Additional file 1

A Data sources

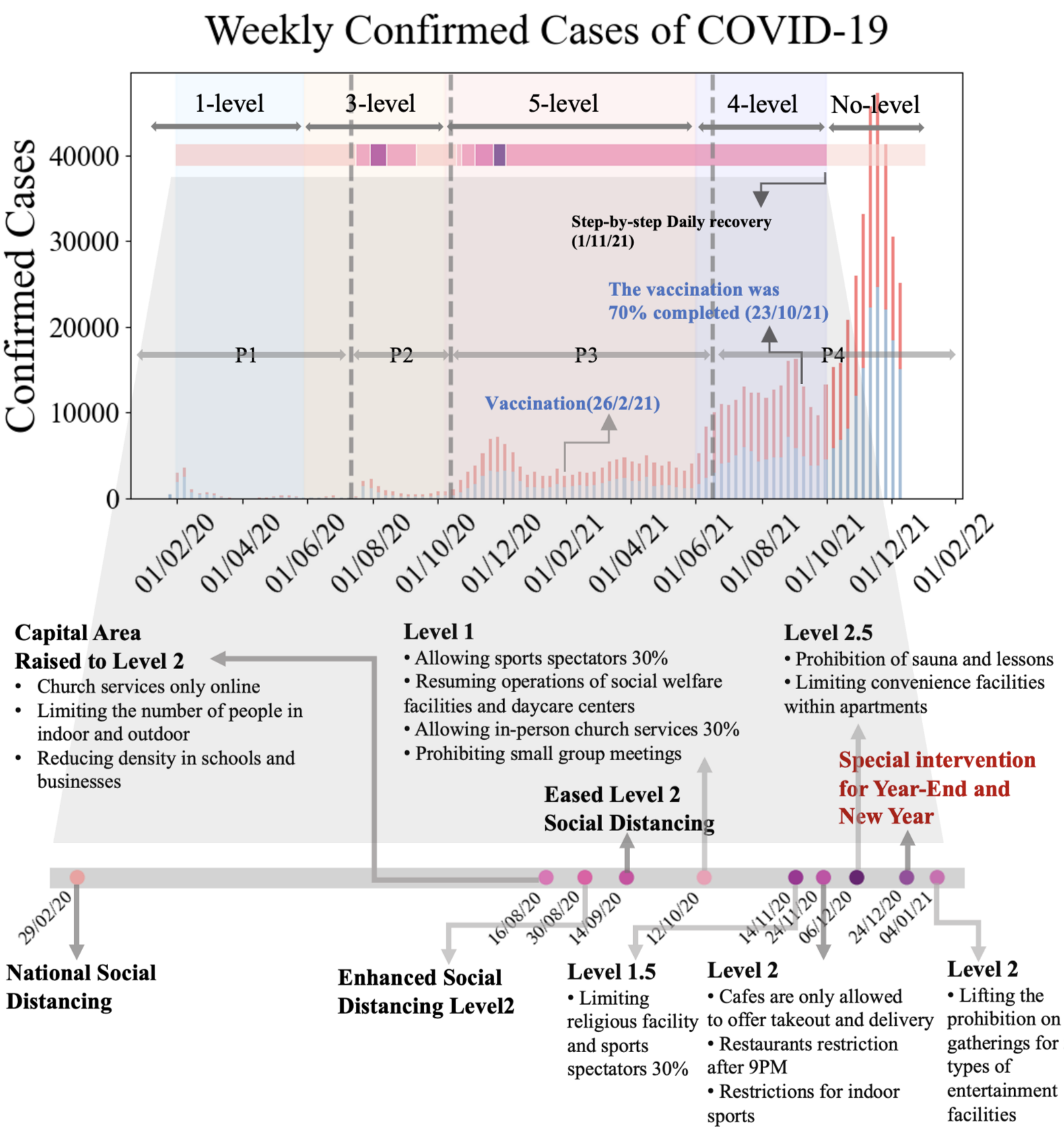

Fig. S1: Covid-19 Social Distancing Timeline in South Korea from February 2020 to February 2021.

Table S1: Interventions according to social distancing measures by period (Feb 29, 2020 – Feb 01, 2021)

Period	Social Distancing Measure	Interventions

P1	National social distancing (Feb 29, 2020 – Aug 15, 2020)	• Recommend refraining from going out, minimizing contact with others, etc.
		• Close daycare center (Feb 27, 2020) and social welfare facilities (Feb 28, 2020), etc.
		• Shift to social distancing in daily life (May 6, 2020)
		  – Daily activities are allowed under adherence to quarantine guidelines.
		• Implement enhanced social distancing in metropolitan area (May 29, 2020 ∼ June 14, 2020)
		  – Recommend to refrain from operating entertainment bars, karaoke, private institutes; suspend operation of public facilities, etc.

P2	Capital area raised to Level 2 (Aug 16, 2020 – Aug 29, 2020)	• Seoul and Gyeonggi province upgraded to the level 2
		  – Mandatory infectious disease prevention and control measures (restrict the number of users, movement within facilities, etc.), refrain from moving to other cities and provinces.
		• Enforce Level 2 throughout the entire metropolitan area and additional action
		  – Prohibit gathering more than 50 people indoors/100 people outdoors, Only non-face-to-face worship allowed, etc.

	Enhanced social distancing Level 2 (Aug 30, 2020 – Sep 13, 2020)	• Only take-out and delivery allowed from 21–05:00 at restaurants and cafes • Indoor sports facilities, reading rooms, and study cafes are prohibited, etc. • Only non-face-to-face classes are allowed at private institutes, Ban visiting to nursing hospitals and facilities, etc.

	Eased social distancing Level 2 (Sep 14, 2020 – Oct 11, 2020)	• Maintain the "Level 2" (mitigating some control measures)
		  – Restaurants, cafes are allowed to operate after 9 o'clock, etc.
		  – Lift of the ban on gatherings at indoor sports facilities, etc. (apply distancing rules between users, etc.)

	Level 1 (Oct 12, 2020 – Nov 18, 2020)	• Allowing sports spectators 30%. • Resuming operations of social welfare facilities and daycare centers. • Allowing in-person church services 30%. • Prohibiting small group meetings.

P3	Level 1.5 (Nov 19, 2020 – Nov 23, 2020)	• Limiting religious facility and sports spectators 30%. • Applied to Seoul and Gyeonggi province (Incheon will be applied from the 23rd) – Add restriction on the number of users in major facilities
	Level 2 (Dec 06, 2020 – Dec 23, 2020)	• Prohibition of sauna and lessons. • Limiting convenience facilities within apartments. • Cessation of operating movie theaters and department stores after 21:00, application of non-face-to-face principles such as worship services, etc. • Strengthen the quarantine measures during the New Year holidays, prohibit gathering of 5 or more people (Dec 24, 2021).
	Level 2.5 (Nov 24, 2020 – Dec 05, 2020)	• Cafes are only allowed to offer takeout and delivery. • Restaurants restriction after 9P. • Restrictions for indoor sports. • Prohibit gatherings entertainment facilities and expand restrictions on the number of users in general facilities, etc.
	Special intervention for year-end and New Year (Dec 24, 2020 – Jan 03, 2021)	• Strengthen the quarantine measures during the New Year holidays, prohibit gathering of 5 or more people.
	Level 2 (Jan 04, 2021 – Oct 31, 2021)	• Downgraded and maintained on level 2. • Announce strengthen control measures in the metropolitan area (Jul 4, 2021) and implementation of additional measures (Jul 7, 2021). – Reinforcement of diagnostic tests (install additional screening centers), Strengthen preemptive tests for high-risk groups, etc.
P4	Level 2 (Jan 04, 2021 – Oct 31, 2021)	• Downgraded and maintained on level 2. • Announce strengthen control measures in the metropolitan area (Jul 4, 2021) and implementation of additional measures (Jul 7, 2021). – Reinforcement of diagnostic tests, Strengthen preemptive tests for high-risk groups, etc.
	Step-by-step daily recovery (Nov 01, 2021 – Feb 01, 2022)	• Transition to step-by-step daily recovery. • Strength measures for medical response in the metropolitan area (Nov 5, 2021). – Secure the number of hospital beds, reinforce control at high risk facilities, etc.

Table S2: Three-level social distancing guidelines (Jun 28, 2020 – Nov 06, 2020)

	Level 1	**Level 2**	**Enhanced Level 2**	**Level 3**
outbreak status	Repeated spread and mitigation of small-scale sporadic cases under the control of the healthcare system	COVID-19 continues to spread in the community beyond what can be afforded by the conventional healthcare system.		Multiple cases of mass outbreaks in the community; COVID-19 is spreading rapidly and on a large scale
Number of daily cases	Less than 50 cases	50 or more but less than 100 cases.		More than 100 people, with a more than twofold increase twice a week.
School Restrictions	Full in-person attendance is possible based on the two-thirds principle of the total capacity	Up to two-thirds of students attending in-person (for high schools), up to one-third of students attending in-person (for elementary and middle schools).	Only 12th-grade students attend in-person (for high schools), fully remote learning or school closure (for elementary and middle schools).	Only 12th-grade students attend in-person (for high schools), fully remote learning or school closure (for elementary and middle schools).
Social gathering	Compliance with health and safety guidelines, limitations on the number of spectators at sports events	Indoor gatherings of 50 or more people prohibited, outdoor gatherings of 100 or more people prohibited, and sports events allowed only without spectators.		Gatherings of 10 or more people prohibited (both indoor and outdoor), and sports events allowed only in a non-face-to-face format.
Work place	Minimize workplace density (public institutions), encourage minimizing workplace density (private companies).	Limit the number of employees (public institutions), recommend limiting the number of employees (private companies).		Only essential personnel report to work (public institutions), remote work recommended for non-essential personnel (private companies).
Transit	Restriction on boarding if not wearing a mask.		Reduced nighttime operations.	Reduced operations and suspension of certain routes.

Table S3: Five-level social distancing guidelines (Nov 07, 2020 – Jul 01, 2021)

	Level 1	**Level 1.5**	**Level 2**	**Level 2.5**	**Level 3**
Outbreak status	Sporadic outbreaks	Community transmission begins	Full-blown community	Nationwide pandemic begins	Full-blown nationwide pandemic
Number of daily cases	Less than 100 cases (capital area), less than 30 cases (other areas)	More than 100 cases (capital area), more than 30 cases (other areas)	More than double the number of cases compared to the Level 1.5 criteria, or continued Level 1.5 spread in two or more regions, or more than 300 cases nationwide.	Nationwide 400–500 or more cases, or a rapid doubling of cases.	Nationwide 800–1000 or more cases, or a rapid doubling of cases.
School Restrictions	The two-thirds rule (full in-person attendance possible).	Adherence to the two-thirds rule.	Adherence to the two-thirds rule (for high schools), the one-third rule (for elementary and middle schools).	Adherence to the one-third rule.	Fully remote learning.
Social gathering	In cases of 500 or more people, consultation with local government is required.	For festivals or certain gatherings, gatherings of 100 or more people are prohibited.	Gatherings of 100 or more people prohibited.	Gatherings of 50 or more people prohibited.	Gatherings of 10 or more people prohibited.
Work place	Mandatory mask-wearing for high-risk groups & recommend implementing remote work at appropriate ratios.	Recommend remote work at appropriate ratios.	Recommend remote work at appropriate ratios.	Recommend remote work or similar measures for at least one-third of the workforce.	Recommend remote work for all non-essential personnel.
Transit	Restriction on boarding if not wearing a mask.	Restriction on boarding if not wearing a mask.	Prohibition on food consumption, except on international flights.	Recommend limiting ticket sales to 50%, excluding aircraft.	Limit ticket sales to 50%, excluding aircraft.

Table S4: Four-level social distancing guidelines (Jul 01, 2021 – Oct 31, 2021)

	Level 1	**Level 2**	**Level 3**	**Level 4**
Outbreak status	Maintain a state of continuous suppression	Regional outbreak	Regional epidemic	Nationwide pandemic
Criteria (Weekly average of confirmed cases per 100,000 people)	Less than 1 case per 100,000 people (Nationwide: fewer than 500 cases; Capital region: fewer than 250 cases)	More than 1 case per 100,000 people (Nationwide: more than 500 cases; Capital region: more than 250 cases)	More than 2 cases per 100,000 people (Nationwide: more than 1000 cases; Capital region: more than 500 cases)	More than 4 cases per 100,000 people (Nationwide: more than 2000 cases; Capital region: more than 1000 cases)
School Restrictions	Full in-person attendance	Full in-person attendance	Full in-person attendance possible	Up to two-thirds attendance (elementary), up to two-thirds attendance (middle), full in-person attendance possible (high schools)
Social gathering	Compliance with health and safety guidelines	Gatherings of 9 or more people prohibited	Gatherings of 5 or more people prohibited	Gatherings of 5 or more people prohibited
Transit	–	For workplaces with 300 or more employees, 10% should work remotely	For workplaces with 50 or more employees, 20% should work remotely	For all workplaces, 30% should work remotely
Work place	Working from home is recommended for a proportion of workers per organization/division	Expansion of working from home is recommended	Expansion of working from home is recommended	Over one-third of employees are recommended to work from home
Common Facility Health Guideline	Display and inform about health guidelines, check for symptoms and restrict entry for symptomatic individuals, maintain an entry log, enforce mask-wearing, prohibit food consumption, promote handwashing, reduce crowding, ensure ventilation, conduct disinfection, designate and manage a health officer.			

B Unified modeling and real-time R(t) estimation

B.1 Quantifying Behavioral Dynamics from Mobility Data

To quantify temporal fluctuations in contact behavior, we constructed a mobility-based behavioral index reflecting population-level responses to distancing policies, using data from the Google COVID-19 Community Mobility Reports

$$\widehat{Y}_j(t) = \text{GAM}_j\big(\text{date}, \text{holiday}, \text{intervention}, \text{DOW}\big) \tag{1}$$

where $j \in \{\text{retail}, \text{transit}, \text{workplace}, \text{parks}, \text{residential}\}$. Among these, we define the household mobility indicator $H(t)$ as the smoothed residential mobility:

$$H(t) = \widehat{Y}_{\text{residential}}(t) \tag{2}$$

To model the behavioral shift in non-household contact activity, we define $\delta(t)$ as the relative change in contact intensity compared to a pre-pandemic baseline. This quantity is represented on a logarithmic scale to ensure additivity and interpretability. Specifically, we model it as a linear combination of mobility trends across non-residential categories:

$$\log \delta(t) = \sum_{j=1}^{4} w_j \cdot \widehat{Y}_j(t). \tag{3}$$

Here, $\widehat{Y}_j(t)$ denotes the smoothed mobility trend for category j, and w_j is a positive-valued coefficient drawn from a prior distribution $w_j \sim \mathcal{N}^+(0,1)$, representing the contribution of category j to behavioral intensity. Since the relationship between mobility and contact frequency is fundamentally multiplicative, the use of a logarithmic transformation renders this relationship additive in the model, enhancing both interpretability and numerical stability. This formulation also guarantees the positivity of derived quantities such as the estimated contact volume $M(t)$ and improves estimation tractability. Accordingly, the estimated number of non-household contacts $M(t)$ is modeled via a log-linear relationship anchored to the baseline number of daily non-household contacts C_N, defined as the pre-pandemic (pre-January 2020) average:

$$\log M(t) = \log C_N + \sum_{j=1}^{4} w_j \cdot \widehat{Y}_j(t), \tag{4}$$

Equation (3) quantifies the proportional influence of activity-specific mobility trends on changes in contact behavior, while Equation (4) scales these proportional changes to yield an absolute, time-varying estimate of non-household contact volume. The probabilistic formulation of w_j allows for uncertainty quantification and enables the model to adaptively weight mobility domains based on their empirical association with contact frequency. This structure also reflects the reality that behavioral responses are shaped by interacting activities rather than isolated effects. Given its log-linear nature, $\log M(t)$ is particularly sensitive to abrupt or sustained increases in mobility, making it a responsive indicator of heightened transmission potential. This mobility-informed modeling framework allows for real-time estimation of contact intensity under varying policy environments and population behaviors.

By substituting Equation (3) into Equation (4), we obtain a compact expression for $M(t)$:

$$\log M(t) = \log C_N + \log \delta(t) = \log\big(C_N \cdot \delta(t)\big), \tag{5}$$

which leads to the multiplicative form:

$$M(t) = C_N \cdot \delta(t). \tag{6}$$

This highlights the interpretation that the estimated non-household contact volume $M(t)$ is a scaled version of the baseline C_N, modulated by time-varying relative change $\delta(t)$ as inferred from mobility data.

In parallel, surveillance performance—particularly the timing of case detection and isolation—plays an equally critical role in modulating realized transmission. We therefore incorporate diagnostic delay into our model to adjust the infectiousness profile accordingly.

B.2 Surveillance Effect and Generation Interval Adjustment

In addition to contact behavior, the timing of diagnosis and isolation plays a critical role in limiting onward transmission. To reflect this, our model explicitly incorporates the effect of case surveillance by adjusting the infectiousness profile using empirically observed delays from infection to diagnosis. Let $f(t, t')$ denote the time-varying probability density function for the delay between infection and diagnosis, where t' is the time since infection. This is estimated using the empirical distribution of the difference between diagnosis date and inferred infection date, with the latter assumed to be five days prior to symptom onset based on the mean incubation period. Delay values outside the plausible range of -5 to 42 days are excluded as outliers.

We define the surveillance adjustment factor $D(t)$ as the expected reduction in infectiousness due to case detection and isolation:

$$D(t) = \sum_{t'=0}^{\infty} f(t, t') \cdot s^*(t'), \tag{7}$$

where $s^*(t')$ is the baseline serial interval distribution, capturing the natural infectiousness profile in the absence of intervention. This convolution reflects the extent to which earlier detection truncates the period during which an individual can transmit the virus.

The surveillance-adjusted generation interval $g(t, t')$ is then defined as:

$$g(t, t') = \frac{f(t, t') \cdot s^*(t')}{D(t)}. \tag{8}$$

This formulation normalizes the adjusted kernel to ensure it integrates to one, preserving its probabilistic interpretation. The shape of $g(t, t')$ varies dynamically with surveillance intensity—becoming concentrated around shorter delays when detection is rapid, and spreading out when delays are long.

We then define the total infectiousness pressure at time t, denoted $I(t)$, as the convolution of $g(t, t')$ with observed local incidence $N(t')$:

$$I(t) = \sum_{t'=0}^{t} g(t, t') \cdot N(t'). \tag{9}$$

This structure allows us to weight past infections by their current contribution to transmission potential under prevailing surveillance conditions. By modeling $I(t)$ in this way, we incorporate real-time variation in diagnostic responsiveness and account for delays in case identification—factors that strongly influence effective reproduction number estimates.

B.3 Estimating the Effective Reproduction Number, R(t)

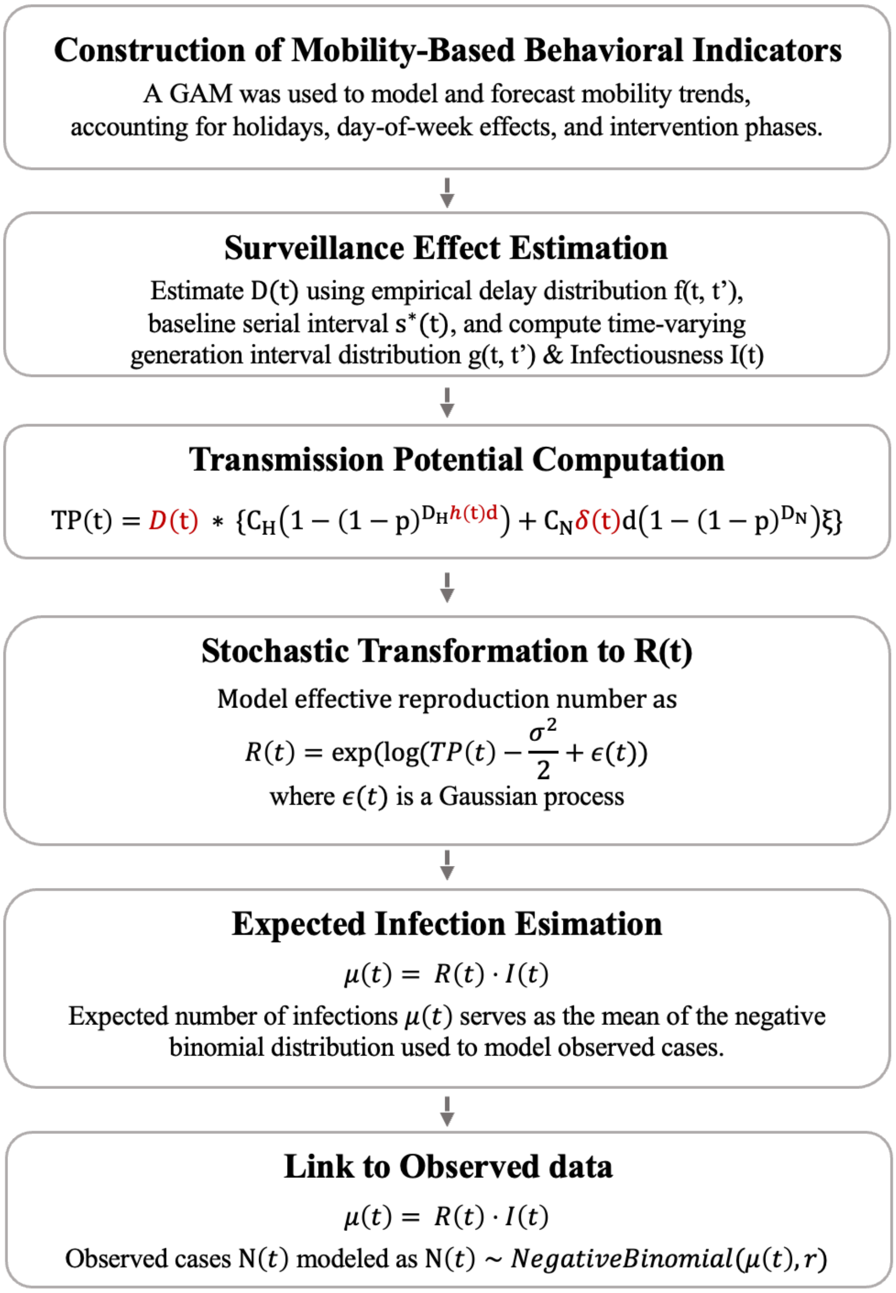

Fig. S2: Schematic of the modeling framework for estimating $R(t)$ and predicting daily case counts. The diagram illustrates the sequential flow from input data—including mobility-based behavioral indicators and surveillance delay distributions—through the computation of transmission potential (TP), the stochastic derivation of R(t), and the generation of case predictions. The layout is structured to emphasize the logical relationships among components and ensure visual clarity.

To quantify the evolving transmissibility of COVID-19, we estimated a time-varying effective reproduction number, $R(t)$, derived from a modified Transmission Potential (TP) model tailored to the Korean epidemic context. The TP model integrates setting-specific contact structures, behavioral dynamics, and surveillance effects into a unified framework. A schematic overview of this modeling framework is presented in Fig. S2, which outlines the sequential process from input variables to case prediction.

We first computed the deterministic transmission potential $TP(t)$, representing the expected number of secondary infections per case under prevailing conditions. This estimate incorporated time-varying inputs: $\delta(t)$: an indicator reflecting changes in non-household contact intensity derived from mobility trends; $H(t)$: an indicator reflecting changes in household contact duration derived from mobility trends; $D(t)$: a surveillance factor summarizing the reduction in infectiousness due to timely detection and isolation.

The TP at time t is calculated as:

$$TP(t) = D(t)\left\{C_H\left[1-(1-p)^{D_H d H(t)}\right] + C_N \delta(t) d\left[1-(1-p)^{D_N}\right]\xi\right\} \tag{10}$$

Here, C_H and C_N represent the baseline number of daily contacts in household and non-household settings, respectively, while D_H and D_N denote the baseline daily contact duration in household and non-household environments, respectively. The infectious period is denoted by d, and p is the per-contact transmission probability. The parameter ξ accounts for the proportion of non-household contacts modified by preventive behaviors such as mask-wearing.

This formulation integrates transmission dynamics in both household and non-household settings. Each component includes the contact rate, contact duration, and the probability of at least one successful transmission, modeled via the binomial expression $1-(1-p)^{\text{duration}}$, assuming independent per-time-unit infection risk.
Since p is not directly observable, we calibrated its value to align TP with a target basic reproduction number $R_0 = 3.22$

$$R_0 = \left(1 + rT_g v^2\right)^{1/v^2} \tag{11}$$

Here, r is the exponential growth rate, T_g is the mean generation interval, and v is the coefficient of variation. With $T_g = 4.52$ and $v = 0.639$, and assuming exponential growth during the early outbreak phase, we estimated $p = 0.1147$ to match the TP with the target R_0. This was then used as the mean of the prior distribution $p \sim N(0.114, 0.08)$.

The infectious period d was defined as the expected value of the baseline serial interval distribution $s^*(t')$, obtained by fitting a truncated gamma distribution (1–14 days) to 22,340 infector–infectee pairs:

$$d = \sum_{t'} t' \cdot s^*(t') \tag{12}$$

The resulting mean value was $d = 4.52$, which we assumed as the average duration of infectiousness in our model calibration.

To account for stochastic fluctuations and unobserved heterogeneity, we modeled $R(t)$ as a log-normal transformation of $TP(t)$:

$$R(t) = \exp\left(\log TP(t) - \frac{\sigma^2}{2} + \epsilon(t)\right) \tag{13}$$

where $\epsilon(t)$ follows a Gaussian process with temporal correlation:

$$\epsilon \sim \mathcal{GP}(0, k(t,t')) \tag{14}$$

$$k(t,t') = \sigma^2\left(1 + \frac{5|t-t'|}{l} + \frac{5(t-t')^2}{3l^2}\right)\exp\left(-\sqrt{5}\frac{|t-t'|}{l}\right) \tag{15}$$

The parameter σ controls the marginal variance of the Gaussian process and regulates the magnitude of short-term fluctuations in $R(t)$. Higher values of σ allow the model to respond more sensitively to rapid changes in transmission, while smaller values yield smoother, more conservative trajectories. This Matérn 5/2 kernel formulation enables $R(t)$ to evolve smoothly while remaining responsive to sudden shifts in epidemic intensity, such as those caused by superspreading events or delayed interventions.

By combining deterministic behavioral drivers with stochastic variability, $R(t)$ offers a flexible and responsive measure of epidemic potential. The resulting $R(t)$ values are subsequently linked to observed incidence data for prediction and inference, as described in the following section.

B.4 Prediction model

Given the estimated effective reproduction number $R(t)$ from the previous section, we constructed a probabilistic model to connect latent transmission dynamics to observed daily case counts.

Specifically, we define the expected number of new infections at time t as:

$$\mu(t) = I(t) * R(t) \tag{16}$$

where $I(t)$ is the total infectiousness contribution from past infections, derived from the surveillance-adjusted generation interval $g(t, t')$.

To reflect the stochastic nature of reported case counts, we assume:

$$N(t) \sim \text{NegBinomial}(\mu(t), r) \tag{17}$$

where r is an overdispersion parameter accounting for reporting variability and cluster-driven fluctuations. This observational model defines the likelihood for Bayesian inference, allowing parameter estimation to be informed by real-time surveillance data.

The combined structure of $R(t)$ and this prediction model enables the full integration of mobility behavior, surveillance efficiency, and stochastic transmission dynamics into a unified epidemic modeling framework.

Posterior inference was performed using Markov Chain Monte Carlo (MCMC) methods implemented in PyMC and ArviZ. All posterior inference was conducted in Python 3.9 using PyMC v4.0 with the No-U-Turn Sampler (NUTS). A total of 10 parallel chains were run, with 1,000 post-burn-in samples collected per chain. Convergence was assessed based on the criteria of $\hat{R} < 1.1$ and an effective sample size (ESS) greater than 1,000.