plays a significant role in the transport models. Several studies indicated attachments of microorganisms to sediments and the impact of this associations on their delayed transportation [**?**, **?**, **?**]. Due to solid mass and subsequent gravitational pull, sediment's velocity differs from the flow velocity and results in solid settlement at the bottom of the stream. These microorganism-attached sediments, later, re-suspend during overflow periods, induced by heavy rainfall and industrial discharges, and act as a reservoir and increase in-stream microorganism's concentration irrespective of new input from the shedding source [**?**, **?**]. Enveloped viruses, such as SARS-CoV-2, dissolve less in water and tends to attach to solids as a result of their hydrophobic envelope [**?**, **?**], and therefore, settlement/resuspension may play a key role in transportation of the SARS-CoV-2 genetic materials in wastewater.

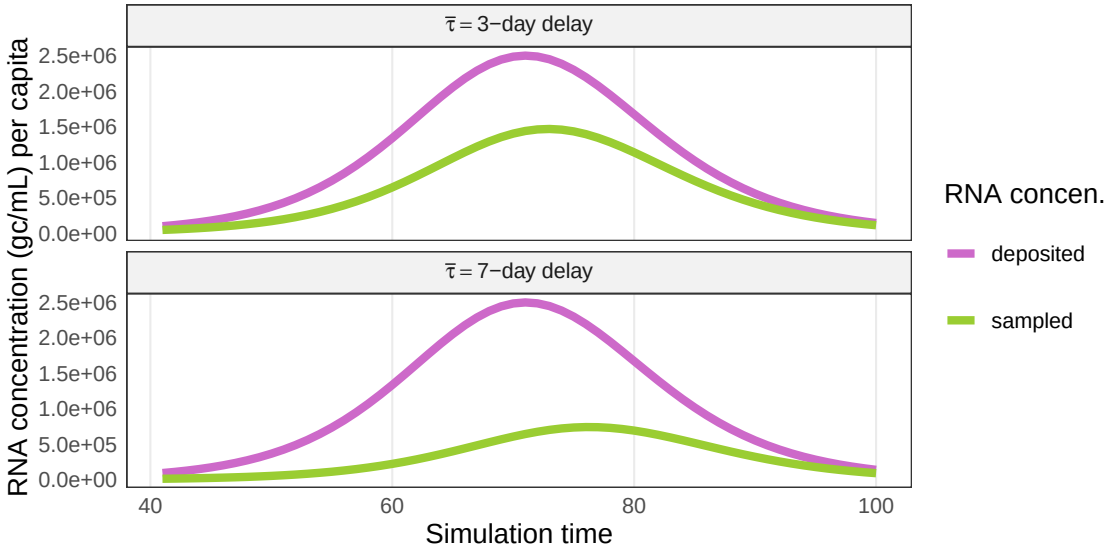

Figure S3: Impact of advection-dispersion-decay model (Eq. A.8). sampled RNA concentration (W_{samp}) may have different viral distribution compared to deposited density (W^*) due to delayed viral genetic materials resulting from hydrodynamical phenomenon in sewer system.

Here, we used a simple advection-dispersion-decay model to simulate the fate of SARS-CoV-2 RNA along their journey from shedding points to the sampling site. We assumed the deposited RNA concentration in the sewage system is a one-time pulse input per day and described the transfer process by a dispersed plug-flow model. As the plug concentration of viral RNA enters into the flow with velocity u, it gradually disperses along the longitudinal axis (flow direction) with dispersion coefficient $\mathcal{D}$ (m^2/s) within the sewage pathway. The length from the shedding location to the sampling site is L. The total RNA mass arrives at the sampling point gradually over time, such that the daily pulse input concentration is delayed. Assuming a small deviation from plug flow ($\mathcal{D}/uL \leq 0.01$), we can use the analytical solution of the 1-dimensional axial dispersed plug flow differential equation [**?**, **?**] as a transfer function for viral RNA in wastewater. For low diffusion limit, the transfer function is approximately symmetrical and defined by a Gaussian distribution $g(\tau)$, which represents the fraction of the deposited concentration at the sampling site τ days after its introduction to the sewage system

$$g(\tau) = \sqrt{\frac{u^3}{4\pi\mathcal{D}L}} \exp\left(-\frac{(L-u\tau)^2}{4\mathcal{D}L/u}\right). \tag{A.6}$$

Defining the mean transit time $\bar{\tau} = L/u$ and its standard deviation $\sigma = \sqrt{2\mathcal{D}L/u^3}$, we can

re-parametrize Equation A.6 based on the first two moments of the transit time:

$$g(\tau) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\tau - \bar{\tau})^2}{2\sigma^2}\right), \tag{A.7}$$

In addition to delay, the genetic materials of SARS-CoV-2 degrades exponentially at a daily rate κ due to the complex environment of wastewater [**?**]. As a result, $W(t)$, the sampled concentration at time t, is a combination of both delayed viral materials and degradation

$$W_{\text{samp}}(t) = \int_0^t W^*(t-\tau)\, g(\tau)\, e^{-\kappa\tau} \mathrm{d}\tau, \tag{A.8}$$

where W^* is the initial daily deposited concentration.

Table 2: Description of the model's compartments and parameters for the SARS-CoV-2 RNA transmission and disease outcome.

Symbole	Definition
S	susceptibles
V_{wa}	vaccine-administered individuals waiting to build immunity
V	vaccinated individuals with possible full protection
E	exposed susceptibles but not infectious
E^V	exposed vaccinated individuals but not infectious
A_k	asymptomatic infectious cases in kth subcompartment
I_k	symptomatic infectious cases in kth subcompartment
J_k	symptomatic infectious cases in kth subcompartment who later admits to hospital
Z_k	non-infectious cases but fecal shedding SARS-CoV-2 RNA in kth subcompartment
H	hospitalized patients
R	recovered cases
D	deceased cases
β_t	transmission rate (per contact)
$1/\varepsilon$	ave. latency time (days)
$1/\nu_k$	ave. duration in kth subcompartment among symptomatics (days)
$1/\mu_k$	ave. duration in kth subcompartment among symptomatics goes to hospital (days)
$1/\theta_k$	ave. duration in kth subcompartment among asymptomatics (days)
$1/\eta_k$	ave. duration in kth subcompartment of shedding after infectiousness (days)
$1/\ell$	ave. length of stay in a hospital (days)
$1/\tau$	ave. duration of waining immunity with zero protection (days)
$1/d$	ave. duration from day one of vaccine administration to full protection (days)
$n_{_I}$	total number of subcompartments in I state
$n_{_J}$	total number of subcompartments in J state
$n_{_A}$	total number of subcompartments in A state
$n_{_Z}$	total number of subcompartments in Z state
r_t	proportional rate of vaccine administration among susceptibles (vaccinated per population per day)
α	proportion of exposed cases that are asymptomatic/mild symptomatic
h	proportion of symptomatic cases that need hospital admission
α'	proportion of exposed cases that are asymptomatic among vaccinated individuals
h'	proportion of symptomatic cases that need hospital admission among vaccinated individuals
δ	proportion of deceased individuals among hospitalized patients
eff_{inf}	vaccine effectiveness against infection given exposure