THE LANCET
Planetary Health

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: McMahan CS, Self S, Rennert L, et al. COVID-19 wastewater epidemiology: a model to estimate infected populations. *Lancet Planet Health* 2021; **5:** e874–81.

COVID-19 Wastewater Epidemiology: A Model to Estimate Infected Population

Supplementary Appendix

SARS-CoV-2 DETECTION AND QUANTIFICATION

Virus Recovery from Sewage

Sewage samples were collected using a programmable peristaltic pump to collect fixed volumes over a 24-hour period. The large jug receiving the incremental samples was kept cold by packing it in ice within the sampling equipment. After completion of sample collection, the composite sample was mixed, and 500 mL were decanted into a factory clean plastic bottle and held on ice in a cooler. The sample was shipped cold with ice or frozen gel packs for next day delivery to the analytical lab.

Upon receipt at the lab, sewage samples (225 mL) were prepared for analysis by pasteurization at 60 ºC for 30 minutes. The pasteurized samples were centrifuged at 6,500 g for 10 minutes at 6 ºC to remove solids. The supernatant was decanted into a chilled beaker in an ice bath. Viruses were purified using the method described by Killington et al. (1996). Reagent grade sodium chloride and polyethylene glycol 6000 (PEG) (Sigma Aldrich, St. Louis, MO) were added directly into the chilled supernatant with constant stirring until all the salt and PEG dissolved to achieve a final concentration of 2.3% salt w/v and 7% PEG. Virus was allowed to precipitate for several hours at 4 ºC. The virus was then collected in the pellet produced by centrifuging at 15,000g for 20 minutes at 6 ºC. The supernatant was decanted and the pellet was dissolved in 2 mL of Tris-EDTA-Salt (TES) buffer (Killington, et al. 1996).

RNA Extraction and RT-qPCR

RNA in the resuspended virus pellet was extracted in TRIzol™ reagent following the manufacturer's instructions (ThermoFisher Scientific, Waltham, MA). The resulting RNA extract was prepared for RT-qPCR using a NucleoMag® Pathogen RNA/DNA Isolation Kit (Macherey-Nagel, Düren, Germany). RT-qPCR was performed in a Bio-Rad IQ5 instrument using 6.25 µL of 4X concentrated TaqPath™ 1-Step Multiplex Master Mix (No ROX™) (ThermoFisher Scientific), 7.5 µL of nuclease-free water, 1.25 µL of COVID-19 Real Time PCR Assay Multiplex (primers/probes mix), and

10 µL of template RNA in clear 96 well plates (VWR International, Mississauga, ON) and Microseal B® 96-well plate adhesive covers (Bio-Rad Laboratories, CA). The thermal program included an initial uracil-N-glycosylase gene (UNG) incubation at 25°C for 2 minutes, one reverse transcription at 53°C for 10 minutes, activation at 95°C for 2 minutes, denaturation at 95°C for 3 seconds, and annealing and extension at 60°C for 30 seconds. The PCR step in the reaction was repeated for 45 thermal cycles but measurements were not made at greater than 40 PCR thermal cycles (Cq >40).

SARS-CoV-2 RNA was detected in sewage sample by RT-qPCR using open reading frame 1ab (ORF1ab), nucleocapsid (N) protein, and spike (S) protein gene primers. The primers and probes for these genes were provided in the TaqPath™ RT-qPCR COVID-19 Kit (ThermoFisher, Waltham, MA). The specific sequences for these primers and probes have not been published and are held as business confidential information by ThermoFisher.

Quantification of gene copies/L was performed using the N gene carried on a plasmid (2019NCoV_N plasmid, Integrated DNA Technology, Coralville, IA). The RT-qPCR reaction as applied has a reliable (95% confidence) limit of detection of 10 gene copies per reaction. The RT-qPCR reaction was calibrated using a seven point calibration ranging from 10 to 100,000 copies per reaction with triplicate measurements at each point. The typical coefficients for the calibration curve, with acceptable ranges in parentheses, were a slope of -3.35 (-3.1 to -3.6), y-intercept of 40.38 (40 to 41) Cq, $R^2 = 0.99$ (>0.98), and an amplification efficiency of 90 to 110%. Using the typical values, the limit of detection at Cq = 40 was 720 gene copies/L of sewage.

Laboratory Controls

MS2 (ATCC® 15597-B1) bacteriophage RNA matrix spikes into the resuspended virus concentrate prior to the TRIzol extraction provide an internal positive control for each sample for the RNA isolation and extraction procedures. In addition to the SARS-CoV-2 genes, the ThermoFisher RT-qPCR

kit has a primer and probe for MS2. MS2 phage RNA added to the RT-qPCR reaction acted as a reverse transcription and PCR positive control. Additions of plasmids containing the SARS-CoV-2 N gene were used in low N gene sequence (1,000 copies per reaction) and high N gene sequence (10,000 copies per reaction) positive controls for the N gene primers. Purified SARS-CoV-2 RNA (TaqPath COVID-19 Control, ThermoFisher, Waltham, MA) was used as low concentration (50 copies per reaction) and high concentration (250 copies per reaction) RNA positive controls. A no template negative control (NTC) containing only nuclease free water was run with each batch of sewage samples. The NTC controlled for contamination during the RNA extraction and RT-qPCR reaction assembly. All samples and positive and negative controls were run in duplicate. Results were reported as gene copies/L of sewage.

Quality Assurance/Quality Control

In the event that the duplicate samples varied by more than one Cq, the internal MS2 RNA failed to amplify or the MS2 amplification occurred at > 30 Cq, a signal was detected in the NTC, the calibration check N-gene standards varied by more than $\pm$ 0.5 Cq, or the positive SARS-CoV-2 controls failed to amplify, corrective actions were taken to resolve the issue. Common remedies included (1) running two-fold dilutions of samples when the template appeared to have an inhibitor, i.e., the internal MS2 RNA failed to amplify correctly; (2) replacing the N-gene standard and rerunning the controls and possibly recalibrating the PCR machine and rerunning all samples associated with the failed calibration check standard if the N-gene control failed to amplify or did not achieve the expected Cq; (3) replacing the RT-qPCR reagents and rerunning all associated samples if the SARS-CoV-2 positive control failed to amplify; (4) or cleaning of all work surfaces and material handing equipment with dilute bleach solution if a detection was noted in the NTC.

SENSITIVITY ANALYSIS FOR THE SEIR MODEL PARAMETERS

The value of β used in the SEIR model was 0.20. β impacted the timing and magnitude of the peak of the epidemic but did not appreciably impact the relationship between active cases and mass rate of gene copies detected in wastewater; see Figure S1 below. Likewise, varying the mean for the maximum viral shedding rate (φ_{ij}) from 7 to 8, the mean for the viral shedding at 25 days (ψ_{ij}) from 3 to 4, and the time to maximum shedding from 4 to 6 only modestly impacted the SEIR model predictions (Supplementary Appendix Figures S2, S3, and S4, respectively).

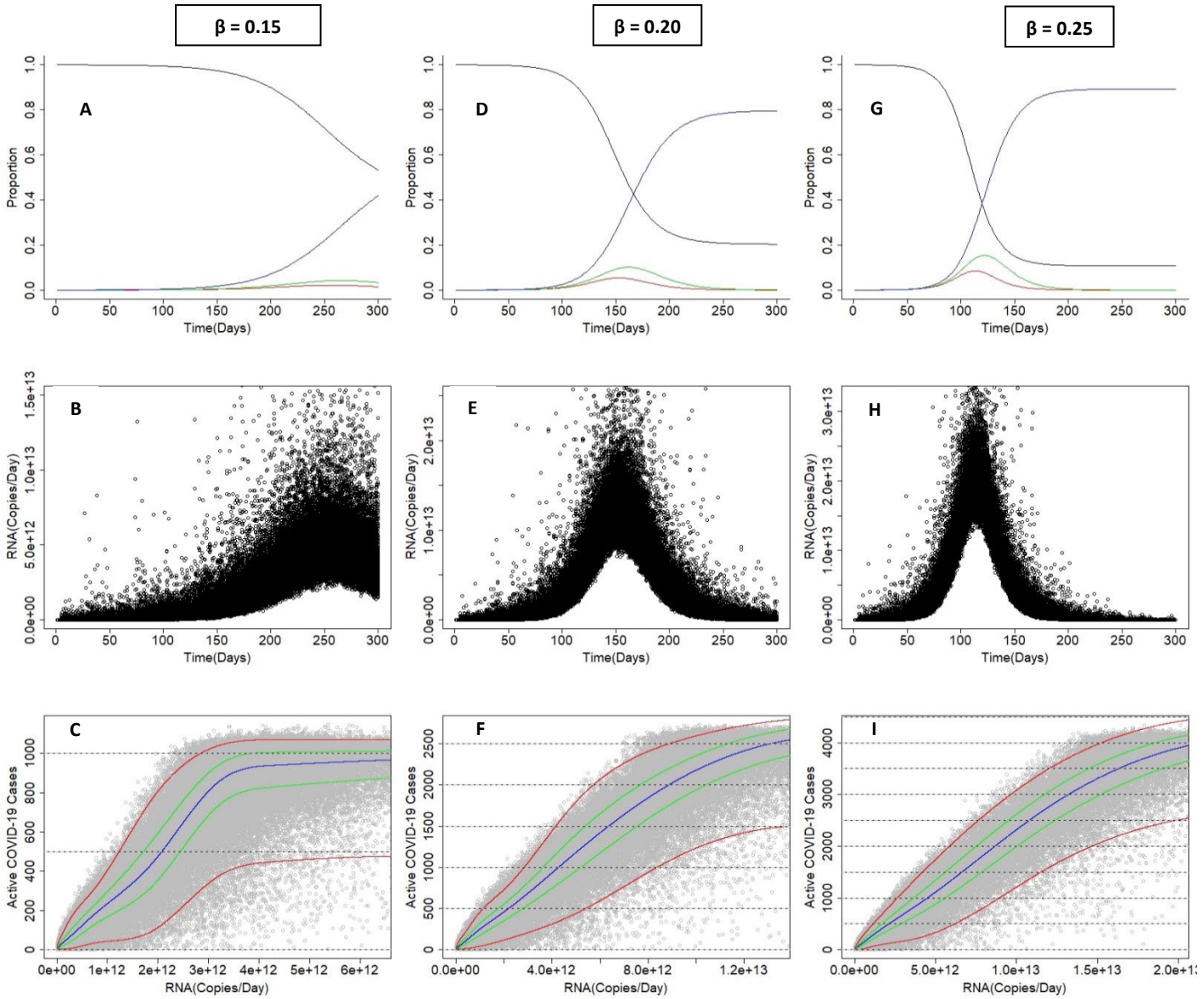

Figure S1. SEIR model for varying β values. Panels A, D, G: Proportions of the population that are susceptible (**black**), exposed (**red**), infectious (**green**), and recovered (**blue**). Panels B, E, H: Model predictions for mass rate of SARS-CoV-2 RNA in wastewater over time. Individual black points represent each stimulation. Panel C, F, I: Predictions of the number of infections versus RNA mass rate. Individual gray points represent each stimulation. The **blue** line represents the median, the **green** and **red** lines represent the 75% and 95% confidence intervals, respectively.

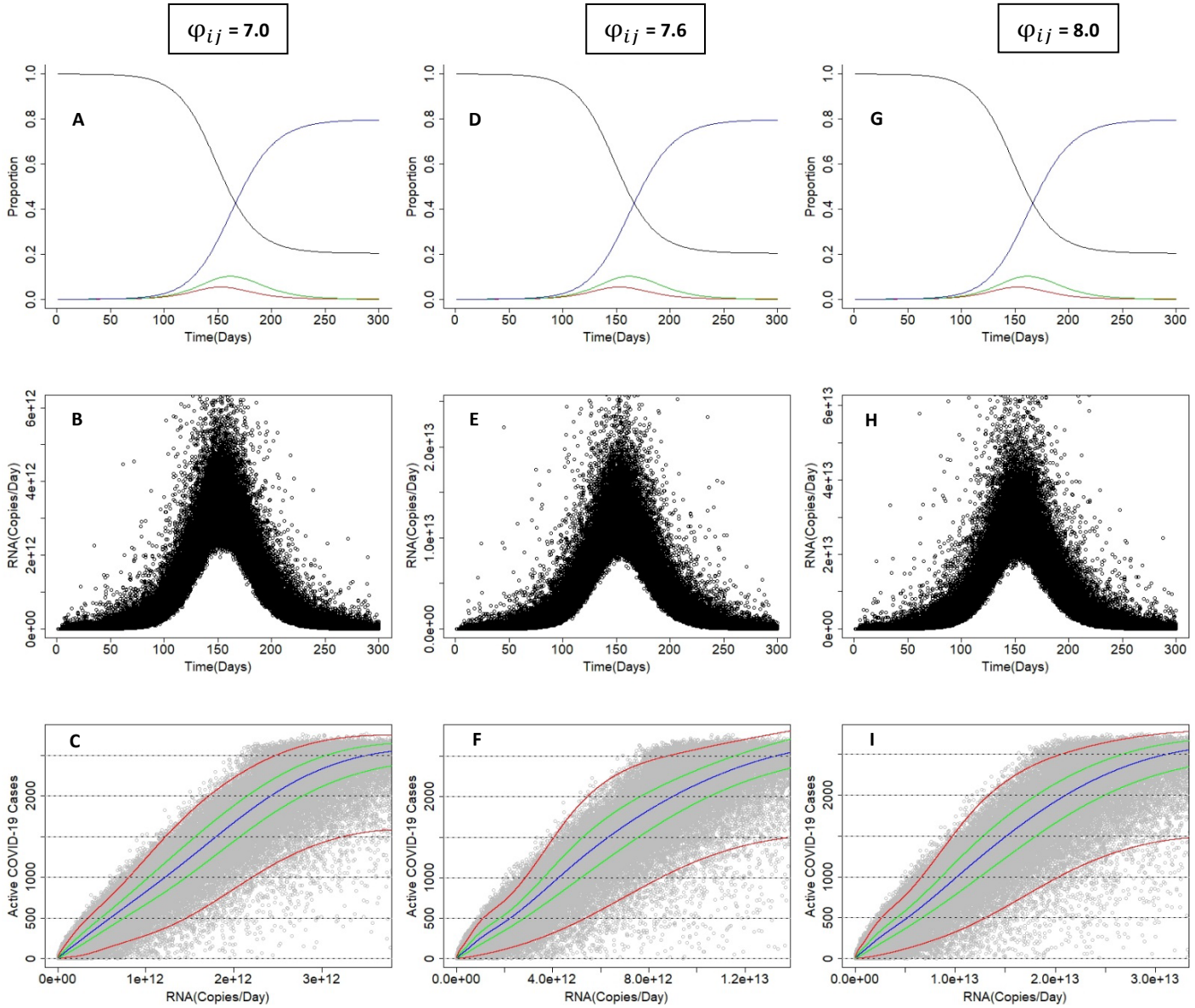

Figure S2. SEIR model for varying maximum viral shedding values. Panels A, D, G: Proportions of the population that are susceptible (**black**), exposed (**red**), infectious (**green**), and recovered (**blue**). Panels B, E, H: Model predictions for mass rate of SARS-CoV-2 RNA in wastewater over time. Individual black points represent each stimulation. Panel C, F, I: Predictions of the number of infections versus RNA mass rate. Individual gray points represent each stimulation. The **blue** line represents the median, the **green** and **red** lines represent the 75% and 95% confidence intervals, respectively.

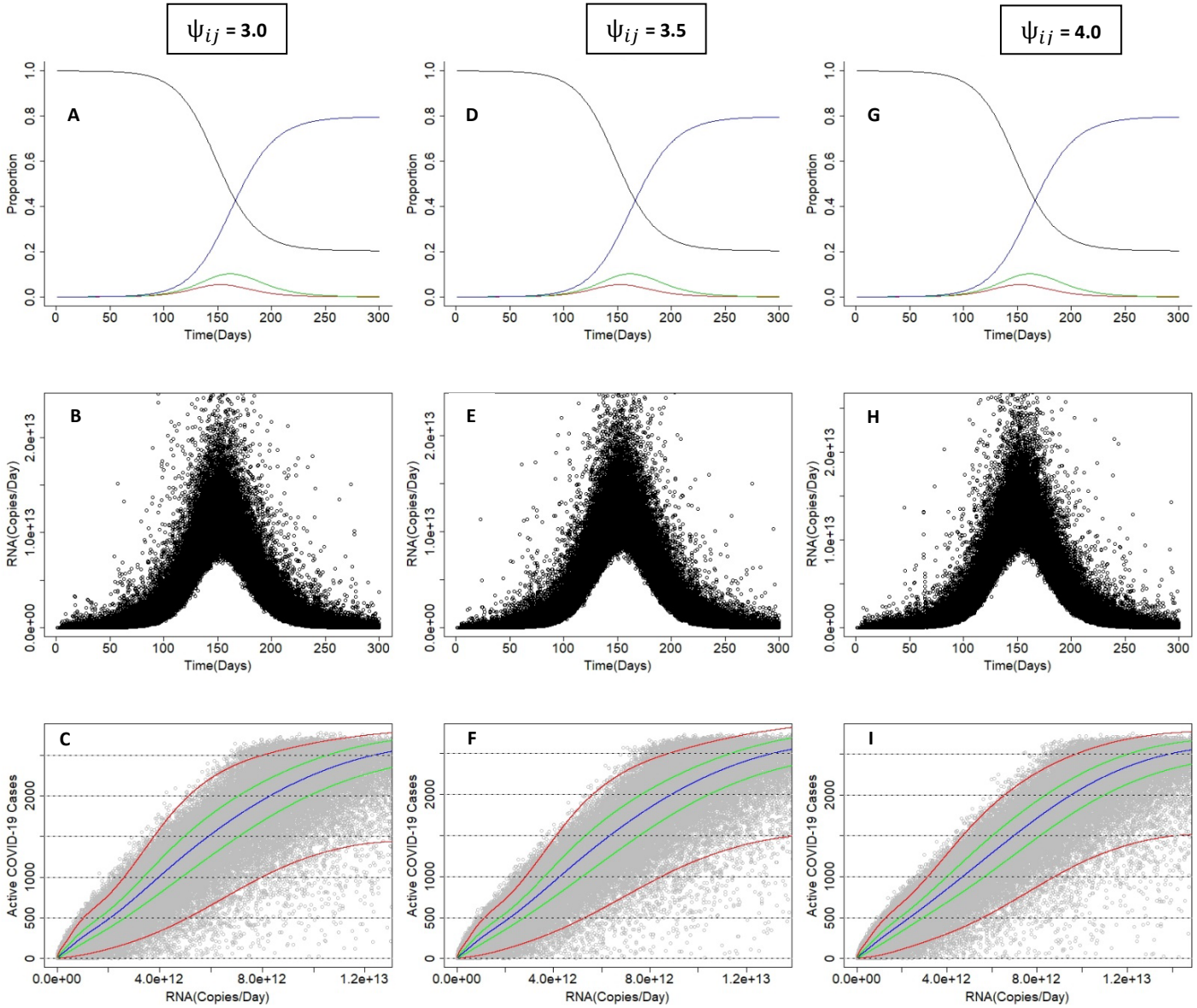

Figure S3. SEIR model for varying mean levels of viral shed at 20 days. Panels A, D, G: Proportions of the population that are susceptible (**black**), exposed (**red**), infectious (**green**), and recovered (**blue**). Panels B, E, H: Model predictions for mass rate of SARS-CoV-2 RNA in wastewater over time. Individual black points represent each stimulation. Panel C, F, I: Predictions of the number of infections versus RNA mass rate. Individual gray points represent each stimulation. The **blue** line represents the median, the **green** and **red** lines represent the 75% and 95% confidence intervals, respectively.

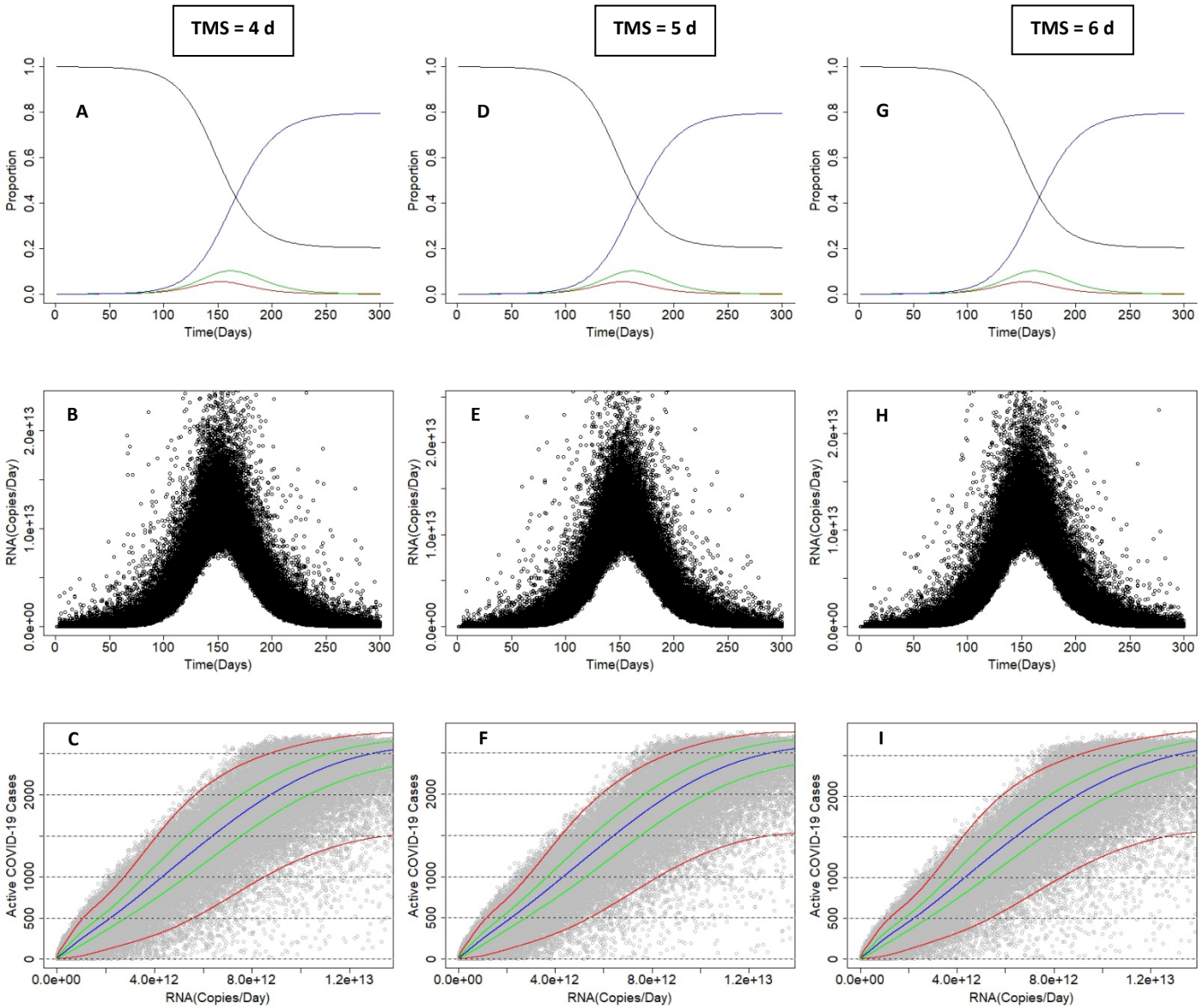

Figure S4. SEIR model for varying times to maximum shedding (TMS). Panels A, D, G: Proportions of the population that are susceptible (**black**), exposed (**red**), infectious (**green**), and recovered (**blue**). Panels B, E, H: Model predictions for mass rate of SARS-CoV-2 RNA in wastewater over time. Individual black points represent each stimulation. Panel C, F, I: Predictions of the number of infections versus RNA mass rate. Individual gray points represent each stimulation. The **blue** line represents the median, the **green** and **red** lines represent the 75% and 95% confidence intervals, respectively.

EXAMINING UNDERREPORTING WITH DATA FROM JUNE THROUGH DECEMBER 2020

Correcting reported case counts for bias to gain an understanding about the true COVID-19 infection burden within a study area is incredibly difficult if even possible at all. Bias in this venue can be introduced due to imperfect testing, lack of testing resources, testing practices, testing hesitancy, rate of asymptomatic infections within the target population, etc. Further, given the population under study, case reporting could be misallocated due to residency; i.e., student cases could be allocated to their county of residence, rather than the county in which Clemson University (CU) resides. Regional mitigation strategies (e.g., testing strategies) have the potential to impact the relation between reported and actual cases by identifying asymptomatic individuals. For example, CU implemented such a strategy which greatly expanded testing capacity and efforts during the Fall 2020 semester, leading to the detection of many asymptomatic infections that would have otherwise gone undetected. For this reason, we split our study period into two different time frames, time frame 1 (6/22/2020-9/1/2020) and time frame 2 (9/1/2020-11/30/2020). We then estimated a corrective factor within each. During the first time frame, CU did not mandate large scale testing, and this time frame should be representative of the general landscape of the state of South Carolina. Noting this allows us to compare our estimated corrective factor to that of Wu et al.[31], who estimated the presence of 15 infections for every confirmed case. During the second time frame, CU mandated large scale testing which led to the discovery and reporting of many asymptomatic cases that otherwise would have gone unreported, leading to a decrease in the corrective factor. During both time frames, the correction factor is determined as the slope estimate that arises from fitting a regression model to our predicted number of cases and the reported case counts. Table S1 summarizes the wastewater data collected during the second time frame for the three sewersheds surveilled in this study; the corresponding data for the first time frame are provided in Table 1 of the corresponding manuscript. Only results in Table S1 are shown for dates when data was available from both the Cochran Road and Pendleton/Clemson

WWTPs, to enable calculation of an RNA mass rate for the 29631 zip code.

Based on these data, the estimated corrective factor during the first time frame (10.86; Std. Error=3.40) agrees with the corrective factor estimated by Wu et al., which was 15. The estimated corrective factor during the second time frame (4.9; Std. Error=4.40) reflects the effect of detecting asymptomatic cases; i.e., expanded testing leads to fewer unreported cases. Figure S5 shows the relationship between the SEIR model predictions and reported cases corrected for underreporting for the second time frame; Figure 3 in the corresponding manuscript provides the same for the first time frame. In Figure S5, four of the points are identified as outliers. Point ① corresponds to 10/14/2020, which was the date of a home football game, when there may have been an influx of infected spectators to the community who went unreported and contributed to an elevated RNA mass rate (and hence prediction). Point ② corresponds to 9/7/2020, when undergraduate students began returning to dormitories. This point highlights a reporting issue at the onset of CU's testing efforts; i.e., the detected cases by CU (https://www.clemson.edu/covid-19/testing/dashboard.html) during this period, far exceeds the reported cases provided by SCDHEC. Points ③ (11/24/2020) and ④ (12/1/2020) correspond to the Thanksgiving holiday break, when an increase in testing (hence, more asymptomatic cases were discovered) was experienced nationwide.

Table S1. SARS-CoV-2 RNA levels in three adjoining sewersheds, fall 2020.

		CU WWTP			Cochran Road WWTP			Pendleton/Clemson WWTP			Zip Code 29631		
Date[a]	Rainfall (cm)	Flow Rate (10^6 L d^{-1})	RNA (copies L^{-1})	RNA Rate (10^{12} copies d^{-1})	Flow Rate (10^6 L d^{-1})	RNA (copies L^{-1})	RNA Rate (10^{12} copies d^{-1})	Flow Rate (10^6 L d^{-1})	RNA (copies L^{-1})	RNA Rate (10^{12} copies d^{-1})	Flow Rate (10^6 L/d)	RNA Rate (10^{12} copies d^{-1})	~Infected Individuals[c]
9/2/20	0.00	1.54	BLD	-	3.47	1.6E+05	0.56	3.75	9.9E+05	3.71	7.22	4.27	704
9/8/20	0.00	1.21	BLD	-	3.40	4.0E+06	13.60	4.17	2.5E+05	1.04	7.57	14.64	2,414
9/15/20	2.03	2.28	3.5E+05	0.7975	3.68	2.3E+06	8.46	4.19	2.4E+05	1.01	7.87	9.47	1,561
9/22/20	0.00	2.40	2.7E+05	0.6489	3.49	4.8E+05	1.67	4.60	1.3E+05	0.60	8.09	2.27	375
9/29/20	0.46	2.85	3.3E+06	9.4053	3.72	1.0E+06	3.72	4.10	5.9E+05	2.42	7.82	6.13	1,012
10/6/20	0.00	3.10	1.7E+06	5.2699	3.62	2.9E+05	1.05	4.46	1.2E+04	0.05	8.08	1.10	182
10/14/20	0.00	2.67	7.0E+05	1.8679	3.78	3.1E+06	11.72	5.70	6.2E+05	3.53	9.48	15.25	2,516
10/20/20	0.00	2.59	1.5E+06	3.8891	3.50	9.3E+05	3.26	4.86	2.2E+05	1.07	8.36	4.33	713
10/27/20	0.00	3.03	2.9E+06	8.7812	3.55	8.8E+05	3.12	4.95	1.2E+05	0.59	8.50	3.72	613
11/10/20	0.00	2.51	1.9E+05	0.4768	3.36	1.4E+05	0.47	4.64	8.9E+04	0.41	8.00	0.88	146
11/12/20	3.33	3.27	1.4E+05	0.4578	4.38	9.1E+04	0.40	8.79	6.2E+03	0.05	13.17	0.45	75
11/17/20	0.00	2.69	3.5E+05	0.9432	3.43	1.6E+05	0.55	5.17	2.1E+05	1.09	8.60	1.64	270
11/19/20	0.00	2.86	9.4E+04	0.2686	3.21	6.0E+04	0.19	5.18	1.4E+04	0.07	8.38	0.26	44
11/24/20	0.00	1.92	4.0E+05	0.7661	2.83	3.7E+04	0.10	5.07	6.8E+04	0.35	7.91	0.45	74
12/1/20	0.00	1.87	BLD	-	3.51	9.1E+04	0.32	8.80	BLD	-	-	0.32	53
12/3/20	0.00	1.85	BLD	-	2.88	1.0E+06	2.88	5.30	1.0E+05	0.53	8.19	3.41	563
12/8/20	0.00	1.60	3.5E+05	0.5590	2.81	1.1E+06	3.09	4.97	3.1E+05	1.54	7.78	4.63	764
12/10/20	0.00	1.65	1.6E+04	0.0264	2.93	5.5E+05	1.61	4.76	4.7E+03	0.02	7.69	1.63	269
12/17/20	0.76	1.89	BLD	-	2.75	1.1E+05	0.30	5.50	3.3E+04	0.18	8.25	0.48	80
12/22/20	0.00	1.37	BLD	-	2.15	1.9E+05	0.41	4.14	8.2E+05	3.39	6.29	3.80	627
12/24/20	1.27	1.39	BLD	-	2.24	2.8E+05	0.63	4.24	2.1E+05	0.89	6.48	1.52	250
12/29/20	0.00	1.44	BLD	-	2.15	3.8E+04	0.08	4.77	2.3E+05	1.10	6.92	1.18	194

[a]Data are shown only for dates when results were available for both Cochran Rd and Pendleton/Clemson, representing the 29631 zip code. Students returned to on-campus housing starting on 9/14/2020.

[b]BLD = below detection level.

[c]Calculated using equation 10.

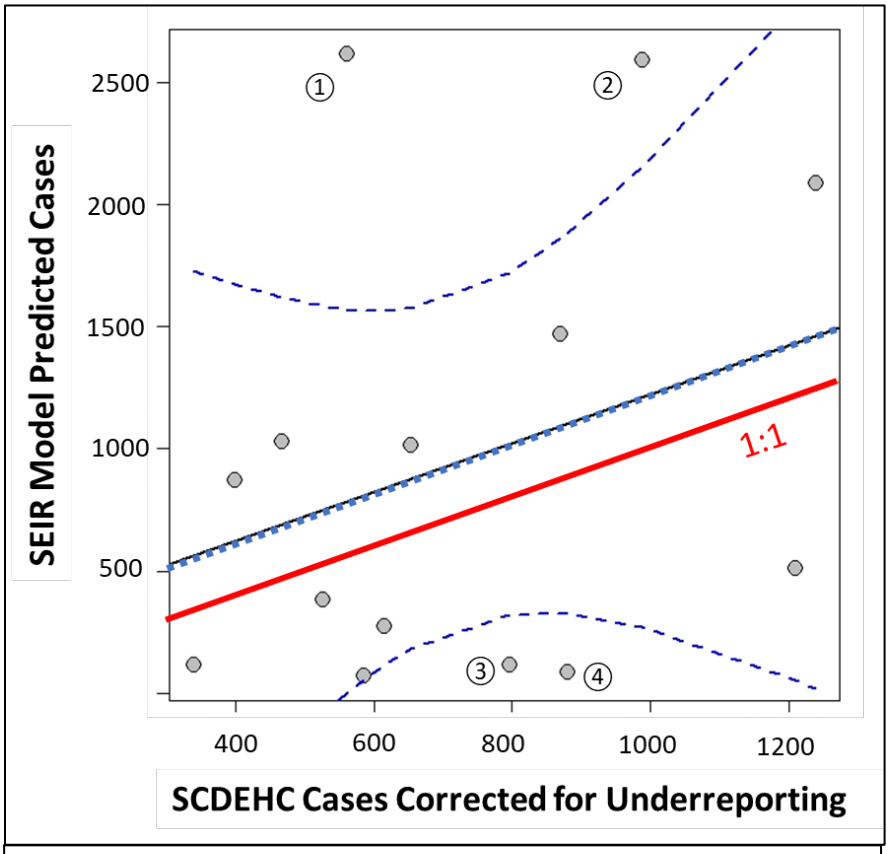

Figure S5. SEIR model predictions of active cases in the 29631 zip code based on RNA mass rates in wastewater versus SCDEHC cases corrected for underreporting based on an estimated ratio of 4.9. The dashed blue line represents the linear regression, bounded by the 95% confidence band. The red line represents a perfect match between the model and active cases. Outliers are identified as ①, ②, ③, and ④, with explanations provided in the text.

Reference

Killington, R. A., Stokes, A., & Hierholzer, J. C. (1996). Chapter 4. Virus purification. In *Virology Methods Manual* (pp. 71-89). Academic Press.