Supplementary Materials for: Delayed introduction and susceptible variation drive spatial asynchrony in pertussis epidemics

Sang Woo Park[1,*]

1 School of Biological Sciences, Seoul National University, Seoul, Korea

*Corresponding author: sangwoopark@snu.ac.kr

Materials and Methods

S1 Data

The weekly surveillance data on pertussis cases across 252 municipalities (1st week of 2024 to the 44th week of 2025) were obtained from a publicly available website, the Infectious Disease Portal, by the Korea Disease Control and Prevention Agency [34]. Population size data as of September 2025 were obtained from publicly available website by the Ministry of the Interior and Safety [35]. Vaccine coverage data were obtained from a publicly available website by the Korea Disease Control and Prevention Agency [36]. Longitude and latitude data for each municipality were obtained from a publicly available GitHub repository [37]. Finally, shape files used for constructing maps were obtained from a publicly available website [38].

S2 Center of gravity

We quantified the center of gravity to characterize variation in the mean timing of epidemic. Typically, the center of gravity is calculated based on annual incidence for recurrent epidemics. Since we are focused on analyzing a single outbreak, we use the entire time series of reported cases C_t to computer the center of gravity for each municipality:

$$\text{Center of gravity} = \frac{\sum_{t=1}^{t_{\max}} C_t \times t}{\sum_{t=1}^{t_{\max}} C_t}. \tag{1}$$

We then calculate the correlation coefficient between the center of gravity and the timing of introduction, defined as the first week when the number of reported cases is greater than 1 in 100,000.

S3 Spatial synchrony of reported cases

We characterized the spatial synchrony of pertussis epidemic, which captures changes in pairwise correlation as a function of distance [7]. We used logged values of reported

cases between May 2024 and March 2025, excluding municipalities that had no reported cases ($n = 250$). Spatial synchrony was calculated using the `ncf` package in R [39]. We used 1,000 bootstraps to generate the median estimate and the corresponding 95% confidence intervals.

S4 Effective reproduction number

We calculated the effective reproduction number $\mathcal{R}(t)$ to characterize changes in pertussis transmission over time. First, we took the case time series C_t and fitted a generalized additive model assuming a Poisson error to smooth the time series [18]. Then, we estimated $\mathcal{R}(t)$ using the method of [16]:

$$\mathcal{R}(t) = \frac{i(t)}{\sum_{k=1}^{n} i(t-k)g(k)}, \tag{2}$$

where $i(t)$ is the smoothed incidence and n is the maximum length of generation interval in weeks (assumed to be 5 weeks). The generation-interval distribution $g(k)$ is assumed to follow a gamma distribution with a mean of 9.47 days and a standard deviation of 6.22 days. The reproduction number estimates can be unrealistically high when the number of infections is close to zero. Therefore, we only use estimates between May 2024 and March 2025 throughout the paper. We truncated all $\mathcal{R}(t)$ estimates exceeding 5 to a maximum value of 5.

First, we compared correlation coefficients between $\mathcal{R}(t)$ estimates and those between logged cases across all pairwise municipality combinations (Figure 2B in the main text). For this analysis, we only use data from municipalities with more than 400 total cases. Then, we compared the spatial synchrony for $\mathcal{R}(t)$ estimates and synchrony for logged cases (Figure 2C in the main text). For this analysis, we used $\mathcal{R}(t)$ estimates from all municipalities without any exclusion.

Finally, we fitted a generalized additive model to logged values of $\mathcal{R}(t)$ to quantify the effect of susceptible depletion and temporal variation in intrinsic transmission [17, 18, 19]:

$$\log\left(\mathcal{R}_{t,m}\right) = \alpha_m + \beta d_{t,m} + s(t), \tag{3}$$

where $\mathcal{R}_{t,m}$ represents $\mathcal{R}(t)$ estimate at time t in municipality m; α_m represents the municipality-specific intercept term; β represents the effect of susceptible depletion; $d_{t,m}$ represents the cumulative cases at time t in municipality m; and $s(t)$ represents the smooth term capturing temporal variation in transmission. For this analysis, we only use $\mathcal{R}(t)$ estimates from municipalities with more than 400 total cases.

S5 Transmission model

Finally, we fitted a simple transmission model using Bayesian inference to test whether variation in introduction timing and susceptible dynamics alone can explain the observed epidemic patterns. Specifically, we extended the SEIR model,

which is commonly used for pertussis transmission [20], to allow for a joint estimation of a single, time-varying transmission term that is shared across all regions and a separate estimation of initial conditions and reporting rates. To do so, we first discretized the SEIR model following the approach of [40]:

$$\beta(t) = \mathcal{R}_0(1 - \exp(-\gamma\Delta t))\delta(t) \tag{4}$$

$$\text{FOI}(t) = \frac{\beta(t)I_m(t-\Delta t)}{N_m} \tag{5}$$

$$\Delta S_m(t) = \left[1 - \exp(-\text{FOI}(t)\Delta t)\right] S_m(t-\Delta t) \tag{6}$$

$$\Delta E_m(t) = \left[1 - \exp(-\sigma\Delta t)\right] E_m(t-\Delta t) \tag{7}$$

$$\Delta I_m(t) = \left[1 - \exp(-\gamma\Delta t)\right] I_m(t-\Delta t) \tag{8}$$

$$S_m(t) = S(t-\Delta t) - \Delta S_m(t) \tag{9}$$

$$E_m(t) = E(t-\Delta t) - \Delta E_m(t) + \Delta S_m(t) \tag{10}$$

$$I_m(t) = I(t-\Delta t) - \Delta I_m(t) + \Delta E_m(t) \tag{11}$$

where $S_m(t)$, $E_m(t)$, and $I_m(t)$ represent the number of susceptible, exposed, and infectious individuals in municipality m at time t, respectively; N_m represents the population size of municipality m; Δt represents the simulation time step, which assumed to be 1 week; $\beta(t)$ represents the shared time-varying transmission rate; $\mathcal{R}_0$ represents the shared basic reproduction number, assumed to equal 17 [41]; $\delta(t)$ represents the normalized time-varying transmission rate; σ represents the rate at which individuals develop infectiousness, which is assumed to be $\sigma = -\log(1 - 7/8)$ such that the mean latent period is 8 days; and γ represents the rate at which individuals recover, which is assumed to be $\sigma = -\log(1 - 7/15)$ such that the mean latent period is 15 days.

Here, we modeled changes in transmission using the $\delta(t)$ term, which is given a normal prior around 1:

$$\delta(t) \sim \text{Normal}(1, 0.2). \tag{12}$$

To allow for smooth variation in transmission, we also incorporated a random walk prior:

$$\delta(t) \sim \text{Normal}(\delta(t-\Delta t), \sigma_\delta), \tag{13}$$

$$\sigma_\delta \sim \text{Half-Normal}(0, 0.1), \tag{14}$$

where a half-normal prior on σ_δ constrains the smoothness of $\delta(t)$.

The municipality-specific observation process is modeled based on a Poisson distribution:

$$\text{cases}_{t,m} \sim \text{Poisson}(\rho_m \Delta S_m(t)) \tag{15}$$

$$\rho_m \sim \text{Uniform}(0, 1) \tag{16}$$

where $\text{cases}_{t,m}$ represents the reported cases at time t in municipality m; ρ_m represents the municipality-specific reporting rate; and $\Delta S_m(t)$ represents the expected number of new infections between time $t - \Delta t$ and t.

Finally, we estimate a municipality-specific initial conditions by imposing the following priors:

$$s_m(0) \sim \text{Uniform}(0, 1), \tag{17}$$
$$i_m(0) \sim \text{Half-Normal}(0, 0.001), \tag{18}$$

where $s_m(0)$ represents the initial fraction susceptible in municipality m, and $i_m(0)$ represents the initial fraction infected in municipality m. Then, the initial conditions for the model is specified as follows:

$$S_m(0) = N_m s_m(0), \tag{19}$$
$$E_m(0) = N_m i_m(0), \tag{20}$$
$$I_m(0) = N_m i_m(0). \tag{21}$$

We simultaneously fit this model time series data between May 2024 and March 2025 from all municipalities with more than 400 total cases (a total of 42 municipalities) using a Bayesian inference software rstan [21].

To assess the model fit, we compute R squared values for each municipality. This is done by calculating the squared value of the correlation coefficient between logged values of the observed cases+1 and a posterior median of the logged predictions $\rho_m \Delta S_m(t)$.

Finally, we evaluate the impact of initial conditions on pertussis epidemic dynamics by varying $s(0)$ between 0.13 and 0.23 and $i(0)$ between 7×10^{-6} and 1.5×10^{-3}. For each simulation, we quantify the center of gravity.

S6 Transmission model validation

We validate our model by fitting the model to remaining municipalities (i.e., those with less than 400 total cases). In doing so, we assume that changes in transmission rate $\delta(t)$ is known and only estimate the initial conditions, $s_m(0)$ and $i_m(0)$, as well as the reporting rate ρ. For this validation, we simply used the optimization function in rstan [21] rather than performing a full Bayesian inference. Then, we compute R squared values for each municipality in the same way as before.

S7 Relationshiip between estimated initial susceptible and vaccine coverage

We assessed the relationshiip between estimated initial susceptible and vaccine coverage using linear regression. Specifically, assuming that the waning of vaccinal immunity is the main cause for pertussis re-emergence, we compared DTaP (Diphtheria-Tetanus-Pertussis) vaccine coverage for 3 years old as of 2015 and the estimated initial susceptible in each municipality and performed a linear regression. We used vaccine coverage as of 2015 because this was the oldest data that were publicly available. We considered using estimates from municipalities with > 400 cases, since the

parameter estimates are expected to be more reliable, but were unable to find any clear signature between vaccine coverage and estimated $s(0)$.

Supplementary Figures

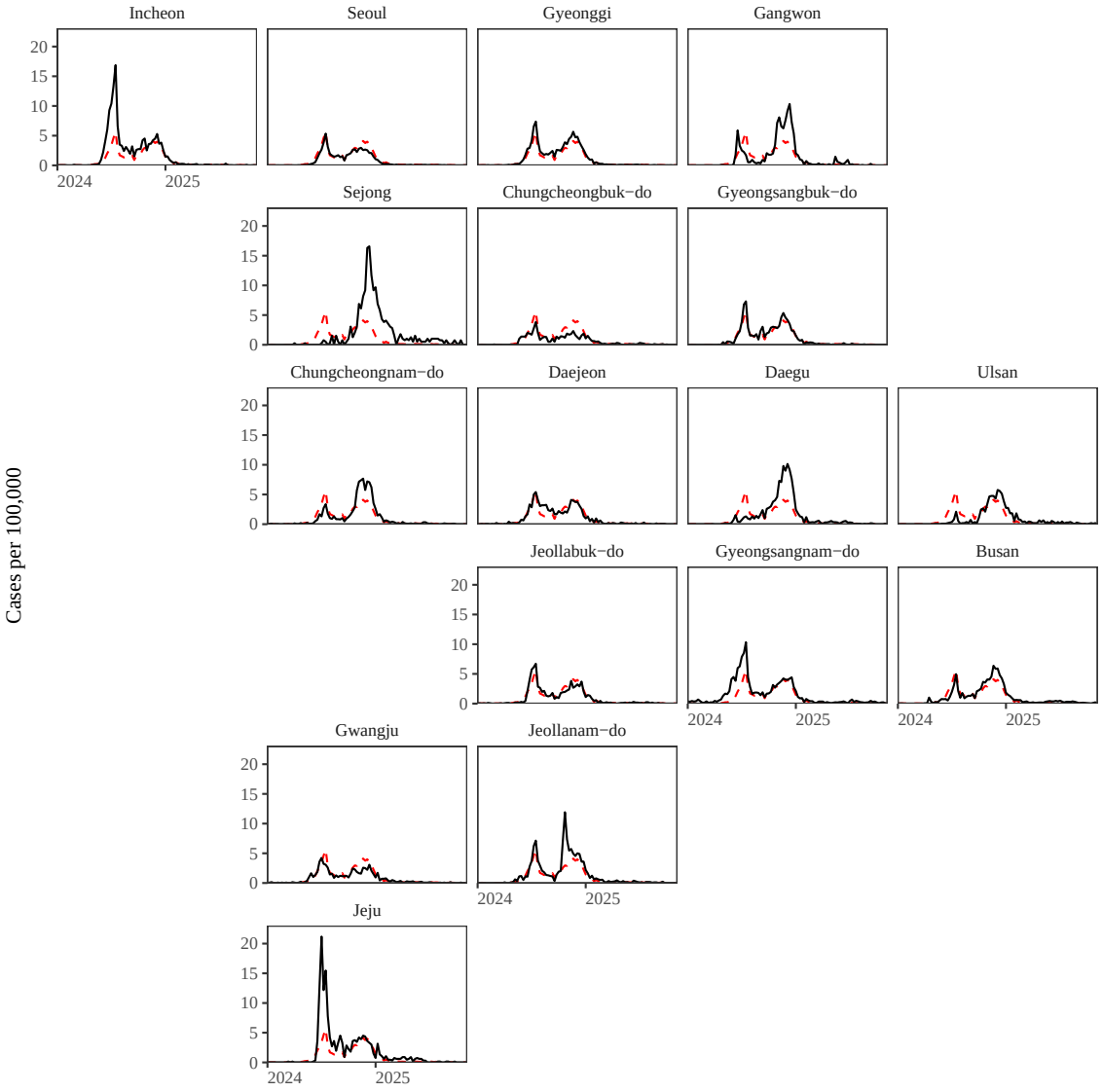

Figure S1: **Spatiotempordal dynamics of pertussis outbreak in Korea across first-level administrative divisions, 2024—2025.** Time series are arranged based on their approximate locations.

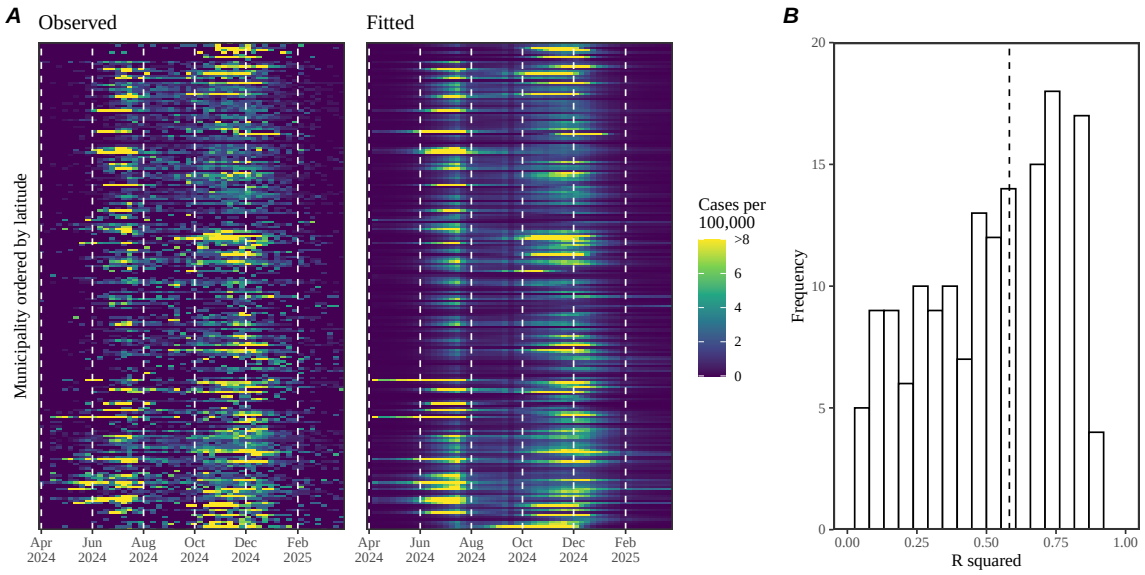

Figure S2: **Validation of our transmission model using time series data from municipalities with less than 400 total cases.** (A) Comparisons between the observed and predicted epidemic dynamics across municipalities with less than 400 total cases, ordered by latitude. (B) The bar plot represents the distribution of R squared values for model fits. The vertical dashed line represents the median.

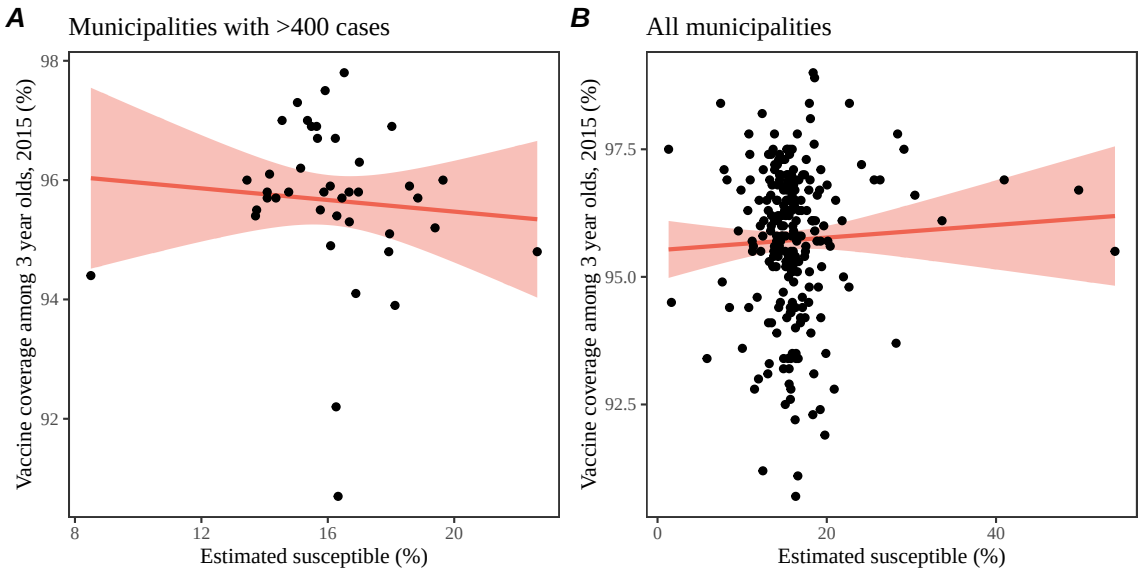

Figure S3: **Relationship between estimated initial susceptible and vaccine coverage for (A) municipalities with $>$ 400 cases and (B) for all municipalities.** Each point represents the estimate initial susceptible and vaccine coverage value for each municipality. Red lines and shaded regions represent the linear regression fit and corresponding 95% confidence intervals.

References

[1] Virginia E Pitzer, Cécile Viboud, Wladimir J Alonso, Tanya Wilcox, C Jessica Metcalf, Claudia A Steiner, Amber K Haynes, and Bryan T Grenfell. Environmental drivers of the spatiotemporal dynamics of respiratory syncytial virus in the United States. *PLoS pathogens*, 11(1):e1004591, 2015.

[2] Benjamin D Dalziel, Stephen Kissler, Julia R Gog, Cecile Viboud, Ottar N Bjørnstad, C Jessica E Metcalf, and Bryan T Grenfell. Urbanization and humidity shape the intensity of influenza epidemics in US cities. *Science*, 362(6410):75–79, 2018.

[3] Margarita Pons-Salort, M Steven Oberste, Mark A Pallansch, Glen R Abedi, Saki Takahashi, Bryan T Grenfell, and Nicholas C Grassly. The seasonality of nonpolio enteroviruses in the united states: Patterns and drivers. *Proceedings of the National Academy of Sciences*, 115(12):3078–3083, 2018.

[4] Rachel E Baker, Ayesha S Mahmud, Caroline E Wagner, Wenchang Yang, Virginia E Pitzer, Cecile Viboud, Gabriel A Vecchi, C Jessica E Metcalf, and Bryan T Grenfell. Epidemic dynamics of respiratory syncytial virus in current and future climates. *Nature communications*, 10(1):5512, 2019.

[5] Michael P Hassell, Hugh N Comins, and Robert M Mayt. Spatial structure and chaos in insect population dynamics. *Nature*, 353(6341):255–258, 1991.

[6] JA Sherratt. Periodic travelling waves in cyclic predator–prey systems. *Ecology Letters*, 4(1):30–37, 2001.

[7] Bryan T Grenfell, Ottar N Bjørnstad, and Jens Kappey. Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414(6865):716–723, 2001.

[8] Derek AT Cummings, Rafael A Irizarry, Norden E Huang, Timothy P Endy, Ananda Nisalak, Kumnuan Ungchusak, and Donald S Burke. Travelling waves in the occurrence of dengue haemorrhagic fever in thailand. *Nature*, 427(6972):344–347, 2004.

[9] Cécile Viboud, Ottar N Bjørnstad, David L Smith, Lone Simonsen, Mark A Miller, and Bryan T Grenfell. Synchrony, waves, and spatial hierarchies in the spread of influenza. *science*, 312(5772):447–451, 2006.

[10] Robert S Paton, Christopher E Overton, and Thomas Ward. The rapid replacement of the SARS-CoV-2 Delta variant by Omicron (B. 1.1. 529) in England. *Science Translational Medicine*, 14(652):eabo5395, 2022.

[11] Hye-Jin Kim, Young Joon Park, Dongkeun Kim, Jihee Lee, Yun Kyoung Kim, Soonryu Seo, Jin Seon Yang, Yeoeun Yun, Eunbyeol Wang, Subin Park, Seo Yeon Ko, Jin Lee, Jeeyeon Shin, Wookeon Lee, and Seonhee Ahn. Pertussis

surveillance and occurrence report in the Republic of Korea: from 2024 to the first half of 2025. *Public Health Weekly Report*, 18(43):1631–1651, 11 2025.

[12] Hyun Mi Kang, Taek-Jin Lee, Su Eun Park, and Soo-Han Choi. Pertussis in the post-COVID-19 era: resurgence, diagnosis, and management. *Infection & Chemotherapy*, 57(1):13, 2024.

[13] Scott M Duke-Sylvester, Luca Bolzoni, and Leslie A Real. Strong seasonality produces spatial asynchrony in the outbreak of infectious diseases. *Journal of the Royal Society Interface*, 8(59):817–825, 2011.

[14] Marc Choisy and Pejman Rohani. Changing spatial epidemiology of pertussis in continental USA. *Proceedings of the Royal Society B: Biological Sciences*, 279(1747):4574–4581, 2012.

[15] Jacco Wallinga and Marc Lipsitch. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):599–604, 2007.

[16] Anne Cori, Neil M Ferguson, Christophe Fraser, and Simon Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology*, 178(9):1505–1512, 2013.

[17] Dennis E te Beest, Michiel van Boven, Mariëtte Hooiveld, Carline van den Dool, and Jacco Wallinga. Driving factors of influenza transmission in the netherlands. *American journal of epidemiology*, 178(9):1469–1477, 2013.

[18] Simon N Wood. *Generalized additive models: an introduction with R*. Chapman and Hall/CRC, 2017.

[19] Stephen M Kissler, Christine Tedijanto, Edward Goldstein, Yonatan H Grad, and Marc Lipsitch. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*, 368(6493):860–868, 2020.

[20] Pejman Rohani, Xue Zhong, and Aaron A King. Contact network structure explains the changing epidemiology of pertussis. *Science*, 330(6006):982–985, 2010.

[21] Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76:1–32, 2017.

[22] Bärbel F Finkenstädt, Ottar N Bjørnstad, and Bryan T Grenfell. A stochastic model for extinction and recurrence of epidemics: estimation and inference for measles outbreaks. *Biostatistics*, 3(4):493–510, 2002.

[23] James O Lloyd-Smith, Sebastian J Schreiber, P Ekkehard Kopp, and Wayne M Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, 2005.

[24] FMG Magpantay, M Domenech De Cellés, P Rohani, and AA King. Pertussis immunity and epidemiology: mode and duration of vaccine-induced immunity. *Parasitology*, 143(7):835–849, 2016.

[25] Helen J Wearing and Pejman Rohani. Estimating the duration of pertussis immunity using epidemiological signatures. *PLoS pathogens*, 5(10):e1000647, 2009.

[26] Tina Tan, Tine Dalby, Kevin Forsyth, Scott A Halperin, Ulrich Heininger, Daniela Hozbor, Stanley Plotkin, Rolando Ulloa-Gutierrez, and Carl Heinz Wirsing Von König. Pertussis across the globe: recent epidemiologic trends from 2000 to 2013. *The Pediatric infectious disease journal*, 34(9):e222–e232, 2015.

[27] Matthieu Domenech de Cellès, Felicia MG Magpantay, Aaron A King, and Pejman Rohani. The impact of past vaccination coverage and immunity on pertussis resurgence. *Science translational medicine*, 10(434):eaaj1748, 2018.

[28] Seonghui Cho, Dong Wook Kim, Chiara Achangwa, Junseo Oh, and Sukhyun Ryu. Pertussis in the elderly: Plausible amplifiers of persistent community transmission of pertussis in South Korea. *Journal of Infection*, 89(3), 2024.

[29] David L Heymann and Guénaël R Rodier. Hot spots in a wired world: Who surveillance of emerging and re-emerging infectious diseases. *The Lancet infectious diseases*, 1(5):345–353, 2001.

[30] Mahmoud M Naguib, Patrik Ellström, Josef D Järhult, Åke Lundkvist, and Björn Olsen. Towards pandemic preparedness beyond COVID-19. *The Lancet Microbe*, 1(5):e185–e186, 2020.

[31] Hai Nguyen-Tran, Sang Woo Park, Kevin Messacar, Samuel R Dominguez, Matthew R Vogt, Sallie Permar, Perdita Permaul, Michelle Hernandez, Daniel C Douek, Adrian B McDermott, et al. Enterovirus D68: a test case for the use of immunological surveillance to develop tools to mitigate the pandemic potential of emerging pathogens. *The Lancet Microbe*, 3(2):e83–e85, 2022.

[32] Michael J Mina, C Jessica E Metcalf, Adrian B McDermott, Daniel C Douek, Jeremy Farrar, and Bryan T Grenfell. A global lmmunological observatory to meet a time of pandemics. *Elife*, 9:e58989, 2020.

[33] Hai Nguyen-Tran, Sang Woo Park, Matthew R Vogt, Perdita Permaul, Alicen B Spaulding, Michelle L Hernandez, Jennifer A Bohl, Sucheta Godbole, Tracy J

Ruckwardt, Peter W Krug, et al. Dynamics of endemic virus re-emergence in children in the USA following the COVID-19 pandemic (2022–23): A prospective, multicentre, longitudinal, immunoepidemiological surveillance study. *The Lancet Infectious Diseases*.

[34] Korea Disease Control and Prevention Agency. Notifiable Infectious Diseases. 2025. `https://dportal.kdca.go.kr/pot/is/rginEDW.do`. Accessed Nov 2, 2025.

[35] Ministry of the Interior and Safety. Resident Population and Household Statistics by Administrative District. 2025. `https://jumin.mois.go.kr/`. Accessed Nov 2, 2025.

[36] Korea Disease Control and Prevention Agency. 2015 National Immunization Coverage Survey. 2025. `https://nip.kdca.go.kr/irhp/infm/goNatnVcntStatView.do`. Accessed Nov 2, 2025.

[37] cubensys. Korea_district. 2018. `https://github.com/cubensys/Korea_District`. Accessed Nov 2, 2025.

[38] Hyung-jun Kim. Download the latest administrative divisions (SHP) of the Republic of Korea, GIS Developer. 2023. `http://www.gisdeveloper.co.kr/?p=2332`. Accessed Nov 2, 2025.

[39] Ottar N. Bjornstad. *ncf: Spatial Covariance Functions*, 2022. R package version 1.3-2.

[40] Daihai He, Edward L Ionides, and Aaron A King. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *Journal of the Royal Society Interface*, 7(43):271–283, 2010.

[41] Roy M Anderson and Robert M May. *Infectious diseases of humans: dynamics and control*. Oxford university press, 1991.