eAppendix

Hand, Foot and Mouth Disease in China:

Patterns of Spread and Transmissibility during 2008-2009

Yu Wang, Zijian Feng, Yang Yang, Steve Self, Yongjun Gao, Ira M. Longini, Jon Wakefield,

Jing Zhang, Liping Wang, Xi Chen, Lena Yao, Jeffrey D. Stanaway, Zijun Wang, Weizhong Yang

Statistical methods

Choosing the distributional model for reported case numbers

Here we derive a statistical transmission model for the surveillance data of HFMD. For stratum k of prefecture i, let $S_i^k(t)$ be the number of susceptible people at the beginning of week t, and $Y_i^k(t)$ the number of cases infected during week t who were identified by symptom onsets in week $t+1$, assuming a one-week incubation period. Let $\boldsymbol{X}_i^k(t)$ be the vector of covariates associated with stratum k of prefecture i in week t. Define A_i as the collection of prefectures that are adjacent to prefecture i. Let γ_0, γ_1 and γ_2 be the baseline rates of disease transmission from an infectious source to a susceptible person for the three types of sources: the unknown reservoir, an infectious person within the same prefecture, and an infectious person in an adjacent prefecture. The exact interpretation of these parameters depends on how covariates are formulated. For example, γ_1 may refer to the transmission rate when the susceptible person and the infectious person are both males in the age group of ≤ 1 year in the same prefecture. Let $\gamma_{ij}^{kl}(t)$ be the effective transmission rate, adjusted for covariates, from an infectious person in stratum l of prefecture j to a susceptible

person in stratum k of prefecture i during week t, such that the probability of the susceptible person escaping infection by the infectious person during that week is $\exp(-\gamma_{ij}^{kl})$. Let γ_{ic}^{k} be the effective transmission rate from the unknown reservoir to a susceptible person in stratum k of prefecture i during week t, where the unknown reservoir accounts for environmental sources, unreported cases and unobserved asymptomatic infections. Covariates can be adjusted for by

$$\begin{aligned} \gamma_{ij}^{kl}(t) &= \gamma_1^{\mathbf{1}_{j=i}} \gamma_2^{\mathbf{1}_{j\in A_i}} \exp\bigl(\boldsymbol{X}_i^k(t)'\boldsymbol{\beta}_S + \boldsymbol{X}_j^l(t)'\boldsymbol{\beta}_I\bigr), \text{ and} \\ \gamma_{ic}^{k}(t) &= \gamma_0 \exp\bigl(\boldsymbol{X}_i^k(t)'\boldsymbol{\alpha}_S\bigr), \end{aligned} \tag{1}$$

where $\mathbf{1}_{\text{condition}}$ indicates whether the condition is true (1) or not (0) , and $\boldsymbol{\beta}_S$, $\boldsymbol{\beta}_I$ and $\boldsymbol{\alpha}_S$ are the covariate effects on susceptibility (with subscript S) and infectiousness (with subscript I).

Assuming random mixing and that the transmission rates are constant regardless of the susceptible population size, the overall transmission risk a susceptible person in stratum k of prefecture i is exposed to during week t is

$$\gamma_i^k(t) = \gamma_{ic}^k(t) + \sum_{j\in\{i\}\cup A_i} \sum_{l=1}^{g} \left\{ \gamma_{ij}^{kl}(t) \sum_{\tau=t-d}^{t-1} Y_j^l(\tau) p_{t-\tau} \right\}, \tag{2}$$

where $p_{t-\tau}$ is the probability a case infected in week τ is infectious in week t. The values of p_l, $l = 1, \cdots, d$, are assumed known and subject to sensitivity analysis. The number of new cases then follows a binomial distribution, i.e., $Y_i^k(t) \sim \text{Binomial}\bigl(S_i^k(t), 1 - \exp\{-\gamma_i^k(t)\}\bigr)$.

Ideally, the values of $S_i^k(t)$'s are informed by seroprevalence rates of enteroviruses before and after the epidemics; however, such information is not available in China. Northern Taiwan and Singapore had similar age-specific EV71 seroprevalence rates before their first large-scale HFMD epidemics (year 1998 in Taiwan and year 2000 in Singapore), about 20%, 10%, 35% and 50–70% in the age groups 0–0.9, 1–2.9, 3–5.9 and 6+.[5,6] A separate serosurveillance in different areas across Taiwan after the 1998 epidemic found higher EV71 seroprevalence rates only in the age groups of 1–2.9 (20%) and 3–5.9 (42%).[5] The median population size over all Chinese prefectures is about 3.2

million, and the median sizes of the five age groups are approximately 42761, 81490, 118600, 142445 and 2833654. Therefore, assuming that China had similar pre-epidemic seroprevalence rates, the values of $S_i^k(t)$'s would be large at the prefecture level. In reality, individual-level contact rate is often found to decrease as the underlying population size gets larger. We further assume that γ_l, $l = 0, 1, 2$, are reciprocally proportional to the size of susceptible subpopulation $S_i^k(t)$ in the asymptotic sense, such that $\gamma_l S_i^k(t) \to \lambda_l$, $l = 0, 1, 2$, as $S_i^k(t) \to \infty$. The binomial distribution can then be approximated by a Poisson distribution:

$$Y_i^k(t) \sim \text{Poisson}(\lambda_i^k(t)), \tag{3}$$

where, analogous to expressions (1) and (2),

$$\begin{aligned}
\lambda_{ij}^{kl}(t) &= \lambda_1^{\mathbf{1}_{j=i}} \lambda_2^{\mathbf{1}_{j\in A_i}} \exp(\boldsymbol{X}_i^k(t)'\boldsymbol{\beta}_S + \boldsymbol{X}_j^l(t)'\boldsymbol{\beta}_I), \\
\lambda_{ic}^k(t) &= \lambda_0 \exp(\boldsymbol{X}_i^k(t)'\boldsymbol{\alpha}_S), \text{ and} \\
\lambda_i^k(t) &= \lambda_{ic}^k(t) + \sum_{j\in\{i\}\cup A_i} \sum_{l=1}^{g} \left\{ \lambda_{ij}^{kl}(t) \sum_{\tau=t-d}^{t-1} Y_j^l(\tau) p(t-\tau) \right\}.
\end{aligned}$$

This model is a special case of the general branching process used to describe disease transmission in large populations.[1,2] λ_0, λ_1 and λ_2 are interpreted as the weekly mean numbers of new infections the infectious source can generate. The λ's are population-level transmission rates, whereas the γ's are individual-level transmission rates.

An overdispersion parameter ϕ can be introduced to allow more variation in $Y_i^k(t)$, that is, $Y_i^k(t)$ follows a negative binomial distribution:

$$\begin{aligned}
\delta_i^k(t) &\sim \text{Gamma}(\phi, \lambda_i^k(t)/\phi), \text{ and} \\
Y_i^k(t) &\sim \text{Poisson}(\delta_i^k(t)).
\end{aligned} \tag{4}$$

We found that the special case with $\phi = \infty$, i.e., no overdispersion, fits the data better in the Southwest and the West regions. Therefore, we use the nonoverdispersed Poisson model for all the analyses.

Choosing the regression model

Risk factors under consideration for human-to-human transmission include age group, gender, population density, and weekly averages of the four climate indices: temperature, relative humidity, wind speed and precipitation. Weekly median climate indices at the national level are plotted over time together with the epicurves in eFigures 11. Because a large proportion of precipitation data were missing and were not reported using a unique standard, we aggregated precipitation to three levels with cut points at the tertiles of all the values. Speculating a relationship between the second rise and school opening in early September after the summer break, we add a time indicator for the school open period versus the school closure period (January 15 – February 15 and July 1-August 31) to the set of potential risk factors. The actual school closure periods differ up to a week across prefectures, but we ignore such variation due to lack of information. We stratify the model by geographic region and fit time-variation in baseline transmission rates with cubic splines to partially account for extra spatial and temporal heterogeneity unexplained by the risk factors. For the temporal variation, specifically, we choose the week $t^\star = 56$ (the first day of the week is January 20, 2009) that separates the 2008 and 2009 epidemics, and create six covariates, $\mathbf{1}_{t<t^\star} \times (t - t^\star)^k$, $k = 1, 2, 3$, and $\mathbf{1}_{t\geq t^\star} \times (t - t^\star)^k$, $k = 1, 2, 3$. The two cubic splines, one for the period from January 1 of 2008 to January 19 of 2009, and the other for the period from January 20 of 2009 to December 31 of 2009, represent the relative strength of the baseline transmission rates over time. We force the coefficients to be equal for the two first-order covariates, $\mathbf{1}_{t<t^\star} \times (t - t^\star)$ and $\mathbf{1}_{t\geq t^\star} \times (t - t^\star)$, such that the two splines are connected at week $t^\star$ with equal first derivatives, a property referred to as the first-order smoothness. Consequently, we have five effective covariates for modeling temporal variation.

We divide the covariates into seven classes as the following:

I. Age group and gender: five indicators;

II. Spatial region defined in eFigure 1: six indicators;

III. Temporal variation (spline) in baseline transmission rates: five continuous variables (see below for details);

IV. Interaction between spatial region and temporal variation: 30 continuous variables;

V. School opening: one indicator (0: weeks covering January 15 – February 15 and July 1 – August 31, 1:otherwise);

VI. Linear and quadratic population density: two continuous variables;

VII. Climate indices, six continuous and two indicator variables.

- Linear and quadratic temperature: two continuous variables;
- Linear and quadratic relative humidity: two continuous variables;
- Linear and quadratic wind speed: two continuous variables;
- Two indicator for precipitation ((0,0):low, (1,0):medium, (0,1):high).

For continuous main factors we consider both linear and quadratic terms; that is why there are two continuous variables for each of population density, temperature, relative humidity and wind speed.

For reservoir-to-human transmission, we assume the transmission rate is constant over time and is only affected by age-gender strata and spatial regions (appearing as $\boldsymbol{X}_i^k(t)'\boldsymbol{\alpha}_S$). The reason is that most covariate effects on this transmission route become non-identifiable as a result of the insufficient number of prefectures that were only exposed to the unobserved sources. We assume the same age and gender effects on changing susceptibility to infection for the two types of transmission routes, but allow the effects of spatial regions to differ.

We compare the following three regression models that differ in the covariates for transmission risk from observed sources (appearing as):

- Model A: age-gender strata, spatial region, temporal variation and school opening. This model serves as the baseline for model comparison.
- Model B: covariates in model A plus spatio-temporal interactions.
- Model C: covariates in model A plus population density and climate indices.

Model B is descriptive in the sense that it describes the spatial and temporal heterogeneity in transmission rates but without etiological interpretation. Model C is exploratory as it tries to explain most of the spatiotemporal heterogeneity by covariates that vary across prefectures and/or time. The model comparison statistics, AIC and BIC, are given in eTable 2 for the three models. Both complex models are significant improvements on the baseline model A. Model B and model C are nearly equivalent in terms of goodness-of-fit, but both AIC and BIC are in favor of the latter. The relative transmission rates (exponential of the related covariate effects) over time in different spatial regions are plotted in eFigure 12 for model B and in eFigure 13 for model C. The patterns of temporal variation based on model C look much more uniform across regions than those based on model B, suggesting that variation in population density and climate indices may not fully account for the spatiotemporal heterogeneity in the data. However, we use model C for primary inference because it is relatively parsimonious and provides reasonable goodness-of-fit. All terms in model C are statistically significant based on the likelihood-ratio test. Interactions among climate indices do not improve the goodness-of-fit of model C substantially and are therefore not included in the final model.

Additional results

Distribution of case severity ratio, case fatality ratio and severe case fatality ratio within infants

It was reported in the 1998 outbreak of Taiwan that, among infants with severe complications, those aged ≤6 months had lower mortality rate, likely due to maternally acquired immunity.[3,4] However, antibodies to EV71 were found to wane rapidly after one month of age.[6] For the Chinese outbreaks, we further broke down case severity ratio, case fatality ratio and severe case fatality ratio by bi-monthly periods for infants ≤11 months (eTable 3). In 2009, infants cases in their fifth or sixth months had higher case severity ratio, case fatality ratio and severe case fatality ratio than those in other months. Infants in their first two months appeared to have lower risks of these unfavorable outcomes.

Transmission model: effects of age and gender on infectiousness

We explored the association of infectiousness of cases with age and gender using the transmission model. The covariate effects on infectiousness are often difficult to estimate,[7] and are therefore not included in the primary analysis. Nevertheless, the fact that case numbers were observed in the age-gender stratum within prefectures makes tentative estimation possible. We found that the 0.0-0.9 year olds and the 3.0-5.9 year olds were 43% (95% CI:38%, 47%) and 35% (95% CI:32%, 38%) more infectious, while the two older groups were 17% (95% CI:10%, 23%) and 63% (95% CI:52%, 72%) less infectious, than the 1.0 2.9 year olds. The reason for the strong infectiousness of infants is not clear, but probably related to close parental contacts with sick infants and their diapers. Male cases were 13% (95% CI: 10%, 17%) more infectious than female cases. Our estimates for covariate effects on susceptibility are robust to the addition of covariate effects on infectiousness to

the model. For example, the OR for school opening versus closure changed slightly from 1.155(95% CI: 1.149, 1.161) to 1.141 (95% CI: 1.135, 1.148). In particular, no change was observed in the estimates for the effects of age group and gender on susceptibility.

Additional discussion

Viral shedding and the infectious period

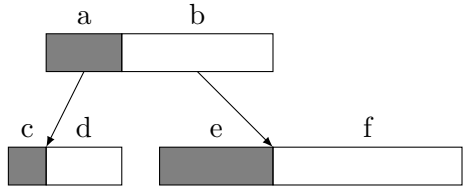
Viral shedding offers an alternative way to determine the infectious period; however, systematic and reliable assessment of the viral shedding period for enteroviruses circulating in recent years is rare. Historical studies suggested that oral/respiratory shedding occurs soon after symptom onset and lasts for up to 2 weeks; fecal shedding occurs later but can be detectable for up to a few months.[9,10,11] In a recent viral shedding study on 34 patients infected with EV71, positive rates in throat swabs collected after symptom onset dropped to 40% in the second week, to 20% in the third, and to nearly 0 in the fourth; the positive rates in feces dropped to 10–20% two weeks after and stayed at that level for more than a month.[12] How levels of viral shedding are associated with transmission is not clear. The short serial intervals in the Singapore outbreak may suggest that sufficient viral shedding levels for human-to-human transmission be present only in the first few weeks after symptom onset. Long-term fecal shedding is partially accounted for in our model via the local reservoir.[8]

How would asymptomatic infections and under-reporting of symptomatic cases affect the estimation of transmission rates

We do not have any information about asymptomatic infections in China, and therefore these cases are not included in the model. Asymptomatic infections exist for many infectious diseases, and their role in the dynamics of transmission can only be fully accounted for with appropriate laboratory

tests. Using only symptomatic cases for analysis is a fairly common practice in transmission studies, for example, in influenza household studies,[13,14] and the estimates of transmissibility measures are interpreted as measuring transmissibility of pathogens that cause symptoms. These analyses all have an implicit assumption that asymptomatic infections do not generate symptomatic new cases. This assumption likely does not hold for the enteroviruses in China, as the model suggested that the mixing of school-age children, who might mostly be asymptomatic after infection, played a role in the spread of the disease. We use a simplified version of our model to explain what assumptions are needed for these analyses, including ours in this paper, to be valid.

Consider the following diagram in which the white and gray bars indicate symptomatic and asymptomatic infections respectively. The two rows represent two consecutive generations, with the second row generated by the first. Specifically, a asymptomatic infectors generate c asymptomatic and d symptomatic new infections, and b symptomatic infectors generate e asymptomatic and f symptomatic new infections.



Let λ be the transmission rate from a symptomatic case, and θ be the relative infectiousness of asymptomatic infections. Assuming a typical branching process with a Poisson offspring distribution and that λ is the only unknown quantity, the likelihood for λ based on this one generation of transmission is given by

$$L = \frac{(a\theta\lambda)^{c+d}\exp(-a\theta\lambda)}{(c+d)!}\frac{(b\lambda)^{e+f}\exp(-b\lambda)}{(e+f)!}. \tag{5}$$

The maximum likelihood estimate (MLE) for λ is $\hat{\lambda} = (c+d+e+f)/(\theta a+b)$. Let α_0 and α_1 be the odds of being asymptomatic given infection by an asymptomatic and a symptomatic infector, respectively, such that $a/b = \alpha_0$ and $c/d = \alpha_1$. It is plausible to assume $\alpha_0 \geq \alpha_1$, i.e., a new

case generated by an asymptomatic infector is equally or more likely to be asymptomatic. For the parent generation whose ancestors are also a blend of asymptomatic and symptomatic infections, we have $e/f = \bar{\alpha}$, where $\alpha_0 \geq \bar{\alpha} \geq \alpha_1$. The MLE of λ is then

$$\hat{\lambda} = \big((1+\alpha_0)d + (1+\alpha_1)f\big)/\big((1+\theta\bar{\alpha})b\big) \begin{cases} \geq \big(\frac{1+\alpha_0}{1+\bar{\alpha}}d + \frac{1+\alpha_1}{1+\bar{\alpha}}f\big)/b, & \theta \leq 1, \\ \leq \big(\frac{1+\alpha_0}{1+\bar{\alpha}}d + \frac{1+\alpha_1}{1+\bar{\alpha}}f\big)/b, & \theta \geq 1, \end{cases} \tag{6}$$

where $\frac{1+\alpha_0}{1+\bar{\alpha}} \geq 1$ and $\frac{1+\alpha_1}{1+\bar{\alpha}} \leq 1$. $\theta > 1$ is possible because asymptomatic infections, who are generally unaware of their own infection status, may have more frequent contacts with the susceptible population than symptomatic cases.

Ignoring the asymptomatic infections, the MLE of λ would be $\hat{\lambda}^\star = (d+f)/b$. We are interested in how the relationship between $\hat{\lambda}$ and $\hat{\lambda}^\star$ is affected by the actually unknown quantities: α_0, α_1, $\bar{\alpha}$ and θ. The dependency of this relationship on α_0, α_1, $\bar{\alpha}$ can be further reduced such that

$$\hat{\lambda} = \big((1+\alpha_0)d + (1+\alpha_1)f\big)/\big((1+\theta\bar{\alpha})b\big) \begin{cases} \geq \hat{\lambda}^\star, & \theta \leq 1, \\ \leq \hat{\lambda}^\star, & \theta \geq 1. \end{cases} \tag{7}$$

Expression (7) holds under either of two conditions: (1) when $\alpha_0 = \alpha_1 = \bar{\alpha}$, i.e., symptom status of infectees is independent of that of the infector, and (2) when the epidemic has passed its initial phase and the proportion of asymptomatic infections among all infections is stable in each generation, such that $a/b = (c+e)/(d+f) = \bar{\alpha}$ and hence $(1+\alpha_0)d + (1+\alpha_1)f = (1+\bar{\alpha})(d+f)$. The second condition is a result of the deterministic version of the assumed branching process. The deterministic model is given by

$$\begin{aligned} Y_n^0 &= \alpha_0\theta\lambda Y_{n-1}^0 + \alpha_1\lambda Y_{n-1}^1 \\ Y_n^1 &= \theta\lambda Y_{n-1}^0 + \lambda Y_{n-1}^1 \end{aligned} \tag{8}$$

where Y_n^0 and Y_n^1 are the numbers of asymptomatic and symptomatic infections at the n^{th} generation. The rate of generating new symptomatic infections is λ for a symptomatic infector and $\theta\lambda$ for

an asymptomatic infector. The rate of generating new asymptomatic infections is $\alpha_1\lambda$ for a symptomatic infector and $\alpha_0\theta\lambda$ for an asymptomatic infector. The deterministic system of $Z_n = Y_n^0/Y_n^1$ derived from (8) has the form

$$Z_n = G(Z_{n-1}|\theta, \alpha_0, \alpha_1) = \frac{\alpha_0\theta Z_{n-1} + \alpha_1}{(\theta Z_{n-1} + 1)}$$

with an equilibrium point $\bar{\alpha} = \frac{1}{2}\Big\{\alpha_0 - \theta^{-1} + \big((\alpha_0 - \theta^{-1})^2 + 4\alpha_1\theta^{-1}\big)^{1/2}\Big\}$ satisfying $\bar{\alpha} = G(\bar{\alpha})$. This equilibrium point is stable, i.e., $0 \leq \frac{d}{dZ}G(Z|\theta, \alpha_0, \alpha_1)|_{Z=\bar{\alpha}} < 1$, for at least a reasonable range of the parameters that we have tested, i.e., $0 \leq \alpha_1 \leq \alpha_0 \leq 2$ and $0 \leq \theta \leq 2$. Within this range, Z_n quickly converges, no matter what the initial numbers Y_1^0 and Y_1^1 are, to $\bar{\alpha}$. This result assures that how well $\hat{\lambda}^\star$ approximates $\hat{\lambda}$ largely depends on θ. When $\theta < 1$ or $\theta > 1$, $\hat{\lambda}^\star$ provides a lower or an upper bound for $\hat{\lambda}$. Their equality holds when $\theta = 1$, i.e., asymptomatic infections are as infectious as symptomatic cases.

The likelihood for the surveillance data is approximately the product of the likelihoods in the form of (5) for all underlying generations over all prefectures and the epidemic duration. Hence, to generalize our conclusion from one generation of transmission to the whole epidemic, we need one additional assumption that the unknown parameters α_0, α_1 and θ are spatially and temporally homogeneous.

The situation of under-reported symptomatic cases on the estimation of transmissibility is simpler than that of asymptomatic infections, because it is reasonable to assume that unreported symptomatic cases are as infectious and pathogenic as reported symptomatic cases. Replacing asymptomatic infections with unreported symptomatic cases in above discussion, we would have $\hat{\lambda}^\star = \hat{\lambda}$.

How would the absence of individual virological data affect the estimation of transmission rates and effects of risk factors

With the absence of individual virological data, we have to model multiple co-circulating pathogens as a single one. However, the estimated transmission rates can be viewed as the average rate of the family of enteroviruses that cause HFMD. Under the assumption that all pathogens share similar covariate effects, the estimated risk ratios also have an interpretation of average effects.

Consider a simplified version of our model, a branching process with two pathogens and one binary covariate. For simplicity, consider only one generation of transmission (branching). Suppose that m_{hr} existing cases of pathogen h and covariate value r generated n_{hr} new cases, $h = 1, 2$ and $r = 0, 1$. We assume a Poisson offspring distribution such that $n_{hr} \sim \text{Poisson}(m_{hr}\lambda_h\theta_h^r)$, where λ_h is the baseline transmission rate and θ_h is the risk ratio of $r = 1$ to $r = 0$. The MLEs for pathogen-specific transmission rates and risk ratios are given by $\hat{\lambda}_h = n_{h0}/m_{h0}$, and $\hat{\theta}_h = n_{h1}m_{h0}/(m_{h1}n_{h0})$, $h = 1, 2$. Without the pathogen-specific case numbers, the MLEs would be $\hat{\lambda} = (n_{10} + n_{20})/(m_{10} + m_{20})$, and $\hat{\theta} = (n_{11} + n_{21})(m_{10} + m_{20})/\big((m_{11} + m_{21})(n_{10} + n_{20})\big)$.

With above setting, we have

$$\hat{\lambda} = \frac{m_{10}}{m_{10} + m_{20}}\hat{\lambda_1} + \frac{m_{20}}{m_{10} + m_{20}}\hat{\lambda_2},$$

an average of the pathogen-specific transmission rates weighted by the population sizes of the pathogens in the parent generation. The MLE for θ can be rearranged as

$$\hat{\theta} = \left(\frac{\hat{\lambda}_1 m_{11}}{\hat{\lambda}_1 m_{11} + \hat{\lambda}_2 m_{21}}\hat{\theta}_1 + \frac{\hat{\lambda}_2 m_{21}}{\hat{\lambda}_1 m_{11} + \hat{\lambda}_2 m_{21}}\hat{\theta}_2\right) \times \frac{\frac{m_{11}}{m_{11}+m_{21}}\hat{\lambda}_1 + \frac{m_{21}}{m_{11}+m_{21}}\hat{\lambda}_2}{\frac{m_{10}}{m_{10}+m_{20}}\hat{\lambda}_1 + \frac{m_{20}}{m_{10}+m_{20}}\hat{\lambda}_2}.$$

The first term on the right side is a weighted average of pathogen-specific risk ratios. Under the assumption that the risk factor effects are similar between the two pathogens, the distributions of risk factor levels would also be similar between the two pathogens in any generation, or equivalently, the distributions of pathogens would be similar across risk factor levels, i.e., $\frac{m_{11}}{m_{11}+m_{21}} \approx \frac{m_{10}}{m_{10}+m_{20}}$,

which implies that the second term approximates 1. Therefore, $\hat{\theta}$ measures the average risk ratio of the family of co-circulating pathogens under the assumption that the risk factor has fairly similar effects on all pathogens.

References

[1] Becker NG. Estimation for discrete time branching processes with applications to epidemics. *Biometrics.* 1977; **33:** 515–522.

[2] Becker NG. On a general epidemic model. *Theoretical Population Biology.* 1977; **11:** 23–36.

[3] Ho M, Chen ER, HSU KH, et al. An epidemic of enterovirus 71 infection in Taiwan. *New England Journal of Medicine.* 1999; **341:** 929–935.

[4] Luo ST, Chiang PS, Chao AS, et al. Enterovirus 71 maternal antibodies in infants, Taiwan. *Emerging Infectious Disease.* 2009; **15:** 581–584.

[5] Chang LY, King CC, Hsu, KH, et al. Risk factors of enterovirus 71 infection and associated hand, foot, and mouth disease/herpangina in children during an epidemic in Taiwan. *Pediatrics.* 2002; **109:** e88–e93.

[6] Ooi EE, Phoon MC, Ishak B, et al. Seroepidemiology of human enterovirus 71, Singapore. *Emerging Infectious Disease.* 2002; **8:** 995–997.

[7] Yang Y, Longini IM, Halloran ME. Design and Evaluation of Prophylactic Intervention Using Infectious Disease Incidence Data from Close Contact Groups. *Applied Statistics.* 2006; **55:** 317–330.

[8] Goh KT, Doraisingham S, Tan JL, et al. An outbreak of hand, foot and mouth disease in Singapore. *Bulletin of the World Health Organization.* 1982; **60:** 965–969.

[9] Kogon A, Spigland I, Frothingham, TE, et al. The virus watch program: a continuing surveillance of viral infections among metropolitan New York families. *American Journal of Epidemiology.* 1969; **89:** 51–61.

[10] Cheng LL, Ng PC, Chan PKS, et al. Probable intrafamilial transmission of Coxsackievirus B3 with vertical transmission, severe early-onset neonatal hepatitis, and prolonged viral RNA shedding. *Pediatrics.* 2006; **118:** e929–e933.

[11] Chung YW, Huang YC, Chang LY, et al. Duration of enterovirus shedding in stool. *Journal of Microbiology.* 2001; **34:** 167–170.

[12] Han J, Ma XJ, Wan JF, et al. Long persistence of EV 71 specific nucleotides in respiratory and feces samples of the patients with Hand-Foot-Mouth Disease after recovery. *BMC Infectious Diseases.* 2010; **10:** 178.

[13] Hayden FG, Belshe R, Villanueva C, et al. Management of influenza in households: a prospective randmized comparison of oseltamivir treatment with or without postexposure prophylaxis. *Journal of Infectious Diseases.* 2004; **189:** 440–449.

[14] Cauchemez S, Donnelly CA, Reed C, et al. Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. *New England Journal of Medicine.* 2009; **361:** 2619–2627.

eTable 1: *Effects of Region on Susceptibility to Transmission Risk from Local Reservoir to Human in the 2008–2009 HFMD Epidemics in China. Presented are Risk Ratios (RR) and 95% Confidence Intervals from the Aggregate Susceptible-Infected-Recovered Model.*

		Assumption about Infectious Period					
Risk		$(1, 0.2, 0)$		$(1, 0.5, 0)$		$(1, 0.6, 0.2)$	
Factor		RR	95% CI	RR	95% CI	RR	95% CI
Region[1]	CN						
	CS	1.25	1.11, 1.42	1.16	1.01, 1.32	1.08	0.94, 1.25
	S	2.02	1.78, 2.30	1.97	1.73, 2.25	1.96	1.71, 2.25
	NE	0.40	0.33, 0.49	0.34	0.28, 0.42	0.26	0.20, 0.34
	SW	0.34	0.29, 0.41	0.29	0.24, 0.36	0.24	0.19, 0.30
	CW	0.46	0.40, 0.55	0.44	0.37, 0.52	0.41	0.35, 0.49
	W	0.10	0.080, 0.13	0.094	0.073, 0.12	0.071	0.052, 0.095

1 CN= Central North, CS=Central South, S=South, NE=Northeast, SW=Southwest, CW= Central West, W=West

eTable 2: *Goodness-of-Fit Statistics of the Three Candidate Regression Models for Human-to-Human Transmission.*

Model	Log-likelihood	Number of Parameters	AIC	BIC
A	-489218	26	978488	978764
B	-487127	56	974366	974960
C	-487138	36	974348	974730

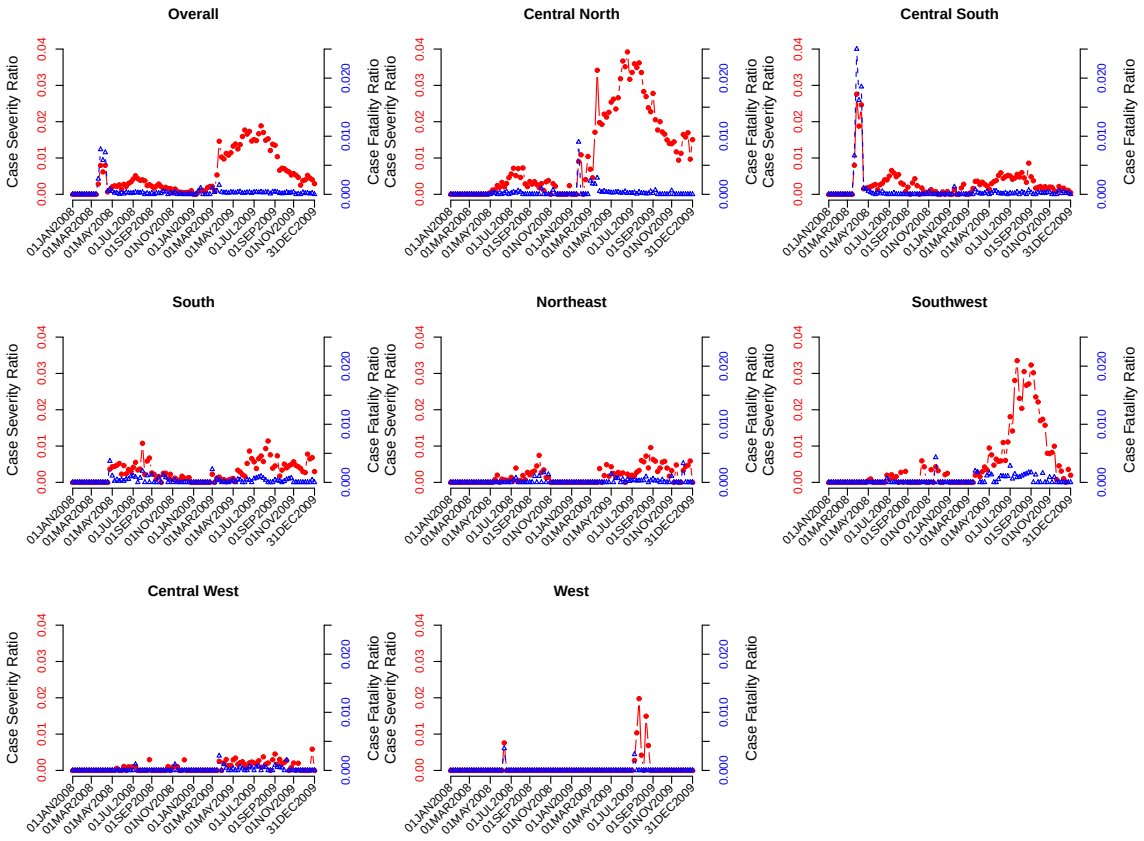
eTable 3: *Reported Numbers of Cases, Severe Cases and Deaths of the HFMD by Age Group among Infants ≤11 Months. Case Severity Ratio (CSR), Case Fatality Ratio (CFR) and Severe Case Fatality Ratio (SCFR) are also Presented.*

Year	Age Group	Number of Cases	Number of Severe Cases	Number of Fatal Cases	CSR (%)	CFR (%)	SCFR (%)
2008	1–2	591	2	0	0.34	0.0	0.0
	3–4	1376	8	1	0.59	0.07	12.50
	5–6	2822	18	4	0.64	0.14	22.22
	7–8	6564	29	6	0.44	0.09	20.69
	9–10	10237	34	11	0.33	0.11	32.35
	11	4831	30	2	0.62	0.04	6.67
2009	1–2	1175	18	0	1.53	0.0	0.0
	3–4	2720	76	1	2.80	0.04	1.32
	5–6	5938	177	13	2.98	0.22	7.34
	7–8	14236	329	17	2.31	0.12	5.17
	9–10	23129	489	14	2.11	0.06	2.86
	11	12398	267	12	2.15	0.10	4.49

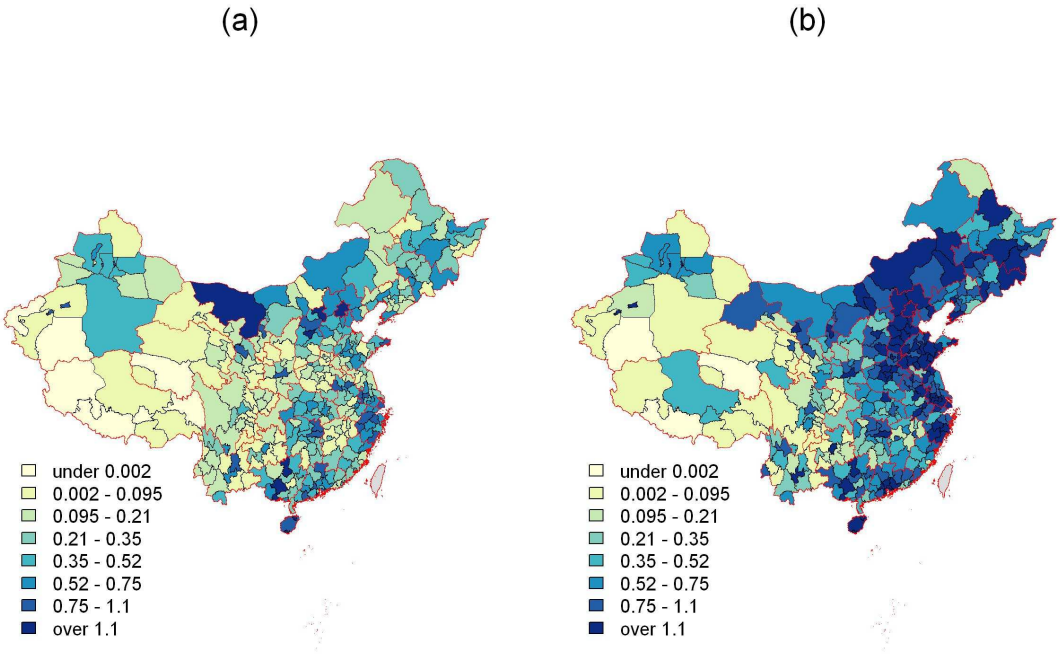
Animation caption

Dynamic changes in prefecture-specific reported weekly attack rates of HFMD during 2008 and 2009 in China. Attack rates are displayed as the number of reported cases per million persons per week. Prefectures with no data available are colored in grey. Red lines indicate provincial boundaries. (The animation is available as another supplementary file with the online version of this paper.)

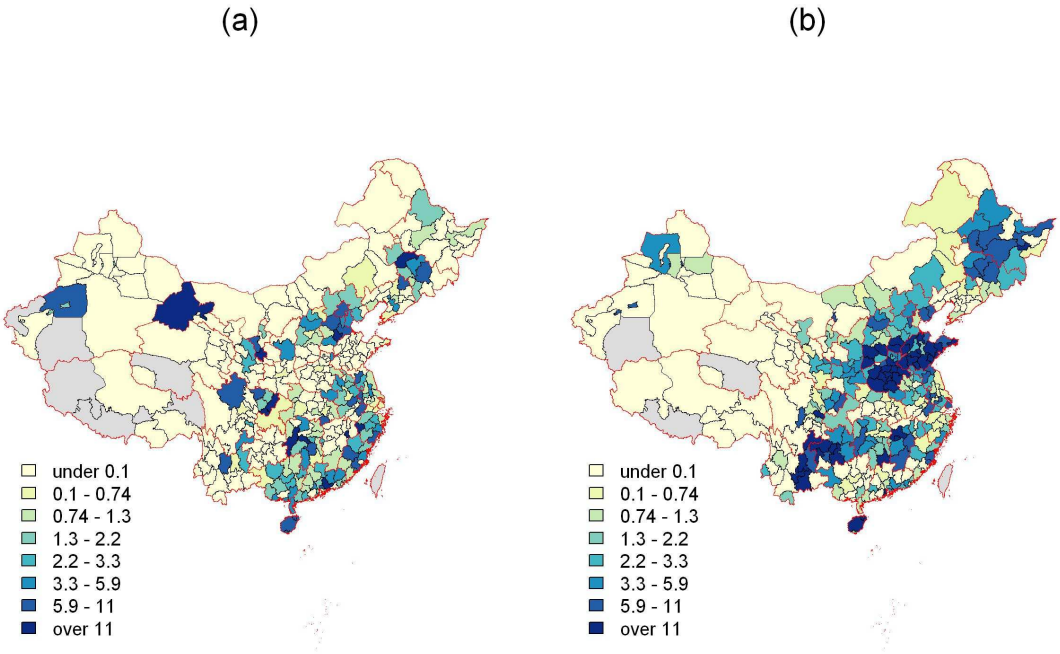
eFigure 1: The provinces in mainland China are divided into seven geographic regions. **Central North**: Beijing, Tianjing, Hebei, Henan, Shandong, Shanxi; **Central South**: Hunan, Hubei, Jiangxi, Anhui, Jiangsu, Shanghai, and Zhejiang; **South**: Guangxi, Guangdong, Fujian, and Hainan; **Northeast**: Heilongjiang, Liaoning, and Jilin; **Southwest**: Sichuan, Yunnan, Guizhou, and Chongqing; **Central west**: Shaanxi, Ningxia, Gansu, and Inner Mongolia; **West**: Tibet, Xinjiang, and Qinghai. The division is determined by geographic location, population density, socioeconomic development, ethnicity, climate, etc.



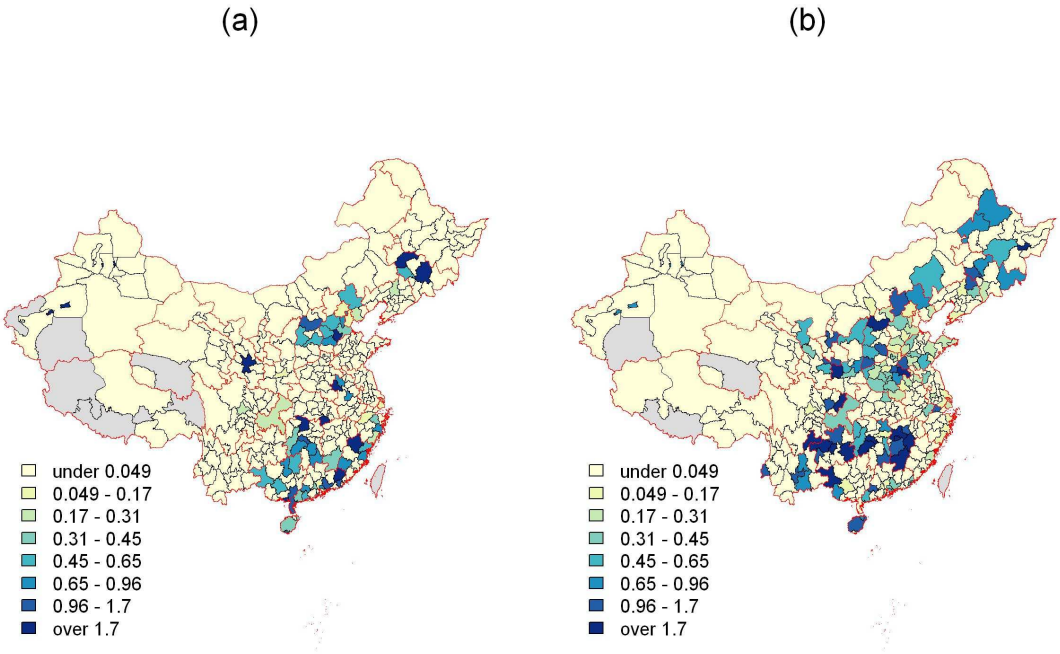
eFigure 2: Variation of case severity rates (red dots) and case fatality rates (blue triangles) over time for the 2008-2009 HFMD epidemics in China and its seven geographic regions. The sharp spikes of both the case severity rate and case fatality rate during March and April, 2008 in the Central South region were due to a large local outbreak in Fuyang, Anhui Province.



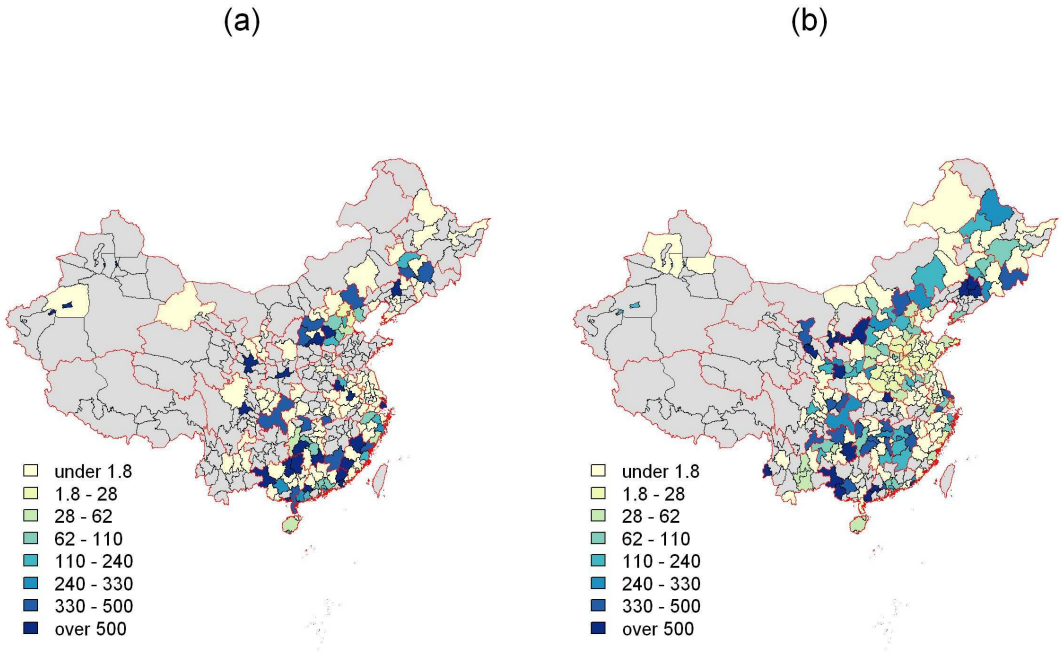
eFigure 3: Geographic distribution of prefecture-specific attack rates (number of reported cases / 1000 people) of the hand, foot and mouth disease in (a) 2008 and (b) 2009 in China. Prefectures with no data available are colored in grey. Red lines indicate provincial boundaries.



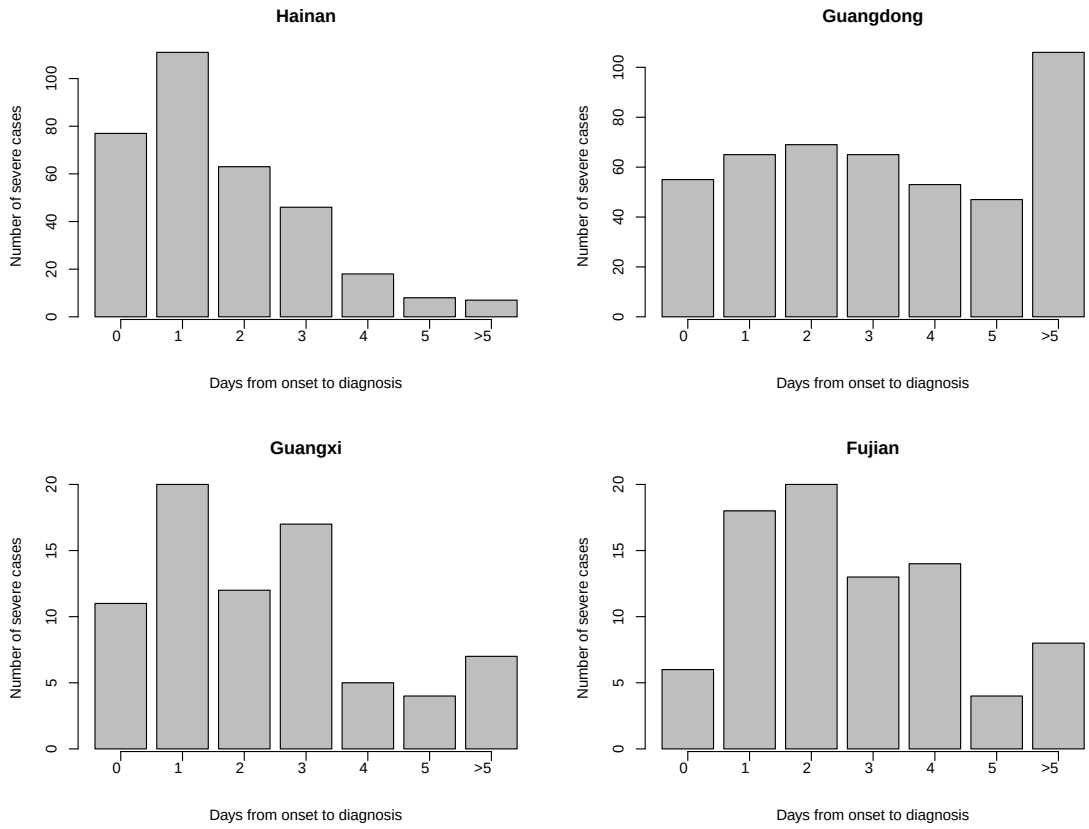
eFigure 4: Geographic distribution of prefecture-specific case severity ratios (number of severe cases / 1000 reported cases) of the hand, foot and mouth disease in (a) 2008 and (b) 2009 in China. Prefectures with no cases reported or no data available are colored in grey. Red lines indicate provincial boundaries.



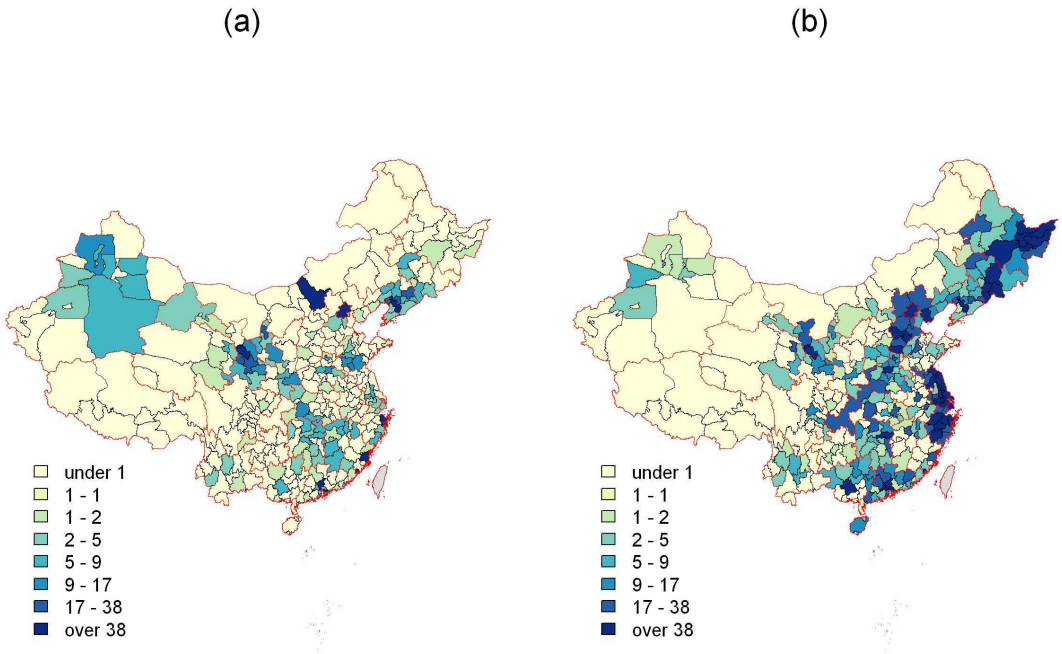
eFigure 5: Geographic distribution of prefecture-specific case fatality ratios (number of fatal cases / 1000 reported cases) of the hand, foot and mouth disease in (a) 2008 and (b) 2009 in China. Prefectures with no cases reported or no data available are colored in grey. Red lines indicate provincial boundaries.



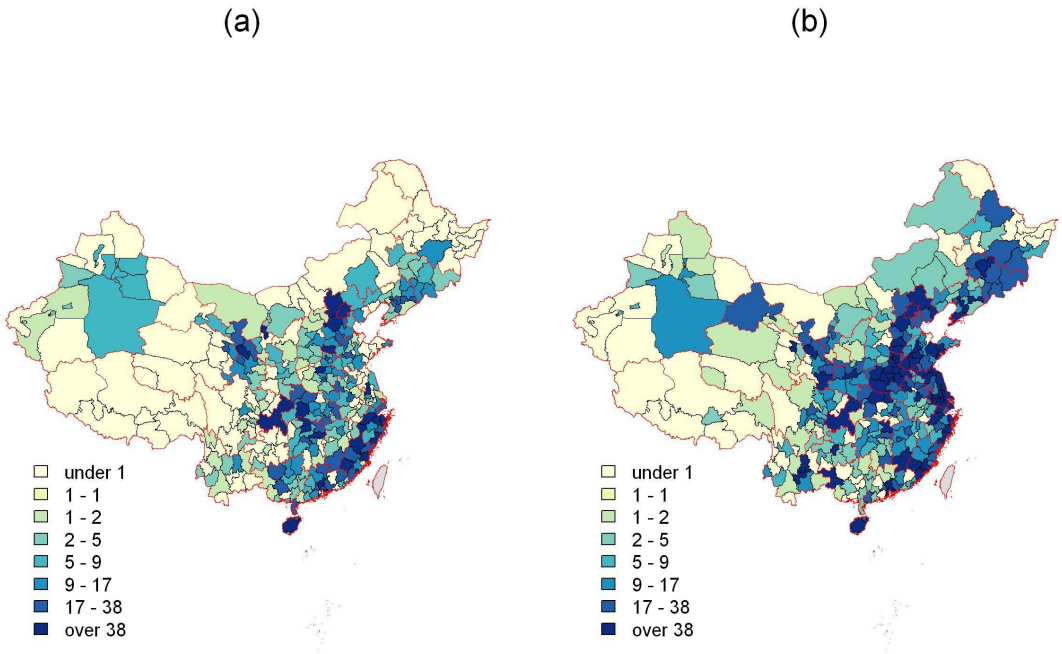
eFigure 6: Geographic distribution of prefecture-specific severe case fatality ratios (number of fatal cases / 1000 reported severe cases) of the hand, foot and mouth disease in (a) 2008 and (b) 2009 in China. Prefectures with no cases reported or no data available are colored in grey. Red lines indicate provincial boundaries.



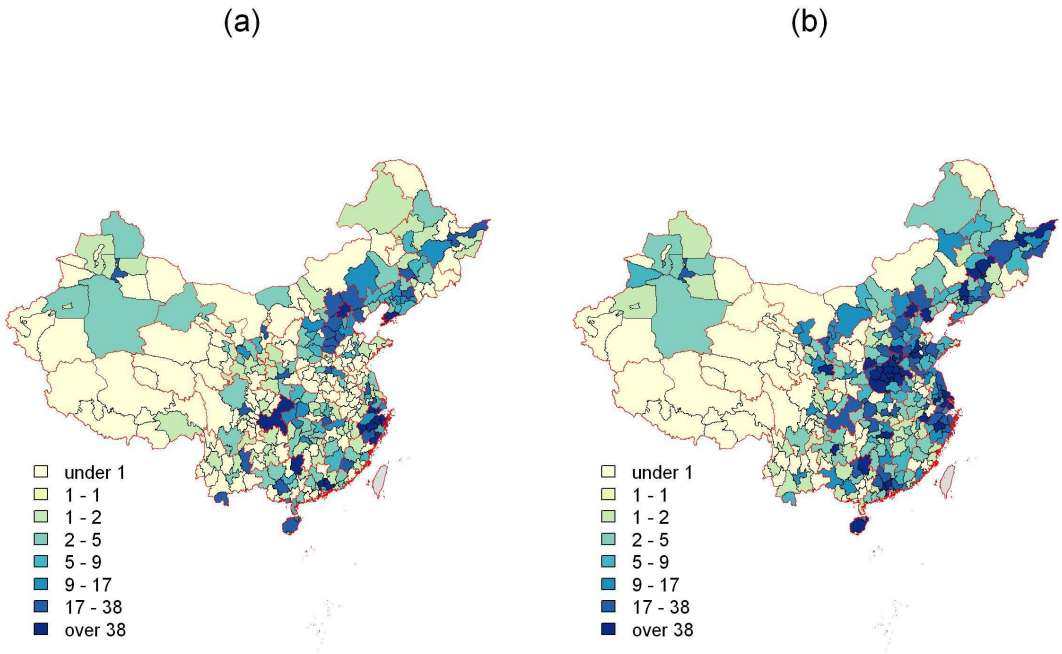
eFigure 7: Distributions of delay from onset to diagnosis (in days) among severe cases during the 2008–2009 epidemics in the four provinces of the South region: Fujian Guangdong, Guangxi, and Hainan.



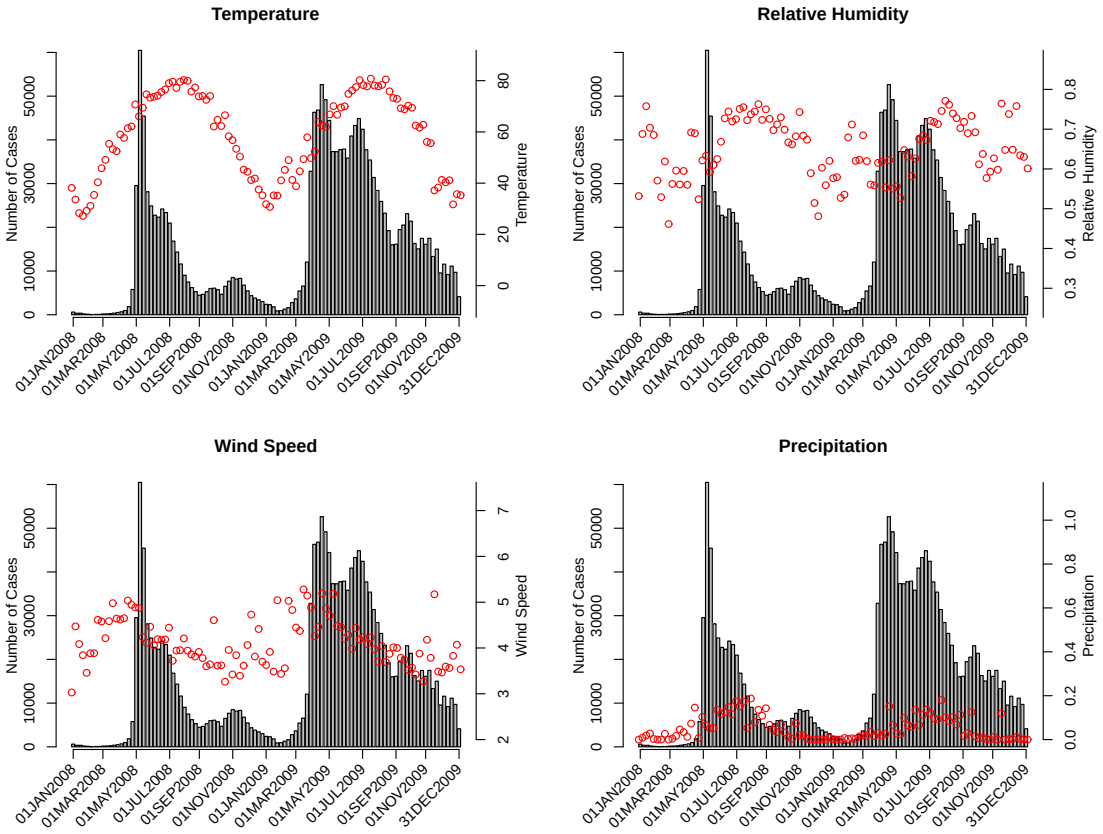
eFigure 8: Geographic distribution of prefecture-specific numbers of HFMD cases lab-confirmed as CA16 in (a) 2008 and (b) 2009 in China. Prefectures with no data available are colored in grey. Red lines indicate provincial boundaries.



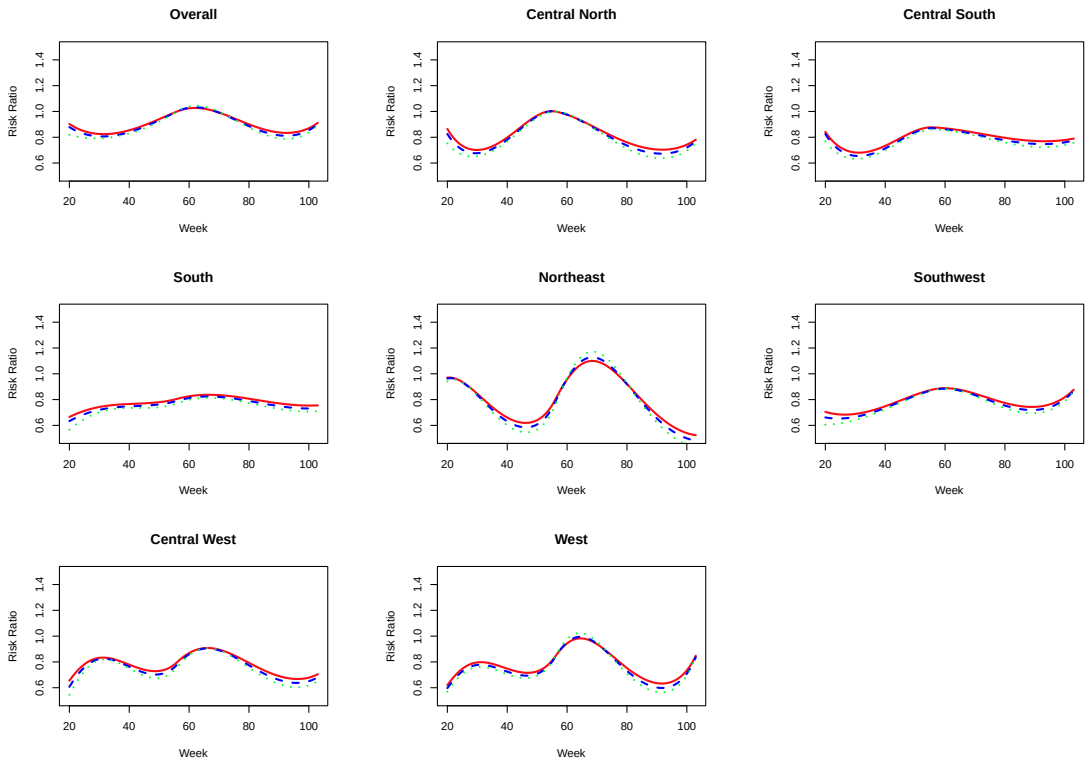
eFigure 9: Geographic distribution of prefecture-specific numbers of HFMD cases lab-confirmed as EV71 in (a) 2008 and (b) 2009 in China. Prefectures with no data available are colored in grey. Red lines indicate provincial boundaries.



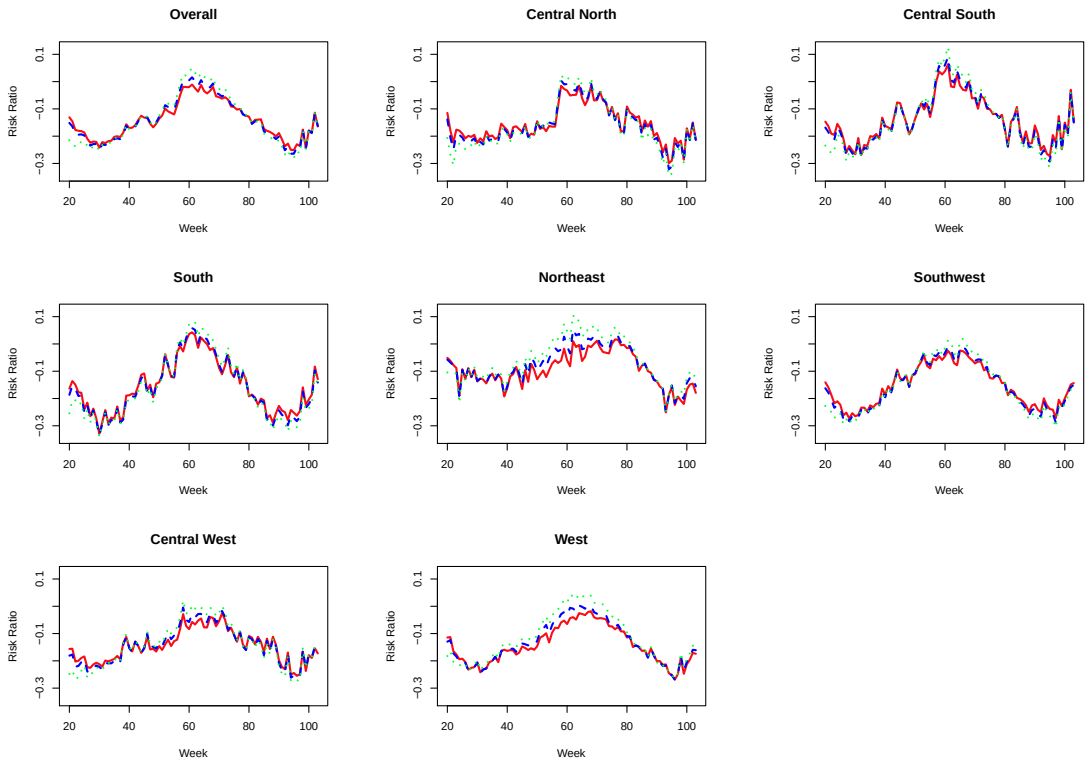
eFigure 10: Geographic distribution of prefecture-specific numbers of HFMD cases lab-confirmed as other enteroviruses in (a) 2008 and (b) 2009 in China. Prefectures with no data available are colored in grey. Red lines indicate provincial boundaries.



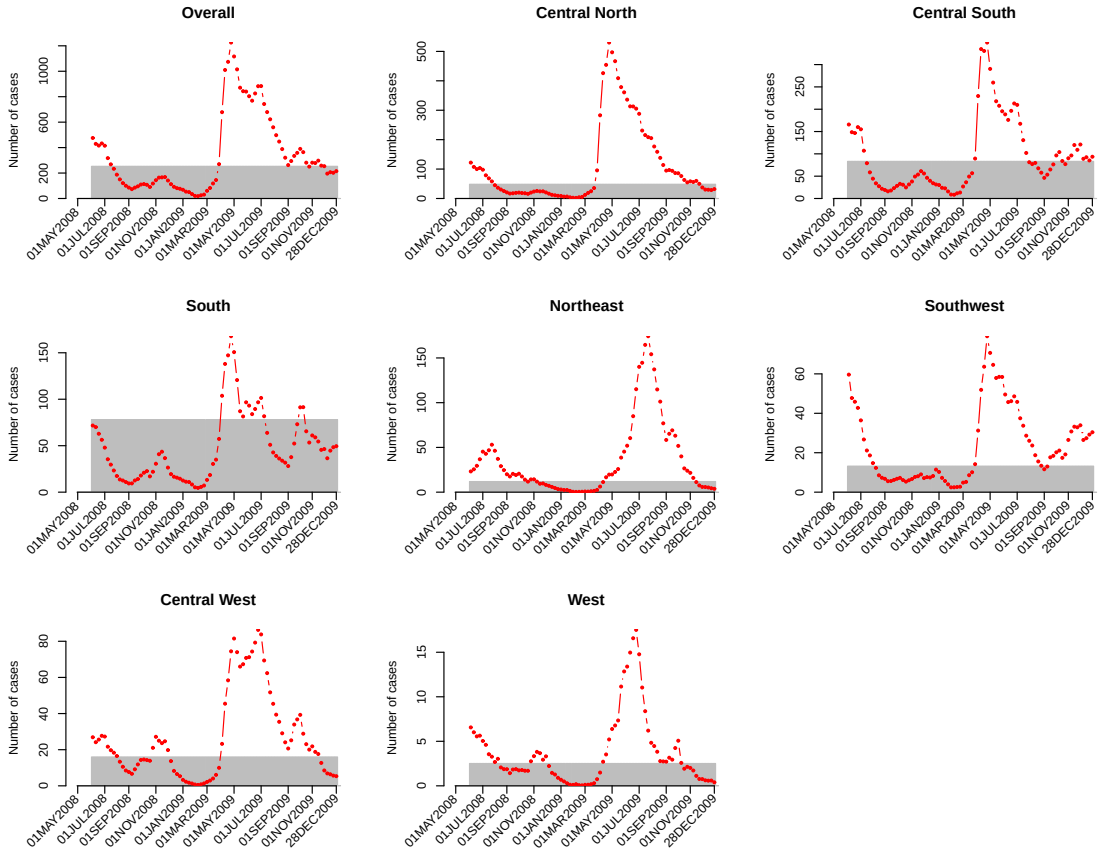
eFigure 11: Weekly average temperatures, relative humidity, wind speed and precipitation plotted over time and overlain with the HFMD epicurve during 2008-2009 in China.



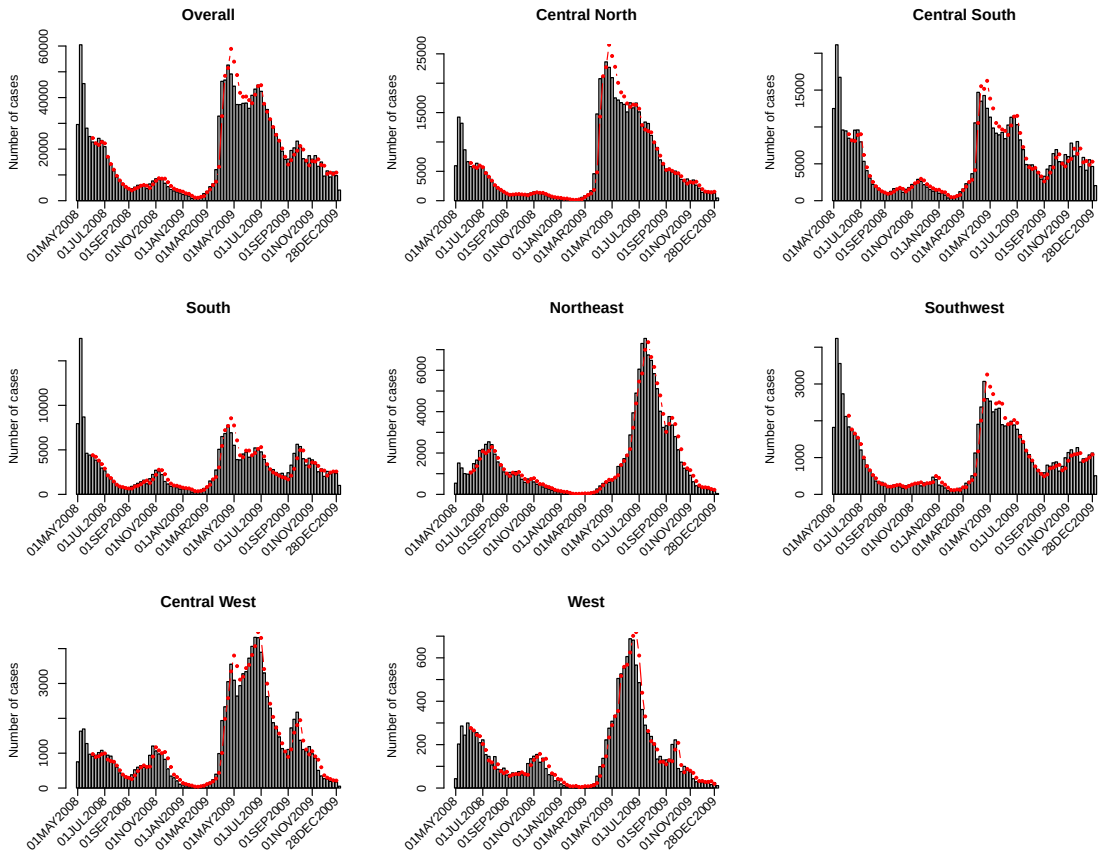
eFigure 12: Spatial and temporal heterogeneity in baseline human-to-human transmission rates, based on the model adjusted for age group, gender, school closure, geographic region, temporal splines, and interaction between geographic region and temporal splines. The red solid, blue dashed, and green dotted lines correspond to the assumptions about the infectious period, (1, 0.2, 0), (1, 0.5, 0) and (1, 0.6, 0.2), respectively.



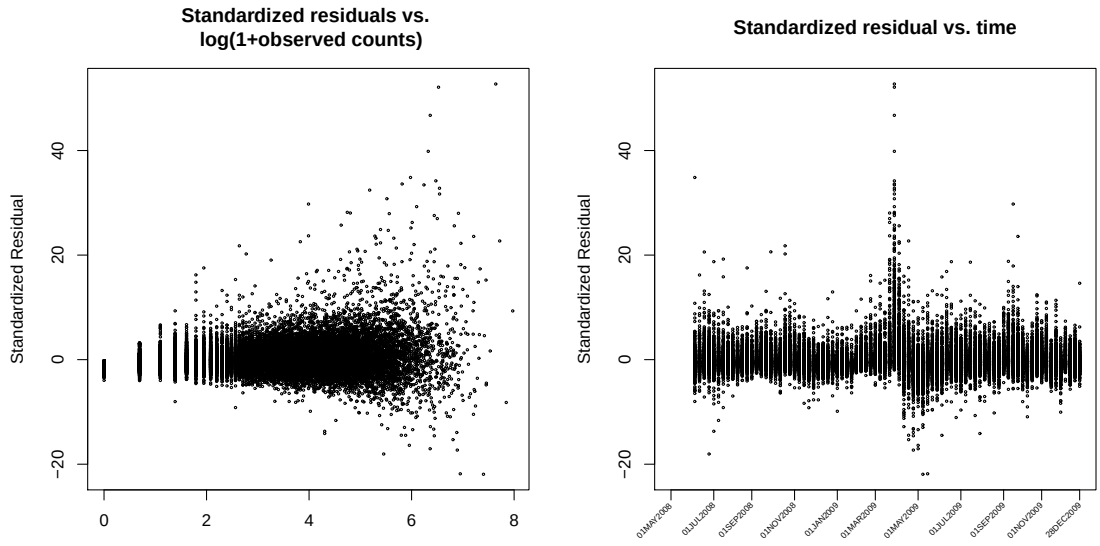
eFigure 13: Spatial and temporal heterogeneity in baseline human-to-human transmission rates, based on the model adjusted for age group, gender, school closure, geographic region, temporal splines, population density, temperature, relative humidity, wind speed, and precipitation. The red solid, blue dashed, and green dotted lines correspond to the assumptions about the infectious period, (1, 0.2, 0), (1, 0.5, 0) and (1, 0.6, 0.2), respectively.



eFigure 14: Model-predicted weekly mean numbers of infection generated by unobserved local reservoir (grey area) and observed between-prefecture human-to-human transmissions (red dots) for the overall nation and seven geographic regions for the 2008-2009 HFMD epidemics in China.



eFigure 15: Model-predicted (red dots) vs. observed weekly numbers of cases (grey bars) for the overall nation and seven geographic regions for the 2008-2009 HFMD epidemics in China.



eFigure 16: Plots of standardized residuals versus observed number of cases (left) and the timeline (right), each circle representing a prefecture and a week.