

Wastewater-based reproduction numbers and projections of COVID-19 cases in multiple cities in Japan, 2022

Shogo Miyazawa^{1†}, TingSam Wong^{2,3†}, Genta Ito¹, Ryo Iwamoto^{1,4}, Kozo Watanabe⁵, Michiel van Boven^{6,7}, Jacco Wallinga^{6,8}, Fuminari Miura^{5*}

1. Data Science Department, Shionogi & Co, Ltd, Osaka, Japan
2. AdvanSentinel Inc., Osaka, Japan
3. SHIMADZU Corporation., Kyoto, Japan.
4. Integrated Disease Care Division, Shionogi & Co, Ltd, Osaka, Japan
5. Center for Marine Environmental Studies (CMES), Ehime University, Ehime, Japan
6. Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands
7. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
8. Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands

† These authors contributed equally

* Correspondence author

Email: miura.fuminari.bt@ehime-u.ac.jp

ORCID: <https://orcid.org/0000-0001-8235-757X>

Abstract (237/250 words)

Background: Wastewater surveillance has expanded globally to monitor the spread of infectious diseases. An inherent challenge is substantial noise and bias in wastewater data due to their sampling and quantification process, leading to the limited applicability of wastewater surveillance as a monitoring tool and the difficulty.

Aim: In this study, we present an analytical framework for capturing the growth trend of circulating infections from wastewater data and conducting scenario analyses to guide policy decisions.

Methods: We developed a mathematical model for translating the observed SARS-CoV-2 viral load in wastewater into effective reproduction numbers. We used an extended Kalman filter to infer underlying transmissions by smoothing out observational noise. We also illustrated the impact of different countermeasures such as expanded vaccinations and non-pharmaceutical interventions on the projected number of cases using three study areas in Japan as an example.

Results: Our analyses showed an adequate fit to the data, regardless of study area and virus quantification method, and the estimated reproduction numbers derived from wastewater data were consistent with notification-based reproduction numbers. Our projections showed that a 10-20% increase in vaccination coverage or a 10% reduction in contact rate may suffice to initiate a declining trend in study areas.

Conclusion: Our study demonstrates how wastewater data can be used to track reproduction numbers and perform scenario modelling to inform policy decisions. The proposed framework complements conventional clinical surveillance, especially when reliable and timely epidemiological data are not available.

Keywords: Wastewater surveillance; wastewater based-epidemiology; scenario modelling; reproduction number; disease monitoring

Main text (3484/3500 words)

Introduction

The COVID-19 pandemic has presented a complex landscape for policymakers to navigate, with interfering factors such as vaccination, the emergence of new variants, and seasonality. Mathematical modelling has been employed by regional and national governments to monitor the disease in real-time, forecast epidemiological situations in near future (e.g., 1-2 weeks ahead), and inform policy decisions by projecting long-term trajectories under different scenarios [1,2]. Scenario modelling, exemplified by various research groups such as the COVID-19 scenario hub in the US and Europe [3,4], has contributed to more realistic and robust projections and a better understanding of epidemiological characteristics. However, accurate and standardized surveillance data are essential to capture temporal changes in disease dynamics, and it has become more challenging to obtain timely and unbiased epidemiological data via (passive) clinical surveillance due to changes in testing policies in many countries [5,6].

Wastewater surveillance has re-emerged as an alternative source of information during the COVID-19 pandemic [7,8]. Wastewater has the potential to monitor the prevalence by measuring virus concentrations excreted by infected individuals, which does not rely on patients' symptoms or medical-seeking behaviour [8,9]. The effectiveness of wastewater monitoring has been demonstrated for various infectious diseases in the past [8,10,11], and the COVID-19 pandemic has accelerated its establishment in many countries [7,12]. Nevertheless, a remaining challenge inherent to wastewater surveillance is the substantial bias and noise in observed data due to the factors related to the sampling and quantification processes (e.g., higher water demand during daytime, dilution due to rainfall, PCR inhibition). To mitigate such biases, new molecular tools and sampling techniques have been developed [13]. Nevertheless, there remains the intrinsic noise in the observation process, and thus extracting true signals of epidemic growth requires data-analytic methods that can disentangle underlying trends from noisy data.

Previous studies have attempted to deal with the noise in wastewater data by using statistical or machine-learning based approaches [14–17]. The strength of these methods lies in the functional flexibility of models, which allows for the smoothing of noisy data (e.g., penalized splines [16], neural network [14,17]). These studies primarily focused on short-term forecasting and were successful in providing near real-time estimates [15,16]. However, a drawback of non-mechanistic models is that they do not necessarily provide biological interpretations, and thus the outputs from such analyses are difficult to use for policy guidance with further scenario analysis.

Mechanistic models have been applied to wastewater data in recent studies, with the primary aim of evaluating the predictive ability of models [18,19] or monitoring growth trends by computing effective reproduction numbers [20,21]. Yet, another important component, scenario modelling, has not been thoroughly explored in the combination of wastewater surveillance. Synthesis of multiple data streams would enhance the robustness of scenario modelling, and more importantly, there is a practical need to inform policymakers of strategic planning of interventions even in the absence of timely and reliable epidemiological data. In the current near-endemic situation of COVID-19, evaluating the potential impact of additional interventions such as vaccination campaigns is one of the key questions, even though notified data are not always fully available [5,6]. To this end, we need to exploit wastewater data and incorporate current transmission mechanisms (e.g., repeated infections related to emerging variants and waning immunity), which have not been explicitly captured in previous work [18,19].

In this study, we develop a modelling approach that accounts for reinfection and vaccination effects and propose a way to infer transmission parameters from wastewater data and integrate them with the scenario modelling framework. As a motivating example, we conducted wastewater monitoring in

three municipalities in Japan between 2021-2022, where there was sufficient access to confirmation testing during the Omicron wave. We applied the proposed modelling approach to the collected data, first to evaluate its predictive performance and then to illustrate the impacts of different interventions (i.e., increasing vaccination coverages and reducing contact rates). Lastly, we discuss potential challenges that may arise when applying the proposed modelling framework to wastewater data for future implementations.

Methods

Wastewater data

We implemented wastewater surveillance in three study sites in Japan; Kyoto (population size: 778,000), Kanagawa (population size: 1,241,200), and City A (population size: 157,000). Wastewater samples were collected 2-3 times per week, and virus concentration in each sample was subsequently quantified with two different molecular methods, i.e. EPISSENS-S and COPMAN [22,23]. The details of sampling methods and experimental procedures are provided in the **Supplementary Materials**. We normalized the observed SARS-CoV-2 concentration by a commonly used faecal indicator, i.e., Pepper mild mottle virus (PMMoV), to adjust for potential bias caused by sampling time and flow rate of influent wastewater. When the measured concentrations were below detection limits, we imputed them as 1 (copy/L) for computational convenience. We then constructed the time series of the normalized SARS-CoV-2 concentration by taking the geometric mean, and the data was used for further analysis.

Epidemiological data

The number of daily confirmed cases within the same periods of wastewater was obtained from the corresponding local government websites (**Table S1**). As the coverage of the wastewater treatment plants does not always match the municipality areas, we calculated the daily number of cases in each catchment area by aggregating case data from multiple municipalities and weighting them by the proportion of the connected population size in each service area.

Transmission model

We developed a compartmental SEIRS model to incorporate reinfections and viral shedding from infected individuals to wastewater, adapting the method of [18]. The disease states are: susceptible $S(t)$, exposed but not yet infectious $E(t)$, infectious $I(t)$, and recovered $R(t)$ (see the conceptual model diagram in **Figure S1**). The model considered reinfections among individuals who have been infected already, by defining the average duration of immunity $1/\omega$ that was assumed to be 180 days (i.e., recovered transition back to the susceptible state at the rate of ω). We assumed fixed values for the mean latent period ($1/\alpha$) and infectious period ($1/\tau$) throughout the analysis. These parameters were defined based on the literature, and the details are provided in **Table S1** and **Table S2**. Two other parameters, the mean duration of virus shedding ($1/\gamma$) and the time-varying transmission rate ($\beta(t)$), were estimated by fitting the model to daily cases and/or wastewater data.

Model fit

To estimate parameters, we converted the above (deterministic) model to a stochastic model and used an extended Kalman filter for the inference task. The Kalman filter or its extended filtering methods have been often used for calibrating a dynamic model with epidemiological surveillance data such as the daily number of reported cases [24,25]. In this study, we employed the filtering method to fit the transmission model to the observed daily cases, virus concentrations in wastewater, or both, by incorporating the observation errors in both data sources. The model dynamics, including the active virus shedding state $A(t)$ and the (stochastic) observation errors, was described in the following system:

$$\begin{aligned} \frac{d}{dt}S(t) &= -\frac{\beta(t)S(t)I(t)}{N} + \omega R(t) - \sqrt{\frac{\beta(t)S(t)I(t)}{N}}w_1(t) + \sqrt{\omega R(t)}w_2(t) \\ \frac{d}{dt}E(t) &= \frac{\beta(t)S(t)I(t)}{N} - \alpha E(t) + \sqrt{\frac{\beta(t)S(t)I(t)}{N}}w_1(t) - \sqrt{\alpha E(t)}w_3(t) \\ \frac{d}{dt}I(t) &= \alpha E(t) - \tau I(t) + \sqrt{\alpha E(t)}w_3(t) - \sqrt{\tau I(t)}w_4(t) \\ \frac{d}{dt}R(t) &= \tau I(t) - \omega R(t) + \sqrt{\tau I(t)}w_4(t) - \sqrt{\omega R(t)}w_2(t) \\ \frac{d}{dt}A(t) &= \frac{\beta(t)S(t)I(t)}{N} - \gamma A(t) + \sqrt{\frac{\beta(t)S(t)I(t)}{N}}w_1(t) - \sqrt{\gamma A(t)}w_5(t) \end{aligned}$$

where the w_i are mutually independent white noise processes. We assume a closed population and thus $N = S(t) + E(t) + I(t) + R(t)$. The transitions here are assumed to follow a binomial process, and the binomial distribution is approximated by the normal distribution (see details in [18]). The model outputs were then compared to the observed data (i.e., case data or wastewater, or both). Firstly, we assumed that the number of daily confirmed cases $y_c(t)$ is a fraction of infected individuals who newly become symptomatic on the date of observation

$$y_c(t) = \mu_t \int_{t-1}^t \alpha E(s) ds$$

where μ_t is the reporting rate of newly confirmed cases out of the total number of infected cases on the observation day t . Since μ_t may change depending on the day of the week and the national holidays, the day-of-week effect were adjusted by estimating μ_t separately, and the holiday effect was further incorporated by reducing the reporting rate by 75% based on the observed maximum change in testing rates in Tokyo during December 2022 [26]. Secondly, the virus concentration in wastewater $y_w(t)$ is assumed to be proportional to the number of individuals shedding viruses $A(t)$:

$$y_w(t) = \nu A(t)$$

where ν is a scaling parameter specific to study regions.

Effective reproduction number

To quantify the growth trend of an epidemic, the (instantaneous) effective reproduction number [27,28], the number of secondary infections caused by a single infected person at time t , is calculated. In this study, the effective reproduction number is obtained by the following equation:

$$R_{eff}(t) = \frac{\beta(t) S(t)}{\tau N}$$

Where the transmission rate $\beta(t)$ is obtained by fitting the model to either notified case data or wastewater data. To distinguish two different reproduction numbers, hereafter we use notification-based reproduction numbers ($R_{eff}^N(t)$) and wastewater-based reproduction numbers ($R_{eff}^W(t)$) for further comparison. We computed the uncertainty in reproduction numbers by using estimates of $\beta(t)$ and its standard deviation (SD) and visualized the uncertainty ranges of two standard deviations. As a reference to standard practice, we used the EpiEstim package [29] to estimate effective reproduction numbers from notified case data, assuming a serial interval is gamma-distributed with a mean of 3.5 days and a standard deviation of 2.4 days [30]. The EpiEstim estimators were then compared to the values computed by our approach.

Forecasting and scenario projections

The model fitting via the Kalman filter allows an adaptive estimation of transmission rate $\beta(t)$ at each time point, and thus we sequentially updated the estimated parameters using the most recent data points. To perform one-week ahead forecasting, we simulated daily reported cases over the next seven

days using the most recent estimates of transmission rates and the number of individuals remaining in each state.

We examined two intervention scenarios; increasing vaccination coverage and reducing contact rates by non-pharmaceutical interventions (NPIs). Initial conditions for projections were determined using the estimated number of individuals in each state by fitting the model to the most recent observed data (**Table S1**). As a baseline scenario (i.e., a scenario without any additional intervention), we projected future cases in 4 months using the most recent estimate of the transmission rate β_0 .

In the scenario in which vaccination coverages are increased, the effect of additional vaccine uptakes was assumed to work as a transition from the susceptible to the recovered state (i.e., the vaccine mode of action was assumed to be “all-or-nothing” [31]). The transitioning proportion was calculated as $S(t)(c_{vac} - c_0)VE$, where $S(t)$ is the susceptible proportion, VE is the vaccine efficacy (assumed to be 60%), and c_0 and c_{vac} are the vaccination coverages before and after the additional vaccination. The baseline vaccination coverage c_0 was set as 70%, and we examined the expected impacts of increased coverage by varying c_{vac} as 80% and 90%. The effect of NPIs was modelled as a reduction in the contact rate, and thus the transmission rate after implementing NPIs was formulated as $\beta_{NPIs} = (1 - \phi)\beta_0$, where ϕ is the reduced ratio of contact rate compared to the baseline. In the main analysis the reduced ratio ϕ was set as 10%, and further reductions were examined in **Supplementary Text**. For both scenario analyses, we used the estimated baseline transmission rate β_0 and its 2SD ranges as the uncertainty ranges of projections.

Results

We implemented the wastewater surveillance system in three study sites in Japan. The observed wastewater data and the collected daily case data are illustrated in **Figure S2**. All data sources are provided in **Table S1**, and further details of two different RNA extraction/detection methods (i.e., COPMAN and EPISEN-S) are described in **Supplementary Text**. While there was a large degree of noise in individual observations of virus concentrations in wastewater, smoothed wastewater data indicated that growing and declining trends roughly matched with those observed in case data (**Figure S2**). This result was consistent regardless of the RNA extraction/detection method used or study areas.

The proposed modelling approach, using only wastewater data, described the epidemic trends in case data well at three study areas in Japan (**Figure 1**). Estimated parameters are listed in **Table S3**. In particular, the initial growth of epidemic waves in January 2022 was captured in a timely manner, which demonstrates the compatibility between notification-based and wastewater-based surveillance. Based on the estimated time-varying transmission rates and reporting rates, both total and reported cases were computed (shown in blue and red lines in **Figure 1**). The estimated ranges of reported cases were mostly consistent with the observed reported cases, and the large difference between the estimated total cases and the observed reported cases indicated that there may have been substantial under-reported cases around the peak of epidemic waves. By comparing different study areas, **Figure 1** illustrated that the uncertainty in estimates increased when the population size of the study area is small.

To further validate our findings, we compared two effective reproduction numbers, i.e. notification-based reproduction number R_{eff}^N and wastewater-based reproduction number R_{eff}^W (**Figure 2**, shown in blue and red). This analysis showed that the computed R_{eff}^N and R_{eff}^W were comparable throughout the study period, suggesting that our modelling approach using wastewater data can provide a reliable proxy for tracking epidemic trends. Besides, this was further supported by the result that the computed R_{eff}^N and R_{eff}^W with our approach were mostly consistent with the values of a standard EpiEstim approach (**Figure 2**, shown in green). In general, however, the estimated values of R_{eff}^W produced smoother curves with respect to time compared to the estimated R_{eff}^N . This indicated that the our approach with wastewater data alone may be less sensitive to abrupt changes in the epidemic, as the inherent noise in the data can hinder the identification of early signals.

We conducted one-week ahead predictions of reported cases under three different conditions: using wastewater data only, case data only, and both wastewater and case data. To account for variations in observation frequency (two or three times per week), we aggregated daily case data over one week and compared the model predictions to the observed weekly number of cases. The prediction accuracy was evaluated by two error metrics (the Root-Mean-Square Error (RMSE) and the Mean-Absolute-Error (MAE)). **Figure 3** shows the one-week ahead prediction of weekly cases across different study sites and RNA extraction/detection methods, and the examined three conditions induced equivalent prediction abilities (**Table S4**). Interestingly, the model using both case and wastewater data did not necessarily show the best prediction performance, despite the utilization of all available data for the prediction. This result suggests that although wastewater data is suitable for capturing an epidemic trend, it may not necessarily improve the accuracy of short-term predictions.

To demonstrate model-based projections, we visualized the potential impacts of two different strategies (i.e., increased vaccination coverages and NPIs) in **Figure 4** and **5**. The forward simulations indicated that both strategies would expedite the decrease in daily cases when compared to the baseline scenario that imposes no additional interventions. While the projected baseline trajectories suggested an overall decreasing trend (green line in **Figure 4** and **5**), the uncertainty intervals in two study sites (Kanagawa and City A) indicated a possible increase in daily cases (green-shaded regions in **Figure 4(A)(B)** and **5(A)(B)**). The same trend in the baseline scenario can be more clearly seen in the projected cumulative cases (**Figure S3** and **S4**). We also performed more stringent NPI scenarios (**Figure S5-S8**); those scenario analyses showed an increase in the vaccination by 10-20% or a reduction in the contact rate by about 10% could alter the upper bound of the projected incidence into a declining trend in our simulation settings. Among the study sites, the largest reduction in projected cases was seen in Kanagawa during January-April 2023 where the incidence of cases was the highest (**Figure 5(D)**).

Discussion

In this study, we showed that wastewater can capture the underlying trend of circulating SARS-CoV-2 infections and presented how scenario analyses can be provided to guide a policy decision by the proposed modelling framework. Our modelling approach translated the observed growth trend in wastewater data into effective reproduction numbers, which were consistent with estimated values by notified case data. As an application example, we further conducted scenario-based modelling analyses to illustrate the impact of different types of interventions on the projected number of cases. This highlighted the benefit of incorporating wastewater data into the current scenario modelling framework, regardless of the virus quantification method, especially when reliable epidemiological data are not obtainable.

The transmission model used in this study provided a good description of wastewater data (**Figure 1**). While previous literature in wastewater surveillance often claimed that machine-learning based models could capture more complex dynamics [14,17], our mechanistic model with parsimonious parameterizations yielded reasonable fitting results (**Figure 2** and **Table S4**). The main strength of our modelling approach is that all parameters have biological or epidemiological interpretations, and thus the outputs can be used for further scenario analysis. The interpretability and explainability are essential for informing policymaking as well as for (external) validity checks, in cases where there is a drastic change in transmission dynamics (e.g., the emergence of new variants).

Real-time monitoring of effective reproduction numbers via wastewater surveillance would be more effectively used if the notified case data suffer from substantial reporting delay or become less reliable. Effective reproduction numbers computed via case data are likely to capture the ‘true’ growth trend in an epidemic, regardless of the level of under-reporting, as long as the reporting rate is constant over the generation time (i.e., the average interval between the infections of an individual and its

secondary case). Our analysis showed that wastewater-based reproduction numbers were consistent with notification-based reproduction numbers (**Figure 2**), suggesting that our approach can effectively monitor the epidemic trend via wastewater surveillance. Various methods have been proposed to compute effective reproduction numbers [29,32,33], and their limitations are widely discussed [27,34]. A common challenge is that those methods are prone to sudden changes in reporting system (e.g., case definition, testing policy, diagnostic capacity, etc.). By contrast, wastewater surveillance is more robust to such transition in the data collection process, and recently several approaches have been proposed to estimate effective reproduction numbers via wastewater data [18,20,21]. We proposed to extend the applicable range of this wastewater-based framework; reproduction numbers estimated by mechanistic modelling approaches, such as ours, would provide a coherent way to simulate possible trajectories of an epidemic by varying other parameters when the epidemiological situation is changing. This usability is important for the iterative policy-making process.

Using Japan as an example, we presented analyses by examining the impact of different intervention scenarios based on the proposed approach with the observed wastewater data. Our model projections showed that, in the case of two study sites (City A and Kanagawa) where the daily incidence was increasing, a 10-20% increase in the vaccination coverage or about 10% reduction in the contact rate may be sufficient to turn the epidemic into a declining trend. These scenario analyses are useful to understand how much additional effort would be needed for controlling the disease on average. However, if more granular scenarios and strategic planning are required (e.g., targeted interventions by age, occupation, etc.), additional epidemiological data would be essential, as wastewater data only captures an aggregated trend over the whole population in the catchment areas. Thus, wastewater surveillance should be used as a complementary tool to support the current epidemiological surveillance.

The present study provided insights for further improvement in wastewater surveillance and its applicability to scenario modelling. Our analysis suggested that the estimated growth trends via wastewater data were more consistent with case data when the prevalence was high and/or the population size covered by the sewage system was large (**Figure 1 and 2**). Conversely, when the prevalence is low, virus concentrations in wastewater would also become low and approach the detection limit, leading to uncertain RNA quantifications with larger variations. Although the sensitivity of molecular methods has been extensively discussed [8,35], the minimization of variations in observations (e.g., experimental errors, variations in water sampling process, etc.) is also the key to capture the underlying epidemic trends. While it is possible to incorporate unobserved variations with various modelling approaches, such as the one proposed in this study, the implementation of experimental and sampling systems with reduced errors (e.g., flow-proportional composite water sampling [13]) would enhance the accuracy of wastewater surveillance and expedite more reliable scenario analysis.

Our scenario analysis should be interpreted with caution. First, our formulation simplified the dynamics, and consequently various pathogen/host factors (e.g., age-dependent contact rates, infectivity and immune-escape effect by variant, seasonality, etc.) were aggregated into the estimated parameters. In particular, the projected impacts of vaccination strategy may vary in practice, due to differences in the timing of vaccinations or differing waning rates by age. Although our aim was to illustrate the proposed framework by using collected data in Japan with minimal parameterization, the model assumptions and possible extensions in the structure, such as age stratification, need to be considered when more data are available. For the best practice of scenario modelling, we should always accommodate alternative candidate models and should not rely on a single model, and scenario analysis needs to be adaptively updated.

In conclusion, we have illustrated how wastewater data can be translated into intuitive epidemiological quantities such as total cases and reproduction numbers, and how we can use

wastewater data as an alternative source of information for scenario modelling to inform future policy. The proposed framework with wastewater surveillance complements and maximizes the benefit of clinical surveillance, especially when reliable and timely epidemiological data is not available.

Ethical statement

Ethical approval was not needed for this study. SARS-CoV-2 RNA concentration data in wastewater do not contain any privacy sensitive information.

Funding statement

The study was financed by the Netherlands Ministry of Health, Welfare and Sport. FM acknowledges funding from Japan Society for the Promotion of Science (JSPS KAKENHI, Grant Number 20J00793).

Data availability

Wastewater and aggregated case data and all codes used for analysis are available on GitHub (<https://github.com/AdvanSentinel/AS-SEIRS>).

Acknowledgements

The authors acknowledge Prof. Byung-Kwang Yoo and Prof. Ung-il Chung at Kanagawa University of Human Services for their support in obtaining permission for wastewater sampling. The authors thank the staff members of the WWTPs at Kanagawa Sewerage Works Foundation and the crisis management office at Kanagawa prefecture for their support and permission to sample the wastewater.

Conflict of Interest

SM, RI, GI, are employees of Shionogi & Co., Ltd. TSW is an employee of SHIMADZU Corporation. FM received research funding from AdvanSentinel Inc.

Authors' contributions

Conceptualisation: SM, FM. Data curation: SM, TSW, GI, RI, FM. Formal analysis: SM, TSW, GI, FM. Methodology: SM, TSM, GI, RI, MvB, JW, FM. Software: SM, TSM, GI. Validation: SM, TSW, GI. Visualisation: SM, TSW, GI. Writing – original draft: SM, TSW, RI, FM. Writing – review & editing: SM, TSW, GI, RI, KW, MvB, JW, FM.

References

- [1] Ainslie KEC, Backer JA, de Boer PT, van Hoek AJ, Klinkenberg D, Altes HK, et al. A scenario modelling analysis to anticipate the impact of COVID-19 vaccination in adolescents and children on disease outcomes in the Netherlands, summer 2021. *Eurosurveillance* 2022;27:2101090.
- [2] Keeling MJ, Dyson L, Tildesley MJ, Hill EM, Moore S. Comparison of the 2021 COVID-19 roadmap projections against public health data in England. *Nat Commun* 2022;13:4924.
- [3] COVID 19 scenario model hub n.d. <https://covid19scenariomodelinghub.org/> (accessed March 5, 2023).
- [4] European Covid-19 Scenario Hub n.d. <https://covid19scenariohub.eu/> (accessed March 5, 2023).
- [5] Leung K, Lau EHY, Wong CKH, Leung GM, Wu JT. Estimating the transmission dynamics of Omicron in Beijing, November to December 2022. *BioRxiv* 2022. <https://doi.org/10.1101/2022.12.15.22283522>.
- [6] Klous G, McDonald S, de Boer P, van Hoek AJ, Franz E, van Rooijen M. Staat van Infectieziekten in Nederland, 2021 2022. <https://doi.org/10.21945/RIVM-2022-0141>.
- [7] Naughton CC, Roman FA, Alvarado AGF, Tariqi AQ, Deeming MA, Kadonsky KF, et al. Show us the data: global COVID-19 wastewater monitoring efforts, equity, and gaps. *FEMS Microbes* 2023;4. <https://doi.org/10.1093/femsmc/xtad003>.
- [8] Hill V, Githinji G, Vogels CBF, Bento AI, Chaguza C, Carrington CVF, et al. Toward a global virus genomic surveillance network. *Cell Host Microbe* 2023. <https://doi.org/10.1016/j.chom.2023.03.003>.
- [9] Kitajima M, Ahmed W, Bibby K, Carducci A, Gerba CP, Hamilton KA, et al. SARS-CoV-2 in wastewater: State of the knowledge and research needs. *Sci Total Environ* 2020;739:139076.
- [10] Wolfe MK, Yu AT, Duong D, Rane MS, Hughes B, Chan-Herur V, et al. Use of Wastewater for Mpox Outbreak Surveillance in California. *N Engl J Med* 2023;388:570–2.
- [11] Ryerson AB, Lang D, Alazawi MA, Neyra M, Hill DT, St George K, et al. Wastewater Testing and Detection of Poliovirus Type 2 Genetically Linked to Virus Isolated from a Paralytic Polio Case - New York, March 9-October 11, 2022. *MMWR Morb Mortal Wkly Rep* 2022;71:1418–24.
- [12] Wastewater SPHERE n.d. <https://sphere.waterpathogens.org/> (accessed March 23, 2023).
- [13] Ahmed W, Bivins A, Bertsch PM, Bibby K, Choi PM, Farkas K, et al. Surveillance of SARS-CoV-2 RNA in wastewater: Methods optimisation and quality control are crucial for generating reliable public health information. *Curr Opin Environ Sci Health* 2020. <https://doi.org/10.1016/j.coesh.2020.09.003>.
- [14] Jiang G, Wu J, Weidhaas J, Li X, Chen Y, Mueller J, et al. Artificial neural network-based estimation of COVID-19 case numbers and effective reproduction rate using wastewater-based epidemiology. *Water Res* 2022;218:118451.
- [15] Morvan M, Jacomo AL, Souque C, Wade MJ, Hoffmann T, Pouwels K, et al. An analysis of 45 large-scale wastewater sites in England to estimate SARS-CoV-2 community prevalence. *Nat Commun* 2022;13:4313.
- [16] van Boven M, Hetebrij WA, Swart AM, Nagelkerke E, van der Beek RF, Stouten S, et al. Modelling patterns of SARS-CoV-2 circulation in the Netherlands, August 2020–February 2022, revealed by a nationwide sewage surveillance program. *MedRxiv* 2022. <https://doi.org/10.1101/2022.05.25.22275569>.
- [17] Li X, Kulandaivelu J, Zhang S, Shi J, Sivakumar M, Mueller J, et al. Data-driven estimation of COVID-19 community prevalence through wastewater-based epidemiology. *Sci Total Environ* 2021;789:147947.
- [18] Proverbio D, Kemp F, Magni S, Ogorzaly L, Cauchie H-M, Gonçalves J, et al. Model-based assessment of COVID-19 epidemic dynamics by wastewater analysis. *Sci Total Environ* 2022;827:154235.
- [19] Phan T, Brozak S, Pell B, Gitter A, Mena KD, Kuang Y, et al. A simple SEIR-V model to estimate COVID-19 prevalence and predict SARS-CoV-2 transmission using wastewater-based surveillance data. *MedRxiv* 2022. <https://doi.org/10.1101/2022.07.17.22277721>.

- [20] Nourbakhsh S, Fazil A, Li M, Mangat CS, Peterson SW, Daigle J, et al. A wastewater-based epidemic model for SARS-CoV-2 with application to three Canadian cities. *Epidemics* 2022;39:100560.
- [21] Huisman JS, Scire J, Caduff L, Fernandez-Cassi X, Ganesanandamoorthy P, Kull A, et al. Wastewater-based estimation of the effective reproductive number of SARS-CoV-2. *Environ Health Perspect* 2022;130:57011.
- [22] Ando H, Iwamoto R, Kobayashi H, Okabe S, Kitajima M. The Efficient and Practical virus Identification System with ENhanced Sensitivity for Solids (EPISSENS-S): A rapid and cost-effective SARS-CoV-2 RNA detection method for routine wastewater surveillance. *Sci Total Environ* 2022;843:157101.
- [23] Adachi Katayama Y, Hayase S, Ando Y, Kuroita T, Okada K, Iwamoto R, et al. COPMAN: A novel high-throughput and highly sensitive method to detect viral nucleic acids including SARS-CoV-2 RNA in wastewater. *Sci Total Environ* 2023;856:158966.
- [24] King AA, Ionides EL, Pascual M, Bouma MJ. Inapparent infections and cholera dynamics. *Nature* 2008;454:877–80.
- [25] Shaman J, Karspeck A, Yang W, Tamerius J, Lipsitch M. Real-time influenza forecasts during the 2012-2013 season. *Nat Commun* 2013;4:2837.
- [26] Tokyo Metropolitan Government. Tokyo Metropolitan Government COVID-19 Information Website. Tokyo Metropolitan Government COVID-19 Information Website n.d. <https://stopcovid19.metro.tokyo.lg.jp/en/monitoring> (accessed March 23, 2023).
- [27] Gostic KM, McGough L, Baskerville EB, Abbott S, Joshi K, Tedijanto C, et al. Practical considerations for measuring the effective reproductive number, Rt. *PLoS Comput Biol* 2020;16:e1008409.
- [28] Fraser C. Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS One* 2007;2:e758.
- [29] Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am J Epidemiol* 2013;178:1505–12.
- [30] Backer JA, Eggink D, Andeweg SP, Veldhuijzen IK, van Maarseveen N, Vermaas K, et al. Shorter serial intervals in SARS-CoV-2 cases with Omicron BA.1 variant compared with Delta variant, the Netherlands, 13 to 26 December 2021. *Eurosurveillance* 2022;27:2200042.
- [31] Halloran ME, Longini Jr. IM, Struchiner CJ. *Design and Analysis of Vaccine Studies*. Springer, New York, NY; 2010.
- [32] Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am J Epidemiol* 2004;160:509–16.
- [33] Parag KV. Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. *PLoS Comput Biol* 2021;17:e1009347.
- [34] Thompson RN, Stockwin JE, van Gaalen RD, Polonsky JA, Kamvar ZN, Demarsh PA, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics* 2019;29:100356.
- [35] Hewitt J, Trowsdale S, Armstrong BA, Chapman JR, Carter KM, Croucher DM, et al. Sensitivity of wastewater-based epidemiology for detection of SARS-CoV-2 RNA in a low prevalence setting. *Water Res* 2022;211:118032.

Figures and Tables

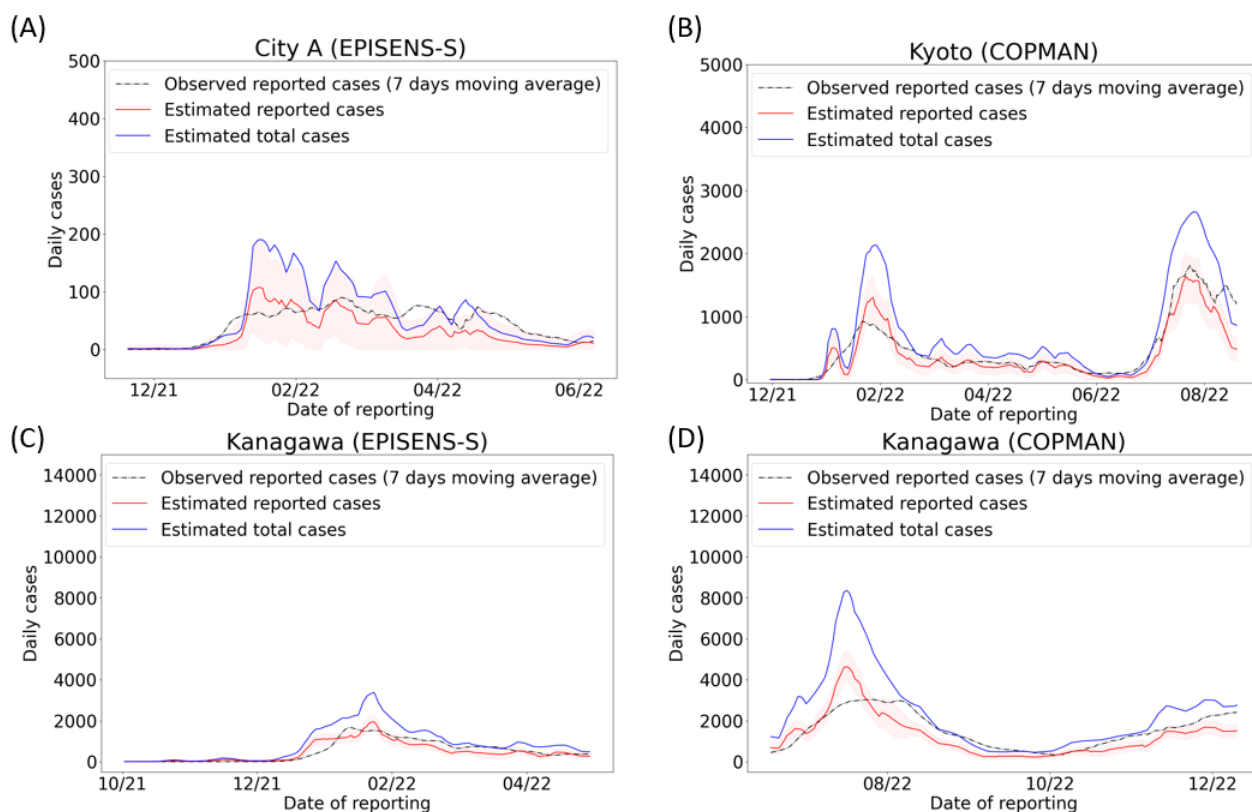


Figure 1. Estimated daily cases using only wastewater data in three areas in Japan. Black line indicates the observed daily reported cases, and red line and ribbon represent estimated daily reported cases with uncertainty bands of two standard deviations. Blue line corresponds to the estimated total cases, which was computed by incorporating under-reporting in the proposed modelling framework. COPMAN and EPISENS-S are different RNA extraction/detection methods, and the COPMAN method has a lower quantification limit of viral RNA.

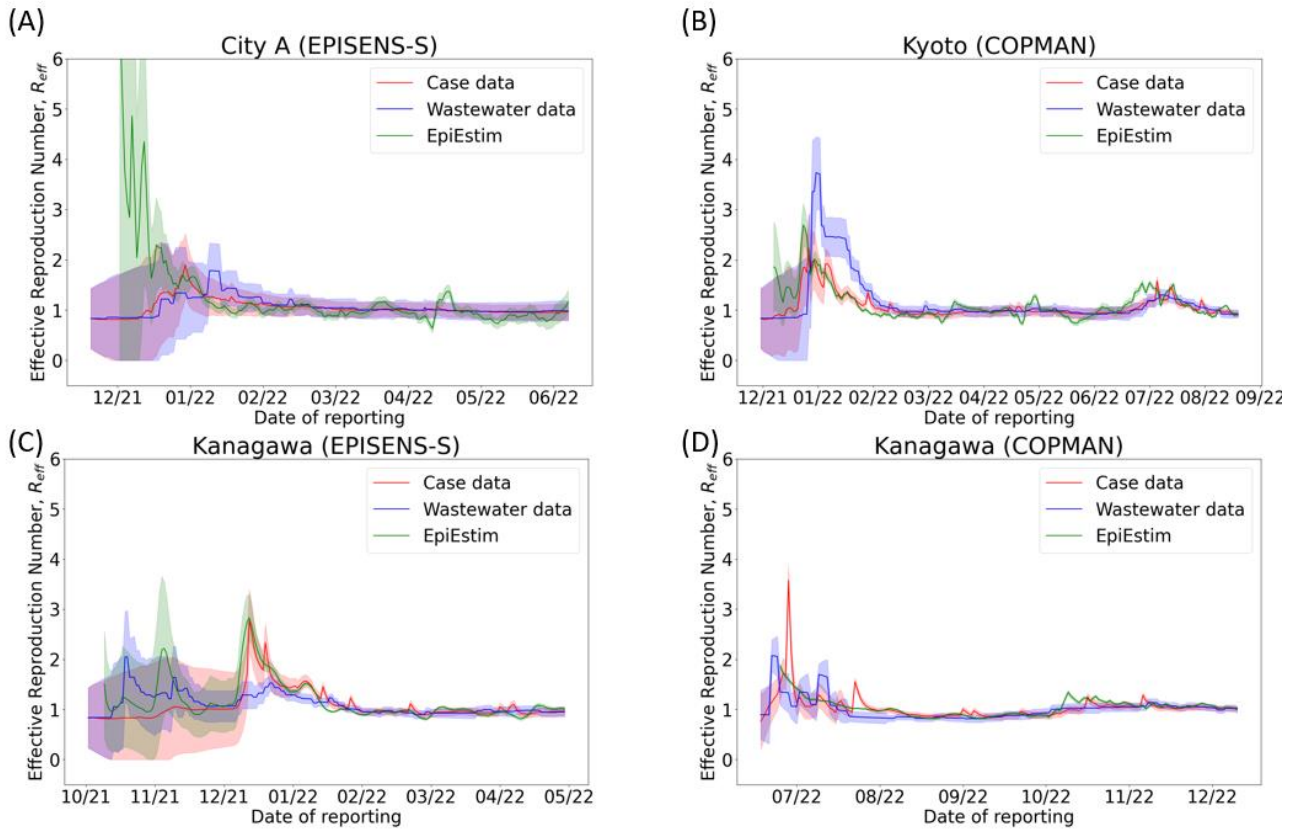


Figure 2. Estimated effective reproduction numbers using the proposed Kalman filter approach with only notified case data (red) and only wastewater data (blue), and using EpiEstim approach with only notified case data (green). Ribbons represent uncertainty ranges of 2 standard deviations (SD) for the proposed Kalman filter approach (red and blue) and 95% credible intervals for EpiEstim approach (green).

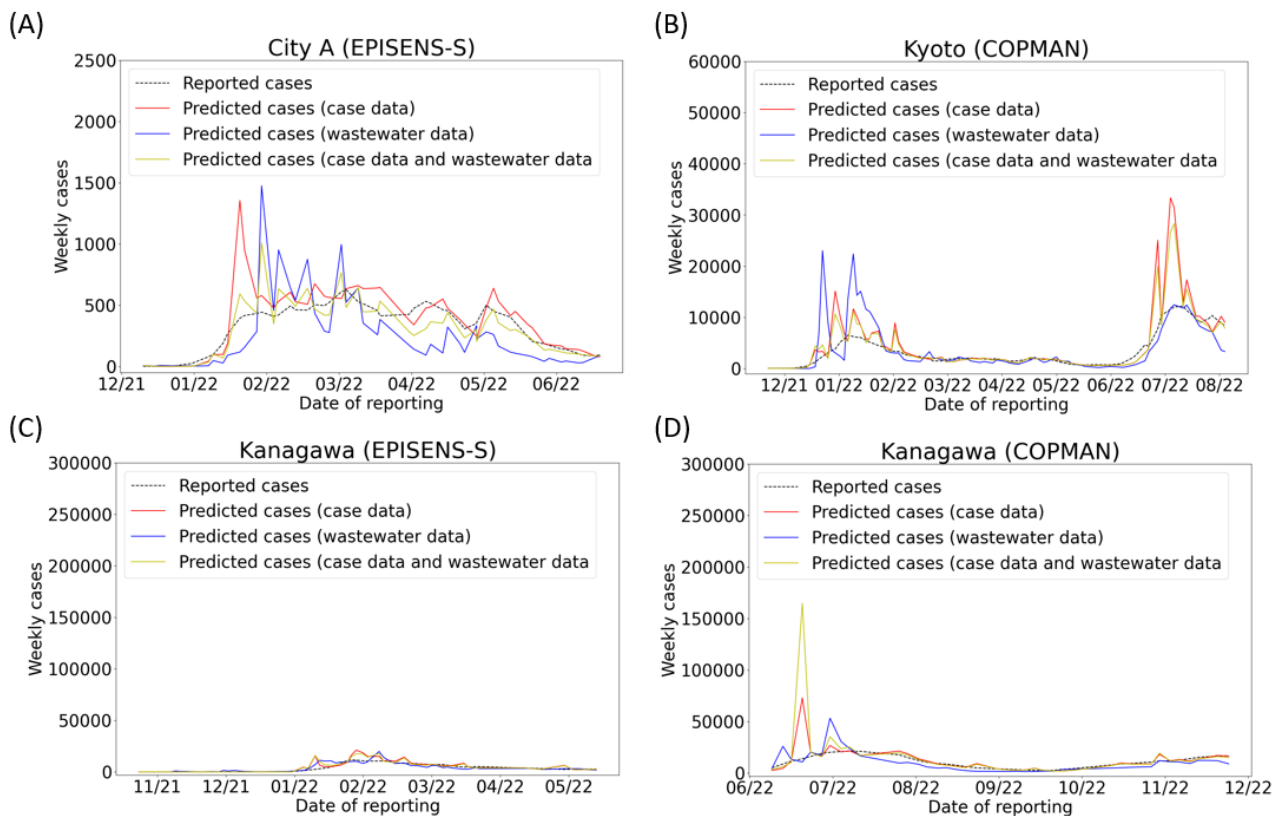


Figure 3. One-week ahead forecasting based on notified case data (red), wastewater data (blue), or both (yellow). The simulated number of cases over a week was adaptively updated via the Kalman filter and was compared to the observed reported cases (black).

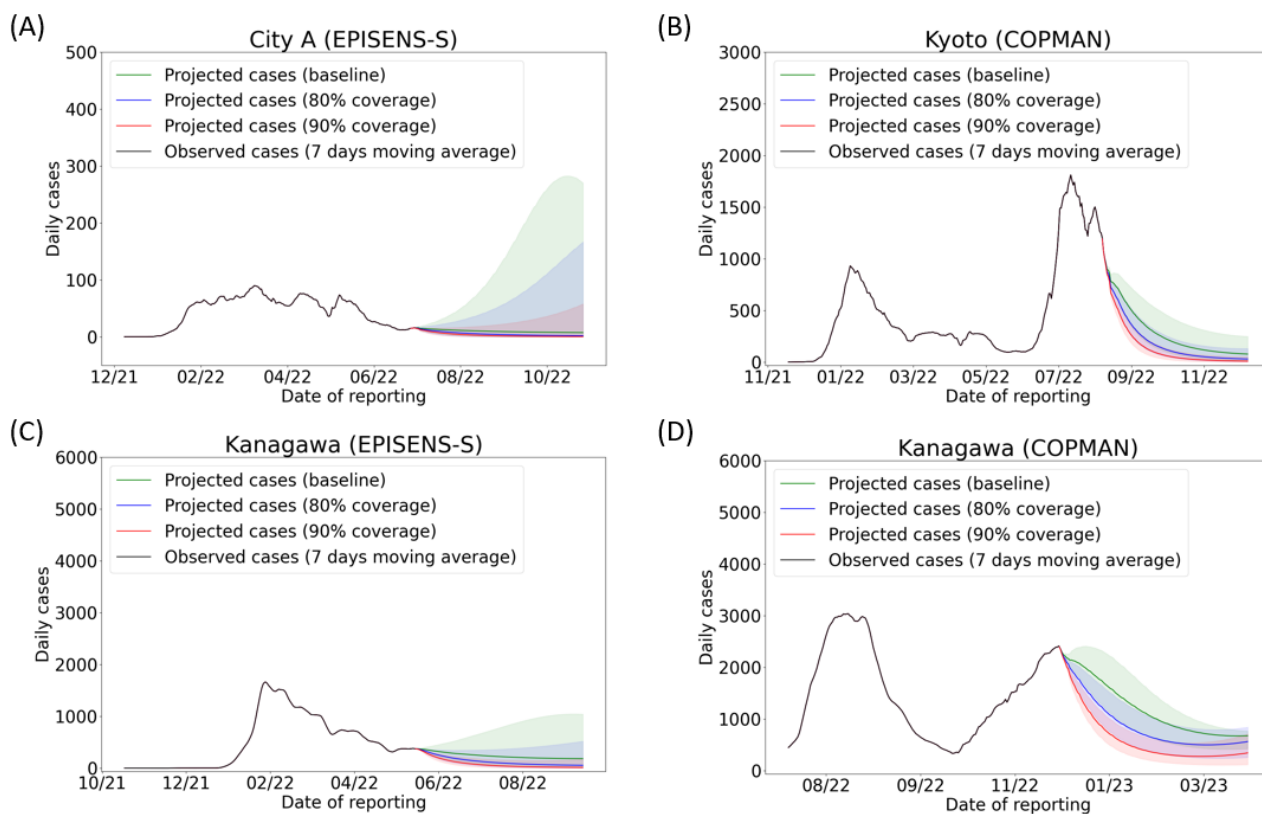


Figure 4. Model projected cases for increased vaccination scenarios. Vaccination coverages are set as 70% of the population for the baseline (green) and 80% (blue) and 90% (red) for scenarios with accelerated vaccine uptakes. Each ribbon represents uncertainty ranges of 2 standard deviations (SD) computed by the estimated variance of the baseline transmission rate.

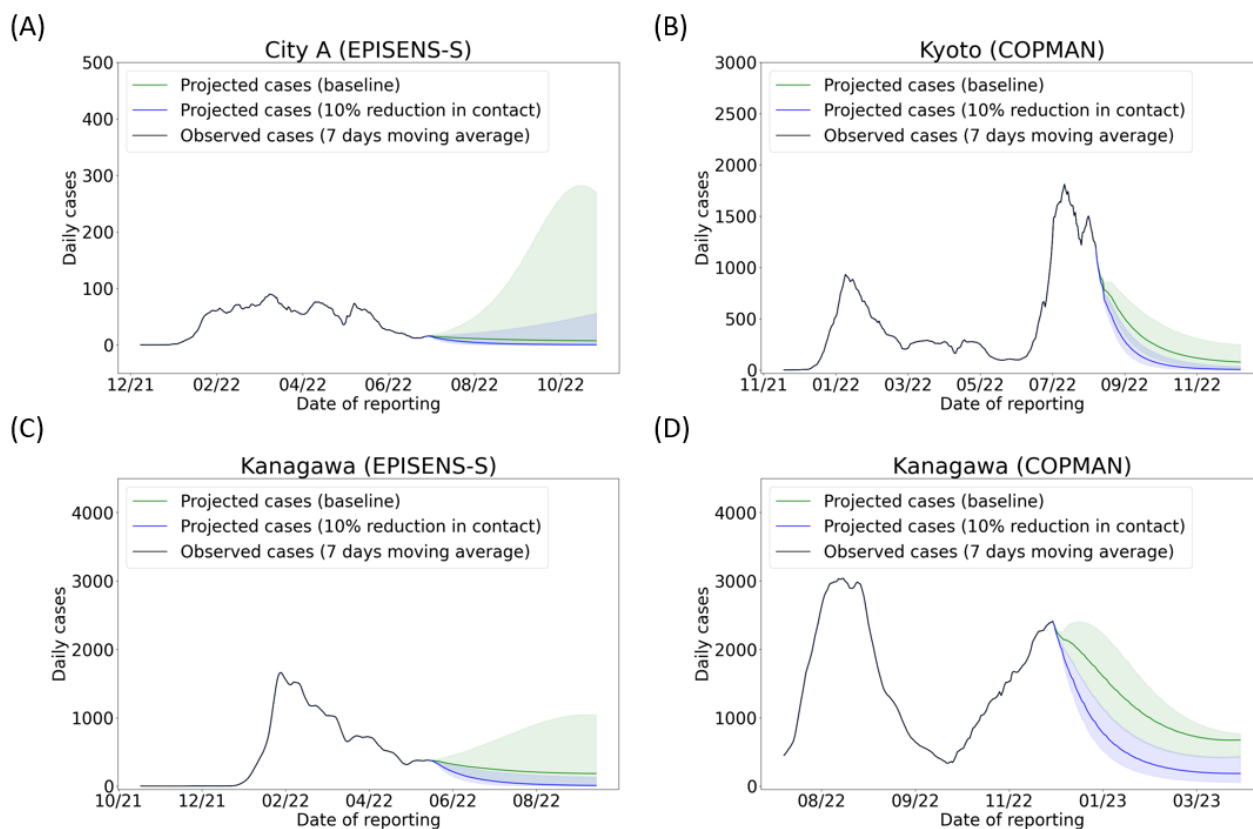


Figure 5. Model projected cases for a non-pharmaceutical intervention scenario. Relative contact rates are set as 1 for the baseline (green) and 0.9 (blue) and for scenarios with reduced contact rates. Each ribbon represents uncertainty ranges of 2 standard deviations (SD) computed by the estimated variance of the baseline transmission rate.

Supplementary Materials for

Wastewater-based reproduction numbers and projections of COVID-19 cases in multiple cities in Japan, 2022

Shogo Miyazawa, TingSam Wong, Genta Ito, Ryo Iwamoto, Kozo Watanabe, Michiel van Boven, Jacco Wallinga, Fuminari Miura.

Correspondence to: miura.fuminari.bt@ehime-u.ac.jp

Table of contents:

Supplementary Text

- Wastewater analysis
- Alternative vaccination scenario

Supplementary figures and tables

- Figure S1.** Schematic illustration of model structure
- Figure S2.** Collected wastewater and case data
- Figure S3.** Projected cumulative cases in vaccination scenarios
- Figure S4.** Projected cumulative cases in non-pharmaceutical intervention scenarios
- Figure S5.** Additional analysis for more stringent NPIs scenario (15% reduction in contact)
- Figure S6.** Additional analysis for more stringent NPIs scenario (20% reduction in contact)
- Figure S7.** Additional analysis for more stringent NPIs scenario (25% reduction in contact)
- Figure S8.** Additional analysis for more stringent NPIs scenario (30% reduction in contact)

- Table S1.** Data source

- Table S2.** List of input parameters

- Table S3.** List of estimated parameters

- Table S4.** Summary statistics for model fits

Supplementary Text

Wastewater analysis

Influent wastewater samples were collected from wastewater treatment plants in three study areas in Japan (see **Methods** and **Table S1**). The wastewater samples were collected in sterile plastic bottles via grab sampling and immediately transported to the laboratory. The samples were processed with the concentration method described below on the day of sample collection.

The collected wastewater samples were analyzed by EPISENS-S [22] and COPMAN [23] to quantify the SARS-CoV-2 and PMMoV RNA concentration in wastewater. Briefly, for EPISENS-S, total RNA was extracted from suspended solid formed via low-speed centrifugation at 3,000 *g* for 10 mins in the 40 mL sewage samples, using the RNeasy PowerMicrobiome kit on a QIAcube system (Qiagen, Hilden, Germany) to obtain a final RNA extracted volume of 50 μ L. One-step RT-preamplification was performed using 13.5 μ L of the RNA extract and the CDC N1 forward and reverse (2019-nCoV_N1-F and 2019-nCoV_N1-R) and PMMoV reverse (PMMV_RP1) primers (Table S1, S2 in the Supplementary Material of [22]) with the iScriptTM Explore One-Step RT and PreAmp Kit (Bio-Rad Laboratories, Hercules, CA, USA). Thermal cycling conditions for RT-preamplification were as follows; 25°C for 5 mins, 45°C for 60 mins, 95°C for 3 mins followed by 10 cycles of 95°C for 15 s and 55°C for 4 mins. qPCR was performed with QuantiTect Probe PCR Master Mix (Qiagen, Hilden, Germany) in a total reaction volume of 25 μ L containing 2.5 μ L of pre-amplified products, which was performed using the primers and probe for SARS-CoV-2 (CDC N1) or PMMoV with a final concentration of 400 nM and 300 nM each, respectively (Table S1 and S2 in the Supplementary Material of [22]). Thermal cycling conditions for qPCR were as follows; 50°C for 2 mins, 95°C for 10 mins followed by 45 cycles of 95°C for 3 s and 55°C for 32 s. qPCR reactions were completed on ABI 7500 Real-Time qPCR system (Applied Biosystems), and the threshold value of relative fluorescent intensity (Δ Rn) was adjusted to be 0.2.

For COPMAN, viruses were coagulated with the addition of 1 μ L of polyaluminum chloride (PAC) followed by vigorous shaking for 30 times, and subsequent gentle shaking at 80–120 rpm for 10 min at 4 °C. The samples were then centrifuged at 3000 \times g for 10 min, and the supernatant was discarded. The samples were centrifuged again at 3000 \times g for 3 min and the remaining liquids were removed by pipetting. The debris was then transferred to a 1.5-mL tube and lysed with 250- μ L SDS-based lysis buffer and digested by 14.25- μ L proteinase K solution at 56 °C for 10 min. Crude RNA of 200 μ L was extracted from the samples with phenol/chloroform/isoamyl alcohol (25:24:1), which was then purified with carboxyl-modified magnetic beads, to obtain a final RNA extract volume of 50 μ L. An aliquot (2 μ g or 13 μ L) of the magnetic bead-purified total RNA was subjected to cDNA synthesis using the Reliance Select cDNA synthesis kit (Bio-Rad Laboratories) under the following conditions: 50 °C for 60 min, 95 °C for 1 min in 20- μ L reaction mix with 2 pmol each of reverse primers of SARS-CoV-2, RSV, and PMMoV. The resultant cDNAs of SARS-CoV-2 and RSV were pre-amplified for 10 cycles by the Biotaq HS (Bioline Reagents Ltd., London, UK) under the following conditions: 95 °C for 10 min, and 10 cycles of 95 °C for 15 s, 55 °C for 15 s, and 72 °C for 30 s, in 30- μ L volume reaction mix containing 9 pmol each of forward and reverse primers. PMMoV cDNA was not preamplified because PMMoV RNA usually exists in wastewater with high amounts. Finally, viral RNA was quantified from 2.5 μ L of the preamp product for SARS-CoV-2 and RSV, and 2.5 μ L of cDNA for PMMoV by qPCR using the TaqMan Environmental Master Mix 2.0 (Thermo Fisher Scientific) under the following conditions: 95 °C for 10 min, and 45 cycles of 95 °C for 15 s and 60 °C for 30 s, in 20- μ L singleplex reaction mix containing 10 pmol each of reverse and forward primers and 7.5 pmol of TaqMan probe.

Alternative vaccination scenario

In additional analysis where vaccination coverages are increased, the effect of vaccines was assumed to work as a proportional reduction in transmission rates (i.e., the vaccine mode of action was assumed to be “leaky” [31]), and the transmission rate after additional vaccination is:

$$\beta_{vac} = \frac{(1 - VE * c_{vac})}{(1 - VE * c_0)} \beta_0$$

where *VE* is the vaccine efficacy (assumed to be 60%) and *c*₀ and *c*_{vac} are the vaccination coverages before and after the additional vaccination, respectively. The baseline vaccination coverage *c*₀ was set as 70%, and we examined the expected impacts of increased coverage by varying *c*_{vac} as 0.8 and 0.9. The results of projections are shown in **Figure-S3**.

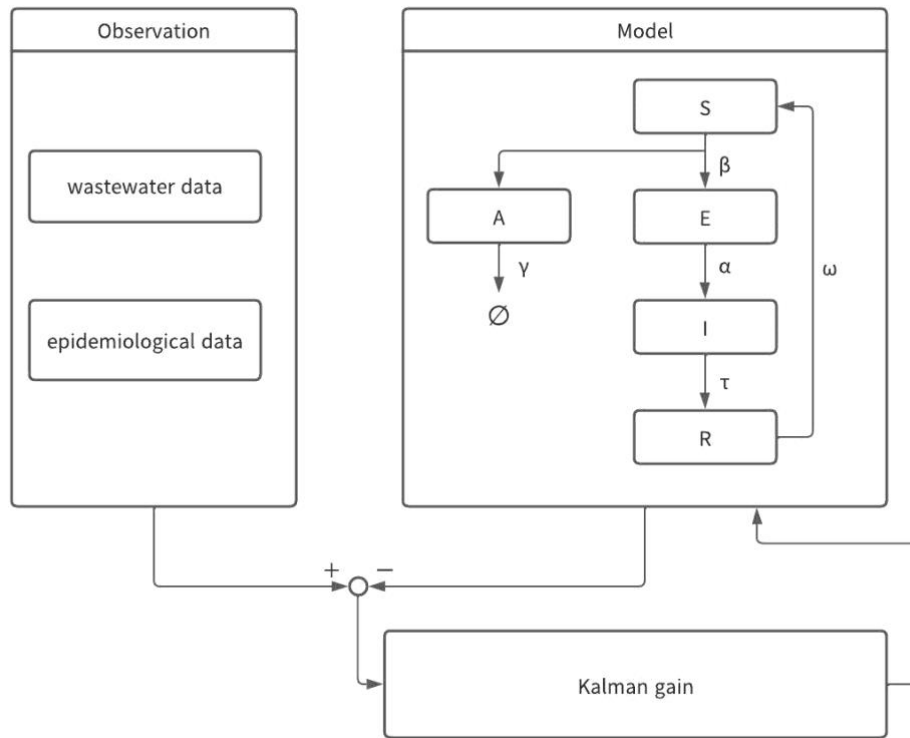


Figure S1. Schematic illustration of model structure.

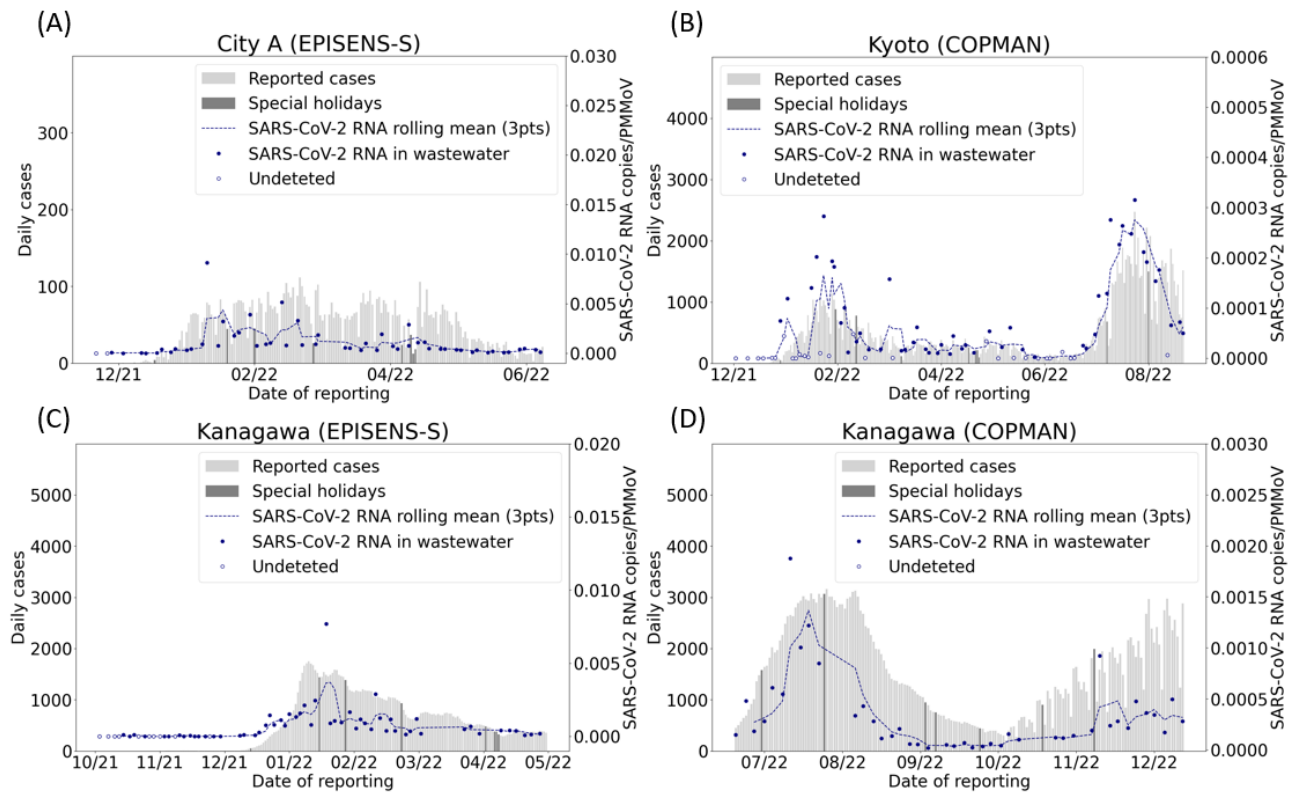


Figure S2. Collected wastewater and daily confirmed case data for different sites. Bars represent the daily number of reported cases, and darker bars indicate the cases reported on special holidays. Dot plots are SARS-CoV-2 RNA concentrations, and its rolling mean over three observed data points is shown as a blue line. White dots represent data points below the detection limit of RNA quantification methods.

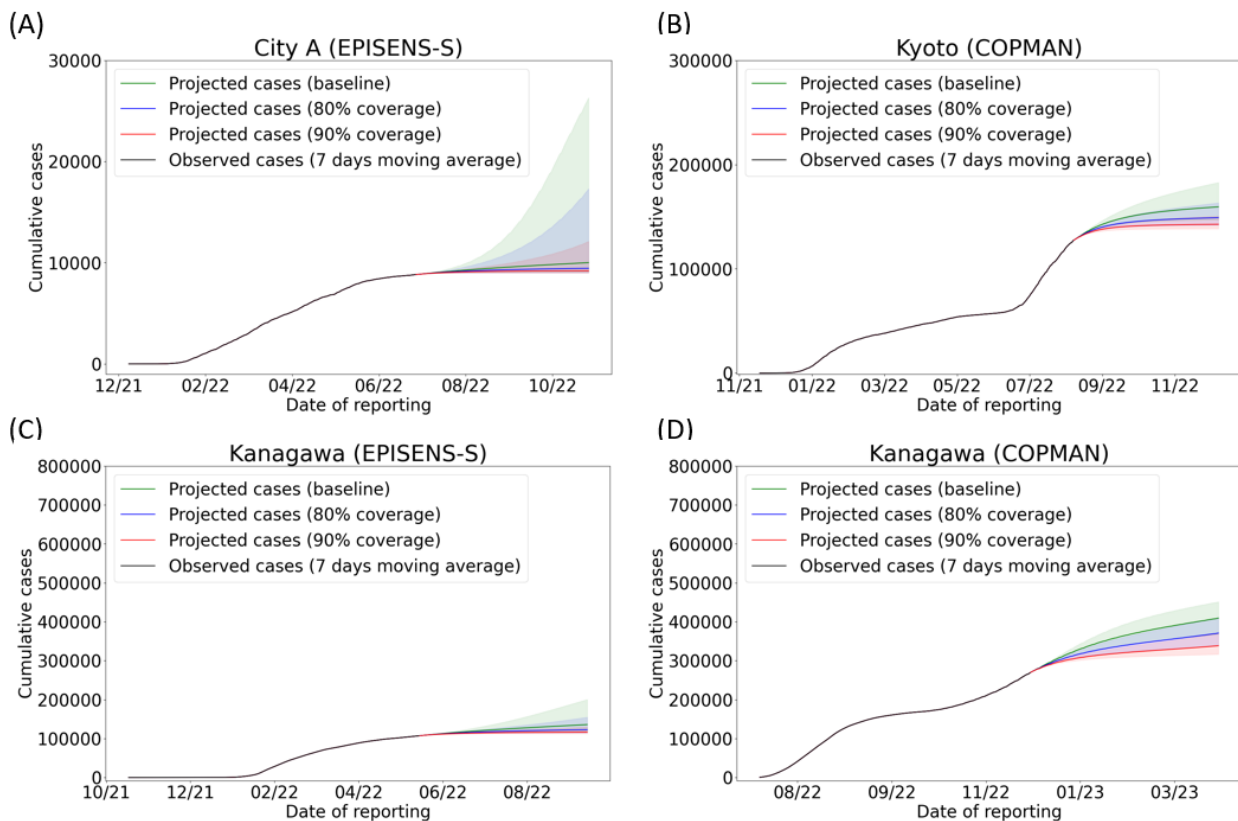


Figure S3. Projected cumulative cases in vaccination scenarios. Vaccine effect is assumed to reduce the baseline transmission rate proportionally (i.e., “leaky” effect). Vaccination coverages are set as 70% of the population for the baseline (green) and 80% (blue) and 90% (red) for scenarios with accelerated vaccine uptakes. Each ribbon represents uncertainty ranges of 2 standard deviations (SD) computed by the estimated variance of the baseline transmission rate.

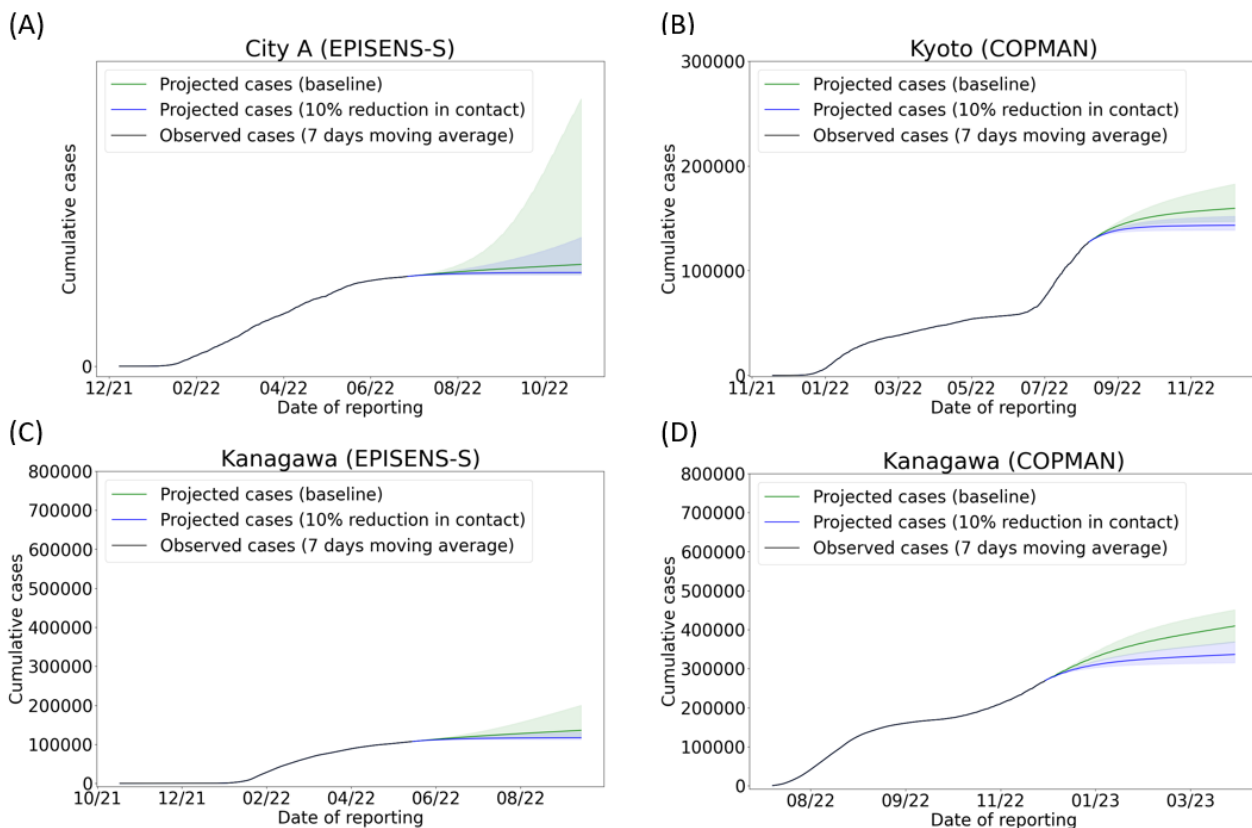


Figure S4. Projected cumulative cases in non-pharmaceutical intervention scenarios. Relative contact rates are set as 1 for the baseline (green) and 0.9 (blue) and for scenarios with reduced contact rates. Each ribbon represents uncertainty ranges of 2 standard deviations (SD) computed by the estimated variance of the baseline transmission rate.

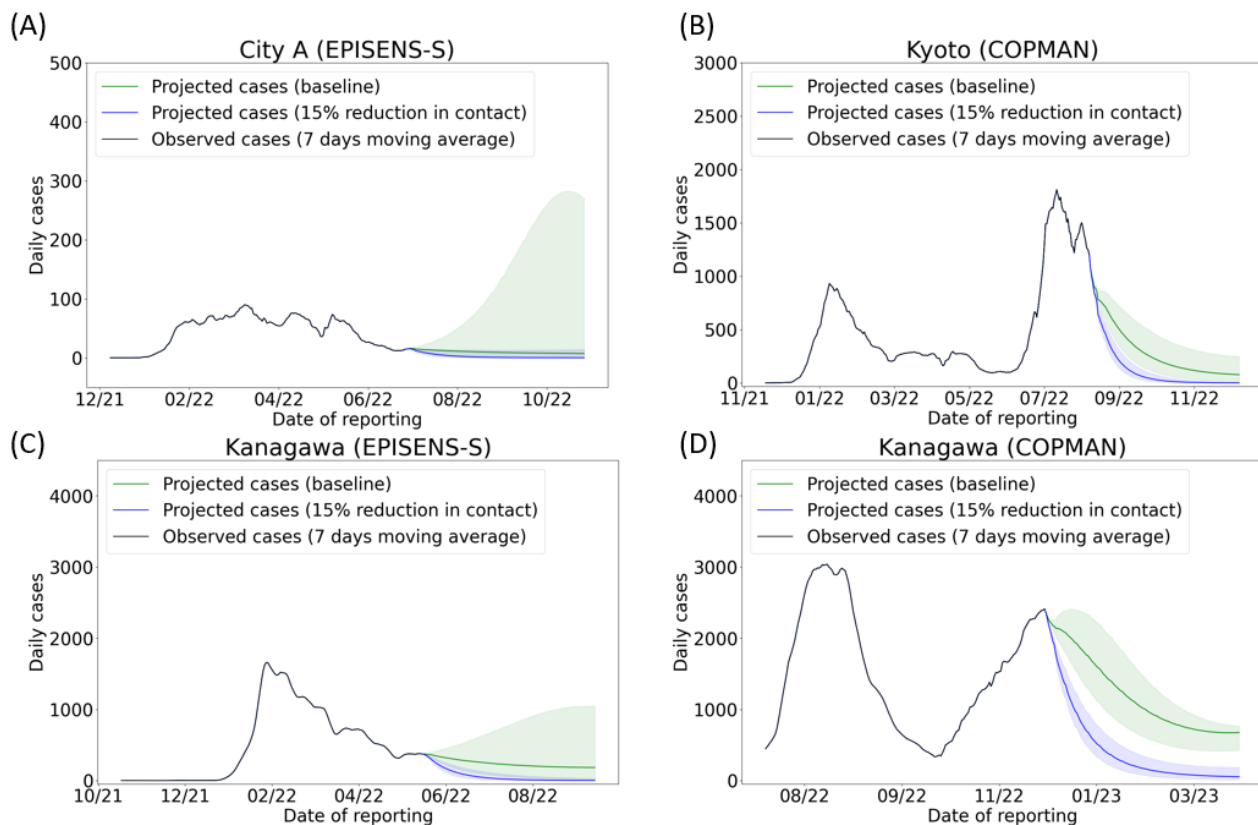


Figure S5. Additional analysis for a non-pharmaceutical intervention scenario. Relative contact rates are set as 1 for the baseline (green) and 0.85 (blue) and for scenarios with reduced contact rates. Each ribbon represents uncertainty ranges of 2 standard deviations (SD) computed by the estimated variance of the baseline transmission rate.

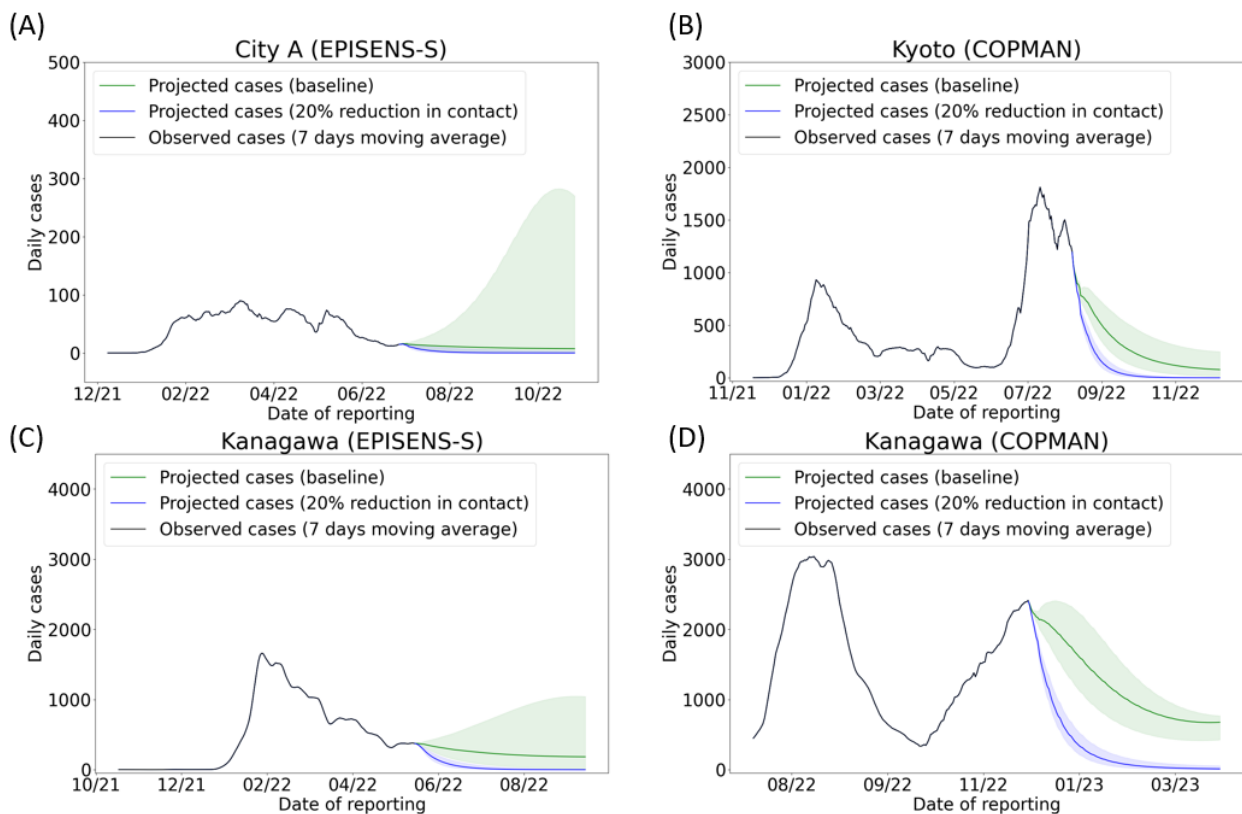


Figure S6. Model projected cases for a non-pharmaceutical intervention scenario. Relative contact rates are set as 1 for the baseline (green) and 0.8 (blue) and for scenarios with reduced contact rates. Each ribbon represents uncertainty ranges of 2 standard deviations (SD) computed by the estimated variance of the baseline transmission rate.

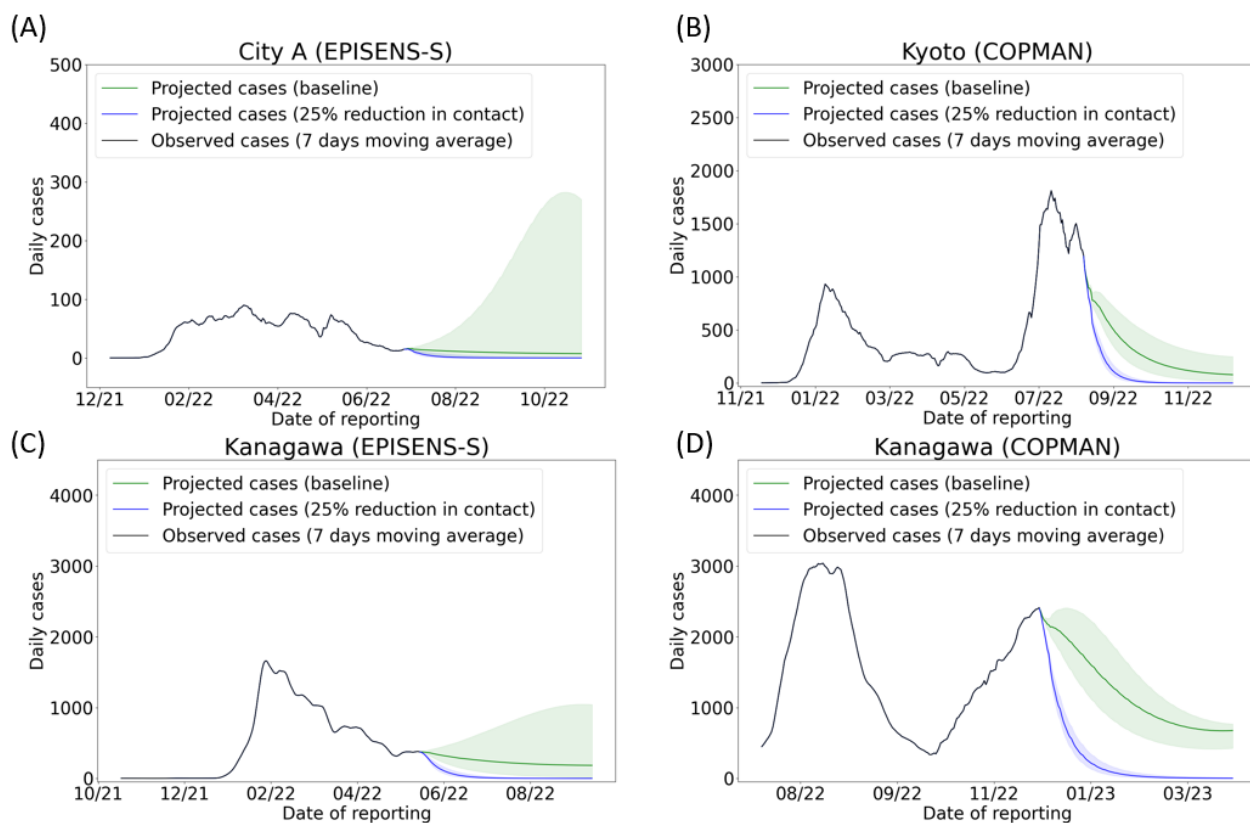


Figure S7. Model projected cases for a non-pharmaceutical intervention scenario. Relative contact rates are set as 1 for the baseline (green) and 0.75 (blue) and for scenarios with reduced contact rates. Each ribbon represents uncertainty ranges of 2 standard deviations (SD) computed by the estimated variance of the baseline transmission rate.

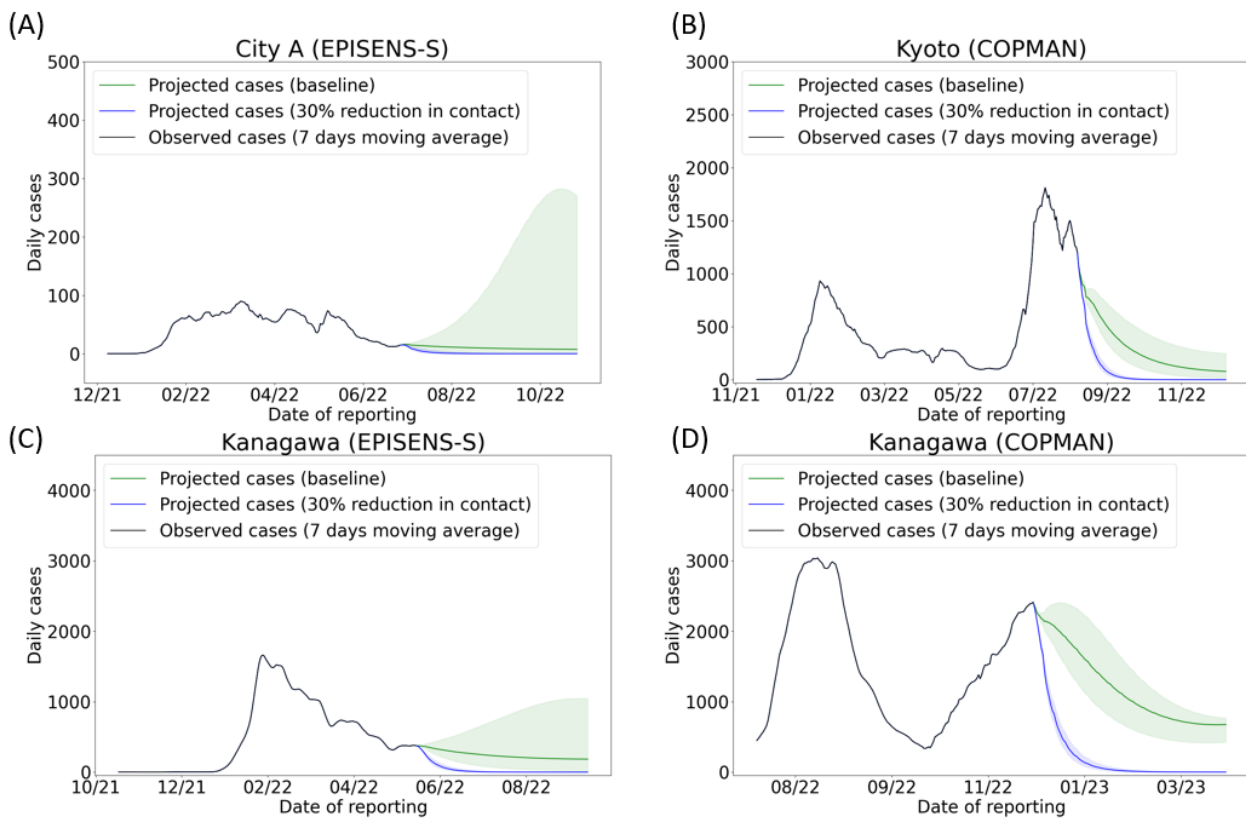


Figure S8. Model projected cases for a non-pharmaceutical intervention scenario. Relative contact rates are set as 1 for the baseline (green) and 0.7 (blue) and for scenarios with reduced contact rates. Each ribbon represents uncertainty ranges of 2 standard deviations (SD) computed by the estimated variance of the baseline transmission rate.

Table S1. Data source.

Municipalities	Wastewater treatment plant	Epidemiological data	Method	Period (Sampling frequency)	Wastewater data
Kyoto city	<ul style="list-style-type: none"> • Toba first WWTP • Toba second WWTP 	https://www.city.kyoto.lg.jp/hokenfukushi/page/0000268303.html	COPMAN	15 Dec 2021 ~ 28 Feb 2022 (3 times/week), 2 March 2022 ~ 31 Aug 2022 (2 times/week)	Not accessible to the general public
City A	<ul style="list-style-type: none"> • East WWTP 	Not accessible to the general public	EPISENS-S	15 Dec 2021 ~ 30 Jun 2022 (2 times/week)	Not accessible to the general public
A part of Kanagawa Prefecture (Sagami river basin)	<ul style="list-style-type: none"> • Right bank of the Sagami River WWTP • Left bank of the Sagami River WWTP 	http://covid-map.bmi-tokai.jp/choroplethmap_kanagawa/	EPISENS-S	1 Nov 2021 ~ 30 Mar 2022 (3 times /week), 19 Apr 2022 ~ 27 May 2022 (2 times/week)	https://www.pref.kanagawa.jp/docs/ga4/covid19/simulation.html
			COPMAN	8 Jul 2022 ~ 27 Dec 2022 (2 times/week)	Not accessible to the general public

Table S2. List of input parameters.

Parameter	Symbol	Kyoto city	City A	A part of Kanagawa Prefecture (COPMAN)	A part of Kanagawa Prefecture (EPISENS-S)
Sewered population size	N	778000	157113	1241200	
Latent period (days)	$1/\alpha$	1.5			
Infectious period (days)	$1/\tau$	2.0			
Immunity duration (days)	$1/\omega$	180			
Special holidays		2021/1/1,2021/1/11,2021/2/11,2021/2/23,2021/3/20,2021/4/29,2021/5/3,2021/5/4,2021/5/5,2021/7/20,2021/8/8,2021/8/9,2021/9/20,2021/9/23,2021/11/3,2021/11/23,2022/1/10,2022/2/11,2022/2/23,2022/3/21,2022/4/29,2022/5/3,2022/5/4,2022/5/5,2022/7/18,2022/8/11,2022/9/19,2022/9/23,2022/10/10,2022/11/3,2022/11/23			

Table S3. List of estimated parameters.

Parameter	Symbol	Kyoto city	City A	A part of Kanagawa Prefecture (EPISENS-S)	A part of Kanagawa Prefecture (COPMAN)
Shedding duration (days)	$1/\gamma$	1.97	0.64	1.71	1.74
Scaling parameter	ν	2.6×10^{-11}	2.0×10^{-9}	1.6×10^{-11}	1.7×10^{-11}
Calibration period		15 Dec 2021 to 31 Aug 2022	15 Dec 2021 to 30 Jun 2022	1 Nov 2021 to 27 May 2022	8 July 2022 to 27 Dec 2022

Table S4. Summary statistics for model fits.

Municipalities		Kyoto city	City A	A part of Kanagawa Prefecture	
Method		COPMAN	EPISENS-S	COPMAN	EPISENS-S
MAE*	Case data	1869.3	91.6	3135.8	1303.3
	Wastewater data	2032.9	178.1	4285.8	1397.5
	Case and wastewater data	1475.9	81.8	5738.7	1305.9
RMSE*	Case data	4447.9	173.4	9632.6	2792.5
	Wastewater data	3981.3	247.1	6969.1	2336.4
	Case and wastewater data	3321.4	124.1	24071.8	2743.0

*MAE and RMSE are calculated by comparing the predicted and actual numbers of cumulative new positive cases up to one week.