Supplementary Information for *"Model-based index reveals highly varied correlation between norovirus wastewater surveillance and gastroenteritis hospital admissions"*

Yifan Zhu [a], Mayuko Saito [b], Takayuki Miura [c], Shinobu Kazama [d], Daisuke Sano [a,e,*]

[a] Department of Frontier Sciences for Advanced Environment, Graduate School of Environmental Studies, Tohoku University, Sendai, Miyagi, Japan
[b] Department of Virology, Graduate School of Medicine, Tohoku University, Sendai, Miyagi, Japan
[c] Department of Environmental Health, National Institute of Public Health, Wako, Saitama, Japan
[d] Department of Urban Engineering, The University of Tokyo, Tokyo, Japan
[e] Department of Civil and Environmental Engineering, Graduate School of Engineering, Tohoku University, Sendai, Miyagi, Japan

Corresponding author: Daisuke Sano, E-mail: daisuke.sano.e1@tohoku.ac.jp

Fig. S1. The norovirus fecal shedding profile used for calculating patient viral load L_p.

The red line represents the median shedding rate while the shaded area is the 95% confidence interval used as the upper and lower limits of the calculation.

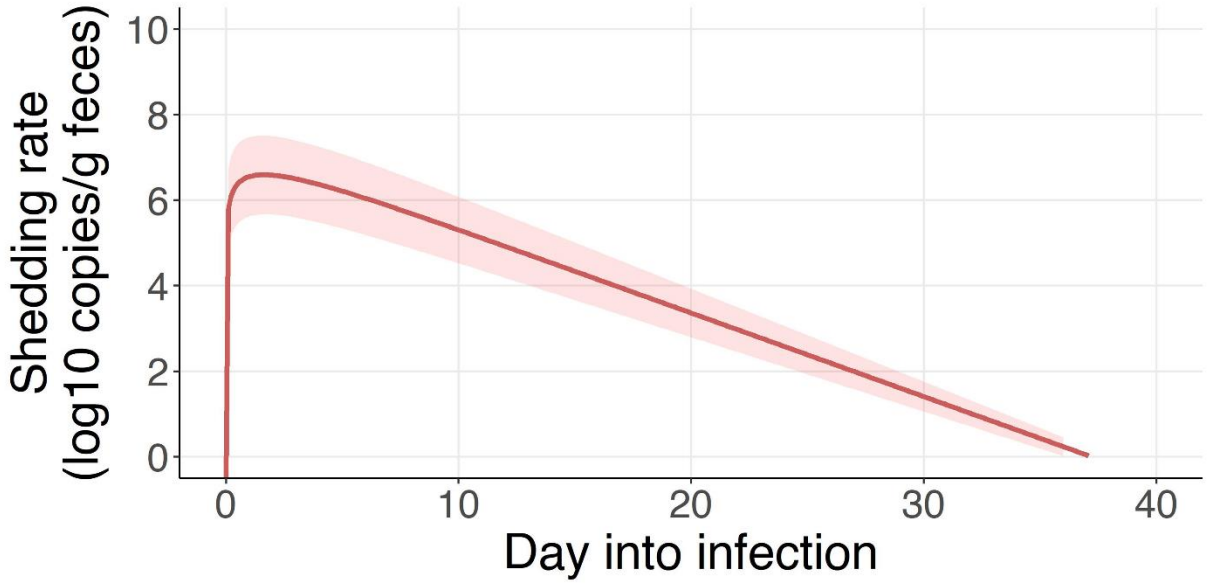

Fig. S2. The posterior distribution of α_m and the monthly average temperature in Sendai.

The red points and error bars are the median value and IQR of α_m. The blue points and shaded area are the average, lowest and highest daytime temperature of each month (data source: Japan Meteorological Agency (https://www.data.jma.go.jp/obd/stats/etrn/view/monthly_s3_en.php?block_no=47590&view=1, accessed on November 25, 2021)). Spearman's rank correlation analysis was performed (ρ = -0.69, p = 0.016).

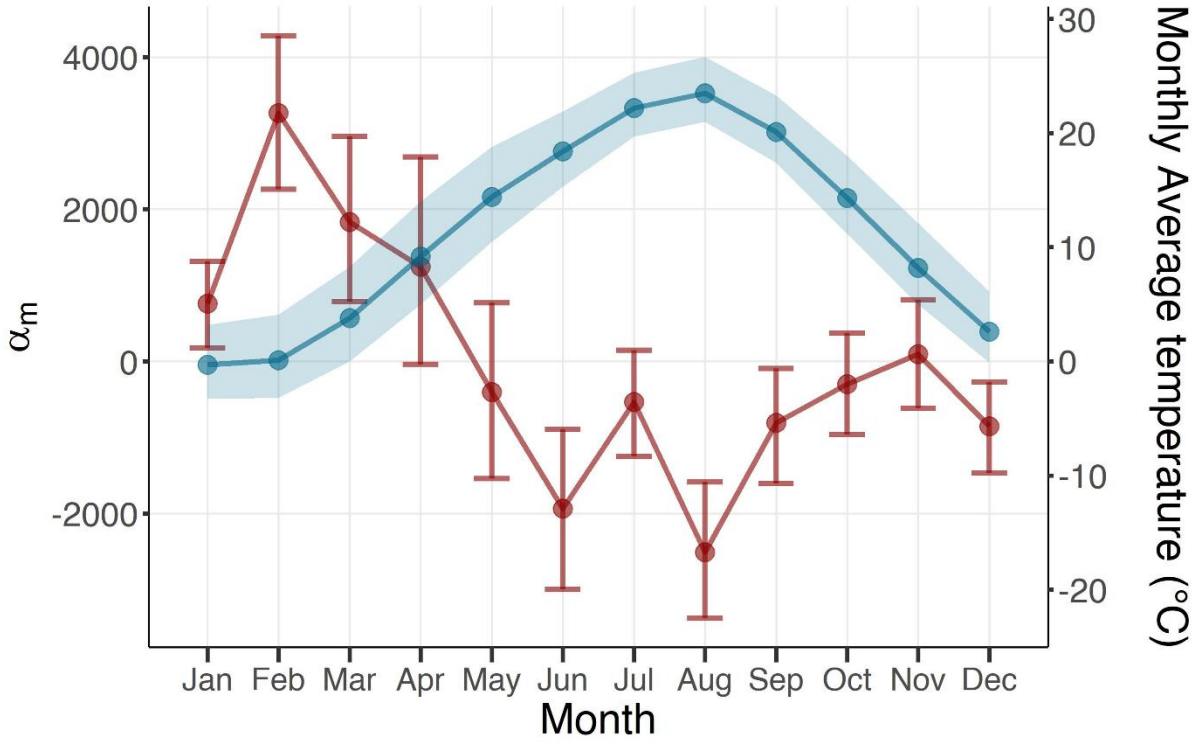

Table S1. Spearman's rank correlation coefficients between L_w and L_p for each epidemic year.

Epidemic year	2016/2017	2017/2018	2018/2019	2020/2021	Overall
ρ	0.7974	-0.0383	0.2815	0.2702	0.2779
p-value	< 0.0001	0.79	0.022	0.031	< 0.0001

Norovirus wastewater detection

Polyethylene glycol (PEG) precipitation

Grab wastewater samples were collected once or twice a week at around 10 a.m. from a municipal wastewater treatment plant (WWTP) in Sendai, Japan. The WWTP has two influent lines that cover the downtown area and residential/industrial area, respectively. Samples were transported to the laboratory in ice. A 40 mL sample was inoculated with 100 µL of pre-prepared murine norovirus (MNV) suspension for recovery evaluation, 3.2 g PEG 6000, and 0.92 g NaCl, the mixture was stirred overnight at 4°C. Then, the mixture was centrifuged at 8,000 g for 30 minutes at 4°C. After centrifugation, the supernatant was removed and the pellet was resuspended with 1 mL ddH2O. After mixed by vortex for 90 seconds, the suspension was centrifuged again at 9,000 g for 10 minutes at 4°C. The pellet was discarded, then the supernatant was used for subsequent steps and its volume was measured for later calculation.

RNA extraction, cDNA synthesis, and RT-qPCR

A 140 µL aliquot obtained from PEG precipitation was used for viral RNA extraction using the QIAamp Viral RNA Mini Kit (QIAGEN, Hilden, Germany) with a resulting volume of 60 µL. Then, a 10 µL extracted RNA was used to synthesize cDNA with the iScript™ cDNA Synthesis Kit (Bio-Rad, Hercules, CA, USA) following the manufacture's instruction, the resulting volume is 20 µL.

For RT-qPCR, we used a previously reported primer/probe set (*1*). For each 20 µL reaction, 10 µL of SsoAdvanced Universal Probes Supermix (Bio-Rad), 5 µL of cDNA, 0.8 µL of 10 µM forward and reverse primers, 0.2 µL of 10 µM probe, and 3.2 µL of PCR-grade water were used. The PCR cycling condition was 30 seconds at 95°C, followed by 40 cycles of 15 seconds at 95°C, 60 seconds at 56°C, and 30 seconds at 72°C.

To evaluate the whole process recovery efficiency, the concentration of MNV in precipitated wastewater samples was compared with that of the stock. For MNV, we used the following assay: forward primer: CGGTGAAGTGCTTCTGAGGTT, reverse primer: GCAGCGTCAGTGCTGTCAA, TaqMan probe: FAM- CGAACCTACATGCGTCAG - TAMRA. The PCR cycling condition was 30 seconds at 95°C, followed by 40 cycles of 5

seconds at 95°C, 20 seconds at 56°C, and 30 seconds at 72°C. RT-qPCR was performed in triplicate and the mean value was used subsequent calculations.

Wastewater concentration calculation

The concentration of norovirus GII viral RNA in influent wastewater sample, $C_{v,w}$ (copies/mL), was converted from qPCR-determined concentration using the following equation:

$$C_{v,w} = 1000 \times \frac{V}{V_w} \times \frac{V_{f,ex}}{V_{s,ex}} \times \frac{V_{f,syn}}{V_{s,syn}} \times C_{qPCR} \quad (1)$$

where V is the volume of sample after PEG precipitation [mL], C_{qPCR} is the concentration of cDNA applied to qPCR [copies/µL], V_w is for the volume of raw wastewater [mL]. $V_{f,ex}$, $V_{s,ex}$, $V_{f,syn}$, and $V_{s,syn}$ are all the volume of samples in the intermediate steps [µL]. Subscripts *f* and *s* stand for final and starting volume, while subscripts *ex* and *syn* stand for RNA extraction and cDNA synthesis, respectively.

Details of the wastewater and patient viral load models

Patient number conversion

Function "na_interpolation" in R package "imputeTS" was used to convert weekly case count into daily case count.

Shedding function

A norovirus fecal shedding function previously described by Teunis et al. (2015) was used (*2*). Briefly, the rate and duration of fecal shedding can be described by the following equation:

$$C_f(t|\alpha,\beta) = C_0 e^{-\alpha t}\left(1 - e^{-(\beta-a)t}\right)(\beta - \alpha) \quad (2)$$

where $C_f(t|\alpha,\beta)$ is the concentration of viral RNA in the feces (copies/g) of infected individuals at time t while α, β, and C_0 are shape coefficients. This shedding function features a high shedding rate in the early stage of infection, followed by a downward trend that lasts days, even weeks before the fecal shedding eventually stops. More details, including the biological explanation, can be found in the original study. The shedding rate curve is shown in Fig. S1.

Viral load calculation

As the total viral load is contributed by individuals in different infection stages, assuming a study period *i* and a shedding period *j*, reported patients, after adjustment for norovirus GII ratio, were assigned into an $i \times j$ matrix called $M(P_N)$ (Equation (3)).

$$M(P_N) = \begin{bmatrix} P_{N[1,1]} & P_{N[1,2]} & \cdots & P_{N[1,j]} \\ P_{N[2,1]} & P_{N[2,2]} & \cdots & P_{N[2,j]} \\ \cdots & \ddots & \ddots & \cdots \\ P_{N[i,1]} & P_{N[i,2]} & \cdots & P_{N[i,j]} \end{bmatrix} \quad (3)$$

In this way, the number of reported patients on their day *n* of the infection course on calendar day *d*, $P_{N[n,d]}$, can be called from the matrix for viral load calculation. Due to the change of infection status, there is a transition $P_{N[n,d]} = P_{N[(n-1),(d-1)]}$.

Then, the viral load from reported patients on a particular day d can be calculated using the following equation:

$$L_{p[d]} = \sum_{n=1}^{j} P_{N,r[n,d]} * C_{f,n} * r_F * m_f * 0.7 \quad (4)$$

where L_p is the daily total patient viral load (copies) and $P_{N,r}$ represents the reported NoV GII infections. Daily viral load $C_{f,n}$ (copies/patient/g feces) was calculated as the mean value between $C_f(n = t)$ and $C_f(n = t + 1)$ using 0.1 days as interval, r_F is the ratio of patients who develop fecal shedding, m_f stands for the average weight of feces (g) excreted from the human body. Although the fecal shedding would last for about 40 days according to the original study, the majority of viral load is contributed by patients in the early stage of infection as the shedding rate drops exponentially after the initial stage. Therefore, in this study, patients were assumed to shed viral RNA for 10 days after the initial virus contraction (n = 10), which is longer than previously reported symptomatic period (*3*). Also, the calculation assumes fecal shedding starts at the same time when an individual was reported. Although norovirus has an incubation period of about 33 hours (*3*), considering the prevalence is reported weekly, the potential lag between infection and shedding can be neglected. Finally, considering the WWTP serves about 70% of the population in Sendai, the calculated patient viral load was multiplied by 0.7.

Equation (5) was used to calculate the daily wastewater viral load from the measured wastewater virus concentration.

$$L_{w[d]} = C_{v,w[d]} * Q_d \quad (5)$$

where the measured concentration of norovirus RNA in the wastewater $C_{v,w}$ (copies/mL) is multiplied by the wastewater flow Q obtain the total load L_w. The average daily wastewater flow rate was acquired from the treatment plant operator.

Characterizing the scaling factor L_w/L_p

If the corresponding relation between L_w and L_p can be established, L_p would be able to be calculated from L_w, leading to the estimation of patient number. To investigate the extent of uncertainty and show whether WBE can deliver reliable quantitative prevalence estimation, we characterized $\alpha = \frac{L_w}{L_p}$ (termed "scaling factor" hereafter) using Bayesian inference.

Briefly, it was considered that α consists of four parts, an overall baseline value, a year-specific add-on value, a month-specific add-on value, and an add-on value for a certain year-month combination (Equation (6)). MCMC simulation was performed in the R programming language (*6*), the year and month were taken as nominal predictors while α_i was the metric output.

$$\alpha_i = \frac{L_{w[i]}}{L_{p[i]}} \sim N\left(\alpha_0 + \alpha_{y[i]} + \alpha_{m[i]} + \alpha_{y,m[i]}, \sigma_i^2\right) \quad (6)$$

MCMC configuration

```
# For hyper-prior on deflections:
gammaShRaFromModeSD = function( mode , sd ) {
  if ( mode <=0 ) stop("mode must be > 0")
  if ( sd <=0 ) stop("sd must be > 0")
  rate = ( mode + sqrt( mode^2 + 4 * sd^2 ) ) / ( 2 * sd^2 )
  shape = 1 + mode * rate
  return( list( shape=shape , rate=rate ) )
}

aGammaShRa = unlist(gammaShRaFromModeSD(mode=sd(y)/2, sd=2*sd(y)))
sGammaShRa = unlist(gammaShRaFromModeSD(mode=medianCellSD, sd=2*sdCellSD))

# Model configuration
model {
```

```
154     for ( i in 1:Ntotal ) {
155       y[i] ~ dt( mu[i] , 1/(ySigma[x1[i],x2[i]])^2 , nu )
156       mu[i] <- a0 + a1[x1[i]] + a2[x2[i]] + a1a2[x1[i],x2[i]]
157     }
158
159     nu ~ dexp(1/30.0)
160
161     for ( j1 in 1:Nx1Lvl ) { for ( j2 in 1:Nx2Lvl ) {
162       sigma[j1,j2] ~ dgamma( sigmaSh , sigmaRa )
163       # Prevent from dropping too close to zero:
164       ySigma[j1,j2] <- max( sigma[j1,j2] , medianCellSD/1000 )
165     } }
166
167     sigmaSh <- 1 + sigmaMode * sigmaRa
168     sigmaRa <- ( sigmaMode + sqrt( sigmaMode^2 + 4*sigmaSD^2 ) ) /(2*sigmaSD^2)
169     sigmaMode ~ dgamma(sGammaShRa[1],sGammaShRa[2])
170     sigmaSD ~ dgamma(sGammaShRa[1],sGammaShRa[2])
171
172
173     a0 ~ dnorm( yMean , 1/(ySD*5)^2 )
174
175     for ( j1 in 1:Nx1Lvl ) { a1[j1] ~ dnorm( 0.0 , 1/a1Sigma^2 ) }
176     a1Sigma ~ dgamma(aGammaShRa[1],aGammaShRa[2])
177
178     for ( j2 in 1:Nx2Lvl ) { a2[j2] ~ dnorm( 0.0 , 1/a2Sigma^2 ) }
179     a2Sigma ~ dgamma(aGammaShRa[1],aGammaShRa[2])
180
181     for ( j1 in 1:Nx1Lvl ) { for ( j2 in 1:Nx2Lvl ) {
```

```
    a1a2[j1,j2] ~ dnorm( 0.0 , 1/a1a2Sigma^2 )
  } }

  a1a2Sigma ~ dgamma(aGammaShRa[1],aGammaShRa[2]) # or try a folded t (Cauchy)

  # Convert a0,a1[],a2[],a1a2[,] to sum-to-zero b0,b1[],b2[],b1b2[,] :
  for ( j1 in 1:Nx1Lvl ) { for ( j2 in 1:Nx2Lvl ) {
    m[j1,j2] <- a0 + a1[j1] + a2[j2] + a1a2[j1,j2] # cell means
  } }

  b0 <- mean( m[1:Nx1Lvl,1:Nx2Lvl] )

  for ( j1 in 1:Nx1Lvl ) { b1[j1] <- mean( m[j1,1:Nx2Lvl] ) - b0 }
  for ( j2 in 1:Nx2Lvl ) { b2[j2] <- mean( m[1:Nx1Lvl,j2] ) - b0 }
  for ( j1 in 1:Nx1Lvl ) { for ( j2 in 1:Nx2Lvl ) {
    b1b2[j1,j2] <- m[j1,j2] - ( b0 + b1[j1] + b2[j2] )
  } }
}
```

References

1. T. Kageyama, S. Kojima, M. Shinohara, K. Uchida, S. Fukushi, F. B. Hoshino, N. Takeda, K. Katayama, Broadly reactive and highly sensitive assay for Norwalk-like viruses based on real-time quantitative reverse transcription-PCR. *J. Clin. Microbiol.* **41**, 1548–1557 (2003).

2. P. F. M. Teunis, F. H. A. Sukhrie, H. Vennema, J. Bogerman, M. F. C. Beersma, M. P. G. Koopmans, Shedding of norovirus in symptomatic and asymptomatic infections. *Epidemiol. Infect.* **143**, 1710–1717 (2015).

3. T. Devasia, B. Lopman, J. Leon, A. Handel, Association of host, agent and environment characteristics and the duration of incubation and symptomatic periods of norovirus gastroenteritis. *Epidemiol. Infect.* **143**, 2308–2314 (2015).

4. A. Campisano, C. Modica, Appropriate resolution timescale to evaluate water saving and retention potential of rainwater harvesting for toilet flushing in single houses. *J. Hydroinformatics*. **17**, 331–346 (2015).

5. D. Butler, E. Friedler, K. Gatt, Characterising the quantity and quality of domestic wastewater inflows.

6. J. Kruschke, Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan (2014).