SI Appendix:
Vaccination strategies for epidemic cholera in Haiti with implications for the developing world

Dennis L. Chao[1], M. Elizabeth Halloran[1,2], and Ira M. Longini, Jr.[1,2,*]

[1]Center for Statistics and Quantitative Infectious Diseases,
Vaccine and Infectious Disease Division,
Fred Hutchinson Cancer Research Center,
1100 Fairview Ave N
Seattle, WA 98109, USA

[2]Department of Biostatistics,
School of Public Health,
University of Washington,
Seattle, WA 98195, USA

[*]To whom correspondence should be addressed; E-mail: longini@scharp.org.

1 Model description

Individuals in the cholera transmission model are either susceptible, exposed, infectious, or recovered (SEIR), as in [1]. The model runs in discrete time and uses a one-day time step. As in [2], infectious individuals shed *Vibrio cholerae* into the environment. Symptomatic infectious people shed one unit of *V. cholerae* into their community per day, and asymptomatic infectious people excrete m units per day (i.e., asymptomatic individuals are m times as infectious). One study found freshly shed cholera to be up to 700 times more infectious, for between 5 and 18 hours [3]. To represent this effect, symptomatic infectious people also shed one unit of hyperinfectious *V. cholerae* per day, and asymptomatic shed m units. A fraction of environmental *V. cholerae* is lost per day (δ), while the hyperinfectious *V. cholerae* disappears completely the day after it is shed. Susceptible individuals are infected from the environment with a probability of $\beta \times ((B + 100B_H)/N)/(\kappa + (B + 100B_H)/N)$, where N is the number of people in the community, B is the level of *V. cholerae* in their community, B_H is the level of hyperinfectious

V. cholerae in their community, κ is the level at which the probability of infection is 50% of the "maximum" and is related to the infectious dose [2], and β is a scaling parameter. β is the maximum probability of infection from the environment per day and is related to the amount of contaminated water consumed by an individual per day. We assume that an individual's level of exposure to environmental *V. cholerae* is proportional to 1/N, which corrects for different community sizes.

Person-to-person transmission occurs in households using a Reed–Frost-like model. Susceptible people are infected by contact with infectious household members with probability p per infectious household member per day. Therefore, the daily probability of infection is $1 - \prod (1 - pV_i)$, where V_i is the relative infectiousness (maximum of 1.0) of individual i in the household. Recall that asymptomatic people are less infectious ($m < 1$). Vaccination, described below, may also reduce infectiousness.

We assume that once infected, an individual has a 1, 2, 3, 4, or 5 day latency period with probabilities of 40%, 40%, 7%, 7%, and 6%, respectively [1]. After latency, an individual is infectious, with a 20% probability of being symptomatic. The infectious period lasts from 7–14 days, with a uniform distribution. See Fig. 1C.

The level of *V. cholerae* in the environment at time $t + 1$ is:

$$B_{t+1} = B_t + I_S + mI_A - \delta B_t \tag{1}$$

where I_S is the number of symptomatic individuals and I_A the number of asymptomatic. Hyperinfectious *V. cholerae* does not persist in the environment for more than one day, and it depends only on the number of currently infectious individuals:

$$B_H = I_S + mI_A \tag{2}$$

The daily probability of infection from the environment is:

$$p_{\text{environment}} = (1 - H)\beta \frac{(B + 100B_H)/N + B_R + 100B_{RH}}{\kappa + (B + 100B_H)/N + B_R + 100B_{RH}} \tag{3}$$

H is the strength of a public health campaign that effectively reduces a person's exposure. We assume that the level of *V. cholerae* in the environment is scaled by the population size, while the level in the river is not. The probability of infection from the household:

$$p_{\text{household}} = (1 - H)[1 - \Pi_i^{family}(1 - pv_i)] \tag{4}$$

Individuals live in communities of approximately 500 individuals (Fig. 1B). Each community has an independent environmental reservoir of *V. cholerae*, so infectious individuals shed into their home communities and possibly the communities of their workplaces, described below. Thus, individuals within each community have the same exposure to pathogen from the environment. Communities are situated within a rectangular lattice consisting of 1km^2 cells. We use LandScan population estimates (described below) to determine how many individuals

should live within each cell, and we create the appropriate number of communities per cell to accommodate these individuals (Fig. 1).

Of the 28% of individuals who work, 30% work in their own cells, and the rest are assigned to work elsewhere according to a gravity model, described below. If there are multiple communities in a cell, including the home cell, then the worker will be assigned to one of them at random. Shedding is also proportional to the time spent in a community – a worker sheds 30% into his or her work community and 70% into their home community and is exposed to environmental *V. cholerae* levels equal to 70% of that from their home community and 30% from his or her work community. 75% of workers with symptomatic cholera stop working after one day of symptom onset [1], which means that they spend all of their time in their home communities.

There are two modes of long-distance travel in the model. The first represents highway travel. Highway locations were obtained from OpenStreetMap ("primary" or "secondary" highways from `http://labs.geofabrik.de/haiti/2010-11-23-17-44.osm.bz2`) and superimposed on the 1km^2 lattice (Fig. 1A). Individuals who live in cells through which a highway runs can travel to any cell, with a daily probability of 5×10^{-5}, that has at least 2000 individuals between 30km and 200km away on the highway. Distance along a highway is computed using a breadth-first search from the source location. Travel is implemented by selecting random individuals at the source and destination locations and swapping their infection and vaccination status. If either person has been symptomatic for more than one day, travel does not occur for this pair. The actual number of swaps per day is drawn from the binomial $\mathcal{B}(5 \times 10^{-5}, N)$, where N is the population of the source cell. This disperses the epidemic geographically while preserving the number of residents in each community. To account for non-highway travel (e.g., to include smaller roads and domestic flights), we allow an individual to travel to any cell in the country with a daily probability of 5×10^{-6}. Note that this will tend to send people to the large population centers, where most of the people are.

We model vaccine protection through three possible mechanisms: $\mathsf{VE_S}$ is the reduction in susceptibility per infectious contact, $\mathsf{VE_I}$ is the reduction in infectiousness (e.g., amount of shedding per day), and $\mathsf{VE_P}$ is the reduction in probability of becoming ill when infected [4]. None of these parameters have been directly estimated in phase III vaccine trials where the primary endpoint is vaccine efficacy against symptomatic cholera with confirmed infection, $\mathsf{VE_{SP}}$ [4]. If we assume that the vaccine effects are multiplicative, then $\mathsf{VE_{SP}} = 1 - (1 - \mathsf{VE_S})(1 - \mathsf{VE_P})$, and we see that the overall effect of the $\mathsf{VE_{SP}}$ is probably more important than the individual $\mathsf{VE_S}$ and $\mathsf{VE_P}$ components (see *Section 5.2.3* in [4]). We assume the maximum efficacy of the cholera vaccine to be $\mathsf{VE_I}$ = 50%, $\mathsf{VE_P}$ = 64%, and $\mathsf{VE_S} = 0$, so that $\mathsf{VE_{SP}}$ = 64% [5]. In the model, individuals are vaccinated by a single dose of vaccine. To mimic the efficacy of a two-dose vaccine, like Shanchol and Dukorol, 50% of maximum efficacy is reached 10 days after vaccination then relative efficacy rises from 50% to 100% from days 14 to 21, as if two doses were given exactly 2 weeks apart (Fig. S1).

Pre-vaccination is simulated by vaccinating the desired fraction of individuals such that these individuals have maximum protection when the simulation starts. Vaccination can also be reactive, after the simulated epidemic starts. In the simulated reactive vaccination campaigns,

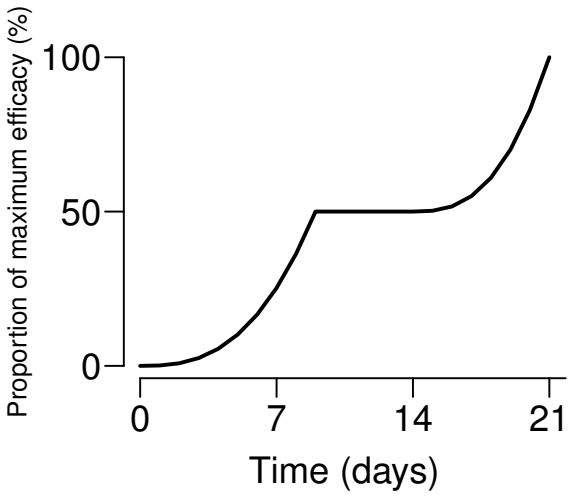

Figure S1: Vaccine efficacy over time in the model. The second dose is automatically administered after 2 weeks.

one can specify the campaign start day and the amount of vaccine made available per day. In a reactive mass vaccination, the simulation attempts to vaccinate all 1km^2 cells with the desired level of coverage (e.g., 70% of individuals). Each day, randomly selected communities are chosen and vaccinated until there is insufficient vaccine to cover additional communities at the desired level. For ring vaccination, one specifies the number of cases that must appear in a 1km^2 cell before that cell is prioritized to be vaccinated. When this cumulative number is reached, this cell is eligible to be vaccinated after a pre-specified number of days (e.g., 5 days). On any given day, randomly selected eligible communities are chosen to be vaccinated to the desired level until vaccine runs out. In the "high exposure" strategy, cells that are not along the river receive vaccine after all cells along the river are vaccinated.

1.1 Gravity model for daily commuting

We use a gravity model to determine how far employed individuals travel to go to work [6, 7]. To determine the distribution of destinations for workers in a given community, we use the following formula from [7]:

$$C_{ij} = \theta \frac{P_i^{\tau_1} P_j^{\tau_2}}{d_{ij}^{\rho}} \tag{5}$$

where C_{ij} is the workflow from community i to j, d_{ij} is the distance between i and j, P_i is the population of community i, θ is a proportionality constant, and τ is used to tune the dispersal. For distances under 119km in the United States, they found $\tau_1 = 0.30$, $\tau_2 = 0.64$, and $\rho = 3.05$. However, mobility might be lower in less developed countries. [6] uses $1/[1+(d/a)^b]$ with a=4km and b=3.8 for rural Thailand, and b=3.0 for England. We chose to use Ferguson

parameter in the Viboud framework, and set ρ to 3.8. In Fig. S2, you can see that workers based in the middle of Port-au-Prince or Mirebalais stay in the city (dark green), those just a couple of kilometers away are a little more dispersed with several venturing into the city (indigo), those who live just east of Port-de-Paix tend to commute west to get to the city (purple), and those in the mountains are a little more spread out (blue). In the model, 30% of employed people are assigned to work in their home cell and the rest are assigned locations by the gravity model confined to a 40km radius.

1.2 Rivers in the model

River locations were obtained from OpenStreetMap (Features tagged with a key of "waterway" and "river", `http://labs.geofabrik.de/haiti/2010-11-23-17-44.osm.bz2`). See Fig. 1A. The rivers are recorded as short disconnected segments, so we linked them together to ensure continuous flow from the mountains to the ocean. Rivers in the model are represented as directional flow between 1km^2 cells. Therefore, multiple rivers entering a single cell are merged.

The river is treated as a second environmental reservoir – individuals shed *V. cholerae* both into their communities and into the river. Infectious individuals shed into the river a proportion of the amount shed into their regular environment. *V. cholerae* in the river moves to the next cell downstream. If there is more than a single adjacent cell downstream, the *V. cholerae* is divided evenly among them (Fig. S3). Because *V. cholerae* may travel more than one cell per time step, the above procedure is repeated n times to allow it to move n kilometers per day. That is, individuals shed n times into the river ($1/n$ of the total amount for each iteration) and it is moved downstream each time. At each step, a small fraction of the *V. cholerae* is lost. In this manner, a single source distributes *V. cholerae* to all communities within n steps downstream. The next day, this same *V. cholerae* will move another n cells downstream after it disappears by the fraction δ, the daily rate of environmental *V. cholerae* loss.

River access is limited to six communities (about 3,000 people) per cell (Fig. 1B). That is, if there are 60,000 individuals in a single cell on a river, then 3,000 of them will shed into and drink from the river, and the rest do not. Fig. S4 shows the simulated incidence of cholera in Haiti. Note that cholera incidence is highest along the rivers.

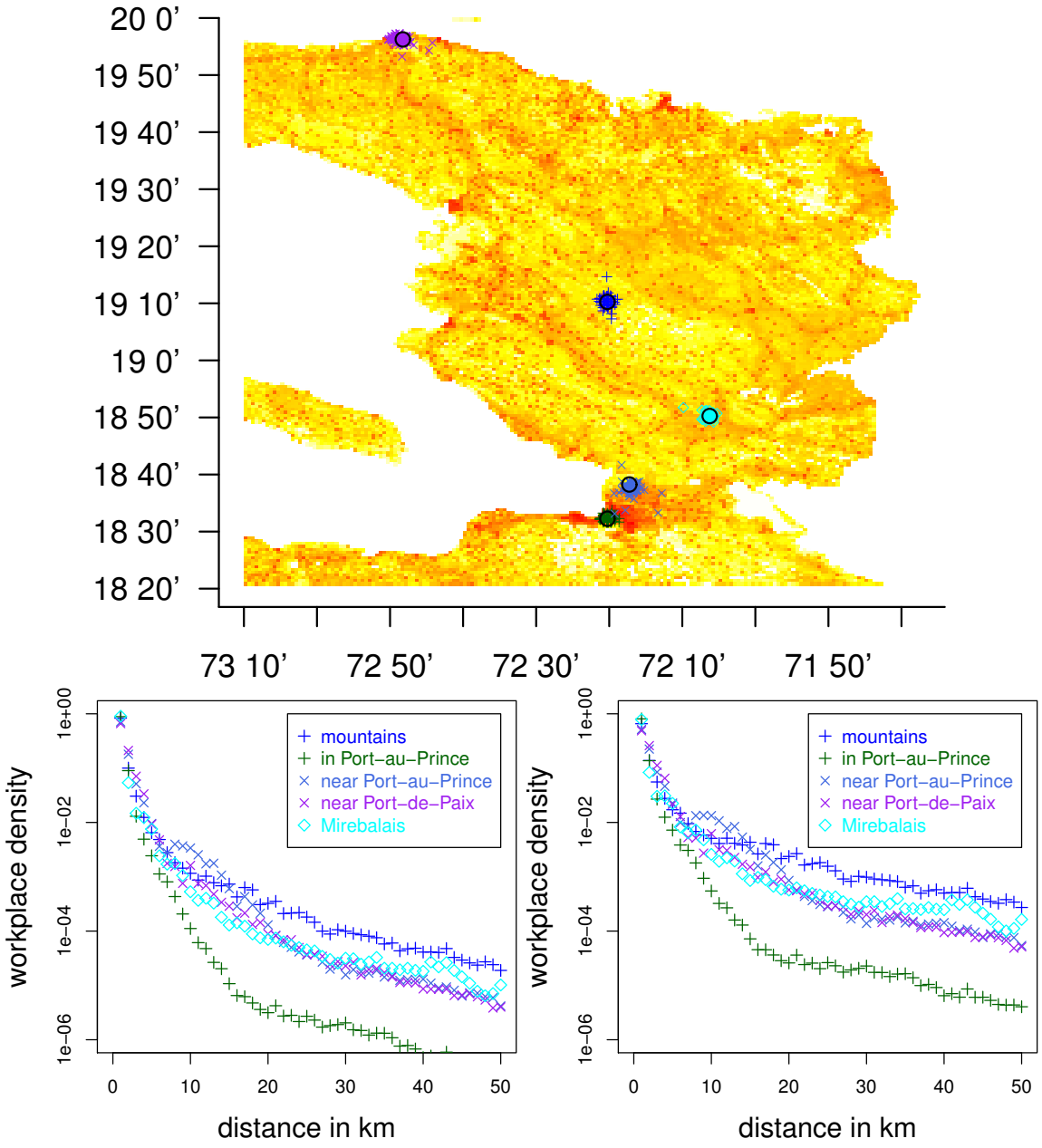

Figure S2: Work distance in Haiti based on a gravity model. The map of Haiti is a heatmap based on population density (per km^2). 10’ in the map is 20 kilometers. For the gravity model of commuter travel, $\tau_1 = 0.30$, $\tau_2 = 0.64$, $\rho = 3.8$, and the maximum distance of travel is 40 km (0 degrees, 25’). Four sample locations were chosen, as indicated by black circles. For each location, the colored crosses represent the workplaces of 200 people who do not work in their home location. The lower left plot shows the gravity model distance that people travel to work in these five locations using the same color scheme. The lower right plot is the same, but for $\rho = 3.05$ to show a contrasting United States-like commuter travel.

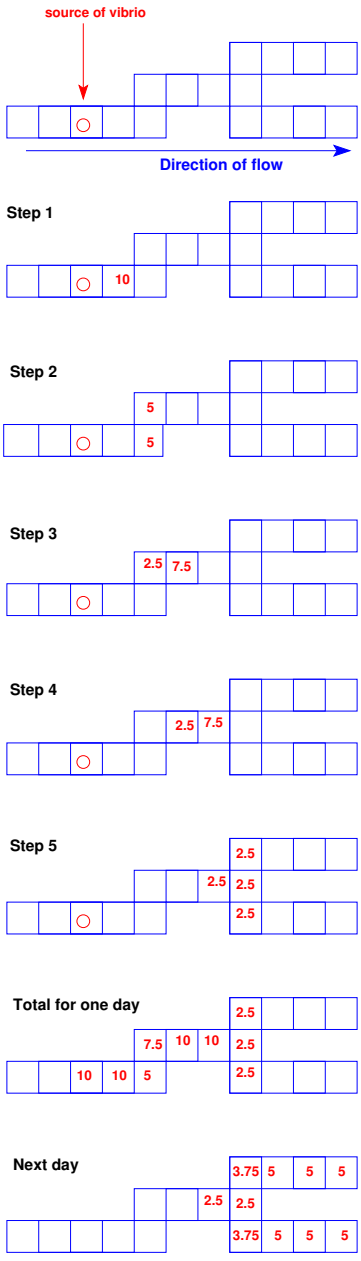

Figure S3: The flow of *V. cholerae* down a river. In this example, 60 units of *V. cholerae* are shed into the river at the source, and we assume that the river flows 5 km per day. We compute the distribution of *V. cholerae* in five steps, one for each kilometer of flow. We start by adding 10 units (1/6 of the total) of *V. cholerae* at the source, then we move it to adjacent downstream communities in each step. In the end, we sum the distribution of *V. cholerae* from all steps to get the distribution of the 60 units of *V. cholerae* shed in one day. The next day, if there is no more shedding at the source, the *V. cholerae* in the river near the source is gone, having flowed another 5km downstream. We can also specify that there is some amount of loss of *V. cholerae* per distance traveled, so that levels drop exponentially as you go downstream.

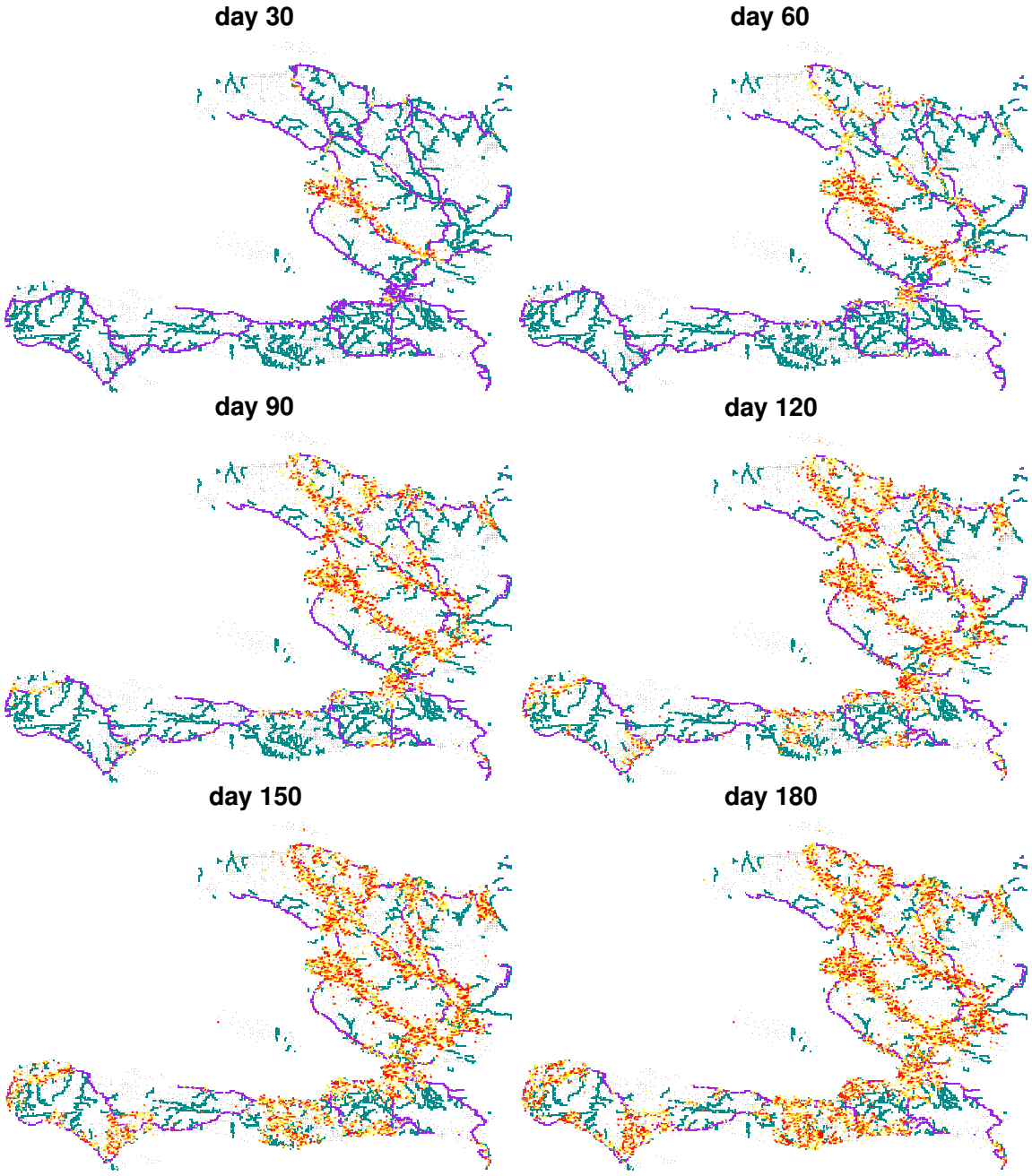

Figure S4: Cumulative illness incidence in a simulated cholera epidemic in Haiti. The plots are heatmaps of symptomatic incidence per population (yellow is low, red is high). Rivers are shown in blue, highways in purple.

Parameter	**Value**	**Reference**
symptomatic fraction	20%	[1, 8]
latency period length	1–5 days	[1, 9, 10]
infectious period length	7–14 days	[1]
fraction of day spent at work	0.3	
m, relative infectiousness of asymptomatic (wrt symptomatic)	10%	[1]
units of cholera shed each day by symptomatic	1	fixed
p, daily person-to-person household contact	0.01	
κ, *V. cholerae* level at which daily infection probability is 50%	70	
β, scaling parameter for environmental source	0.25	
δ, decay rate of cholera in the environment	1/30 day^{-1}	[11]
multiplier for infectiousness of freshly shed vibrio	100	[3, 11]
$\mathsf{VE_S}$	0.0	[5]
$\mathsf{VE_I}$	0.5	[1]
$\mathsf{VE_P}$	0.64	[5]
H, relative personal hygiene	0–0.3	baseline is 0, increases to 0.3 in a public health campaign
daily highway travel probability	5×10^{-5} per day	
daily long-distance travel probability	5×10^{-6} per day	
τ_1	0.3	gravity model parameter [7]
τ_2	0.64	gravity model parameter [7]
ρ	3.8	gravity model parameter [6]

Table S1: Summary of model parameters.

1.3 Population density and total data

We use population estimates at a resolution of 30 $\times$ 30 degree cells, which is approximately 1km^2 (From http://www.ornl.gov/sci/landscan/, downloaded Nov 11, 2010). In the model, the population of a single 1km^2 cell is divided into communities with equal size populations of approximately 500 people (Fig. 1B). If a cell contains fewer than 500 people, then the residents are placed in a single community.

We found the departments to have the following populations in the LandScan and Census 2009 projection data [12]:

Department	LandScan	Haiti Census projection for 2009
Grand Anse	392,670	425,878
Nord Ouest	587,823	662,777
Sud	665,347	704,760
Artibonite	1,391,690	1,571,020
Centre	636,143	678,626
Nippes	284,582	311,497
Nord Est	330,703	358,277
Nord	922,645	970,495
Ouest	3,157,476	3,664,620
Port-au-Prince	2,223,771	2,509,939
Sud Est	653,565	575,293

We assign individuals to households, to match the distribution of household sizes in Haiti. From the IHSI document "Enquête sur les conditions de vie en Haïti Vol 1" (Tableau 2.5.1.1), we find that 9.3% of households have 1 person, 29.0% have 2–3 people, 41.2% have 4–6 people, 16.8% have 7–9 people, and 3.7% have 10 or more. If we assume that the distribution of household sizes is uniform within each range of people (e.g., 14.5% of households have 2 people and 14.5% have 3) and that 10 is the maximum household size, 2% of people live alone, 16% of people live in households of 2 or 3 (8% for 2, 8% for 3, respectively), 45% in households of 4–6, 29 in households of 7–9, and 8% in households of 10.

2 Measuring the basic reproductive number in the model

We estimated R_0 in the simulation by infecting a single randomly selected person in the population then running a simulation in which secondary cases are not infectious (do not shed or transmit person-to-person). We repeated this process 20,000 times to obtain a distribution of outcomes. We find that a randomly chosen individual will generate a mean of 2.6 (median of 0.0) secondary infections, so $R_0 = 2.6$. If the individual is symptomatic, then the mean is 9.6 and the median is 3.0. If the individual lives on a river, the mean is 10.0 and the median is 2.0. If the individual is not on a river, the mean is 0.8 and the median is 0.0. See Fig. S5.

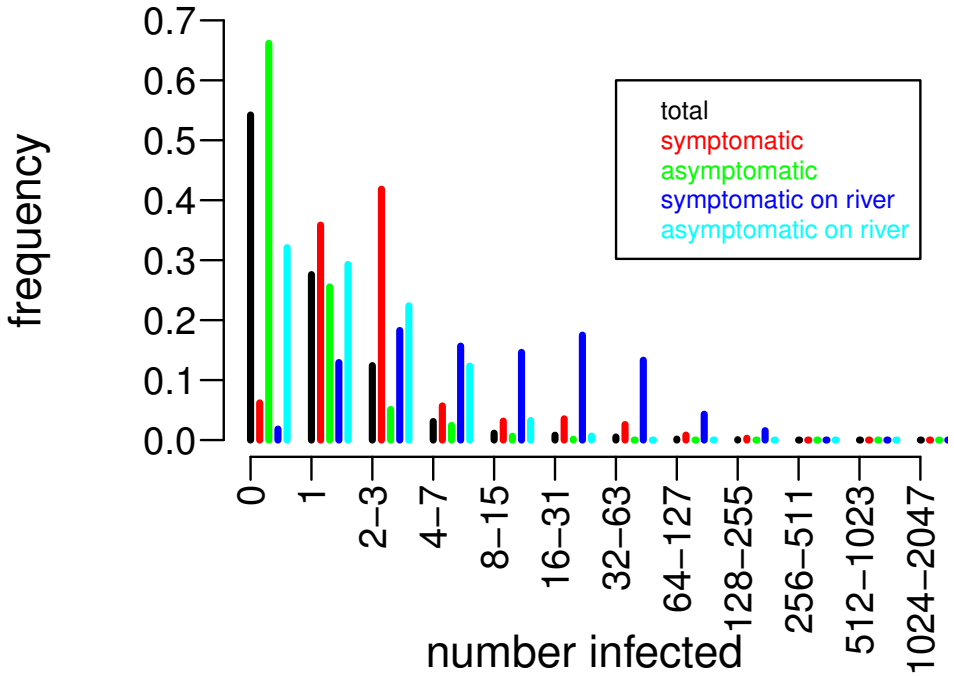

Figure S5: Transmissibility for various populations in the simulator. The simulation was run 20,000 times, infecting a single person and ensuring that secondary cases were not infectious. The total number of people infected by the index case was counted (in black). We measured the number of infections if the index case was symptomatic (red), asymptomatic (green), symptomatic and living on a river (blue), and asymptomatic and living on a river (cyan).

3 Sensitivity analyses

We simulated mass vaccination with different final amounts of vaccine (Fig. S6A). We also simulated a *pro rata* approach in which all cells were vaccinated in proportion to their populations rather than 70% or nothing (Fig. S6B). In these simulations, we first determined the final vaccine coverage (e.g., 30%), then randomly selected cells were vaccinated at this level each day. Coverage of less than 50% appeared to give slightly higher attack rates but with slightly less variability among stochastic runs than the non *pro rata* strategy that covers 70% of each cell.

We simulated reactive ring vaccination with different delays after the two cases appeared in a cell (Fig. S7A). That is, cells were eligible for vaccination n days after two residents become symptomatic. The effectiveness of the strategy diminishes with the delay length, and we chose a 5-day delay as the default in the main text. We also simulated vaccination five days after n cases appear in a cell (Fig. S7B). We can interpret these thresholds as different case detection ratios.

We qualitatively assess the effect of varying various simulation parameters on the relative timing of the epidemic among departments. As β is increased or κ is decreased, transmissibility is increased and the peaks among departments are more clustered (Figs. S8 and S9). Increasing the amount that infected individuals shed into the river also speeds transmission, and one can see that both Nord and Port-au-Prince have much earlier peaks (Fig. S10). The long-distance travel and driving probabilities particularly affect the importation of cholera in the more remote departments, such as Nippes and Grande Anse (Figs. S11 and S12). If these probabilities are set too high, then these departments have epidemic peaks that are much earlier than actually occurred. Increasing these probabilities also causes the peak in Port-au-Prince to occur much sooner. Changing the daily within-household transmission probability, p, increases the speed of the epidemic but the effect is the same in all departments since they are assumed to have the same household structure (Fig. S13). The relative infectiousness of asymptomatic individuals with respect to symptomatic individuals, m, is not known. We set this value to 10% in our model. Lower values resulted in slower spread of the epidemic to other departments (Fig. S14), indicating that asymptomatic infectious individuals play an important role in propagating the epidemic in the model. Recall that symptomatic individuals are less likely to travel in the model. We also tested the model with various values for the symptomatic fraction, which is the fraction of infected individuals who become symptomatic. We found that the size of the epidemic was sensitive to this parameter, with more cases appearing with higher values of the symptomatic fraction (Fig. S15). The higher symptomatic fraction results not only in more infected individuals being reported as cases, but these symptomatic individuals are also more infectious than asymptomatic. The incidence of symptomatic cholera in small localities within the model often approached the symptomatic fraction, consistent with 100% of the people being infected in these regions.

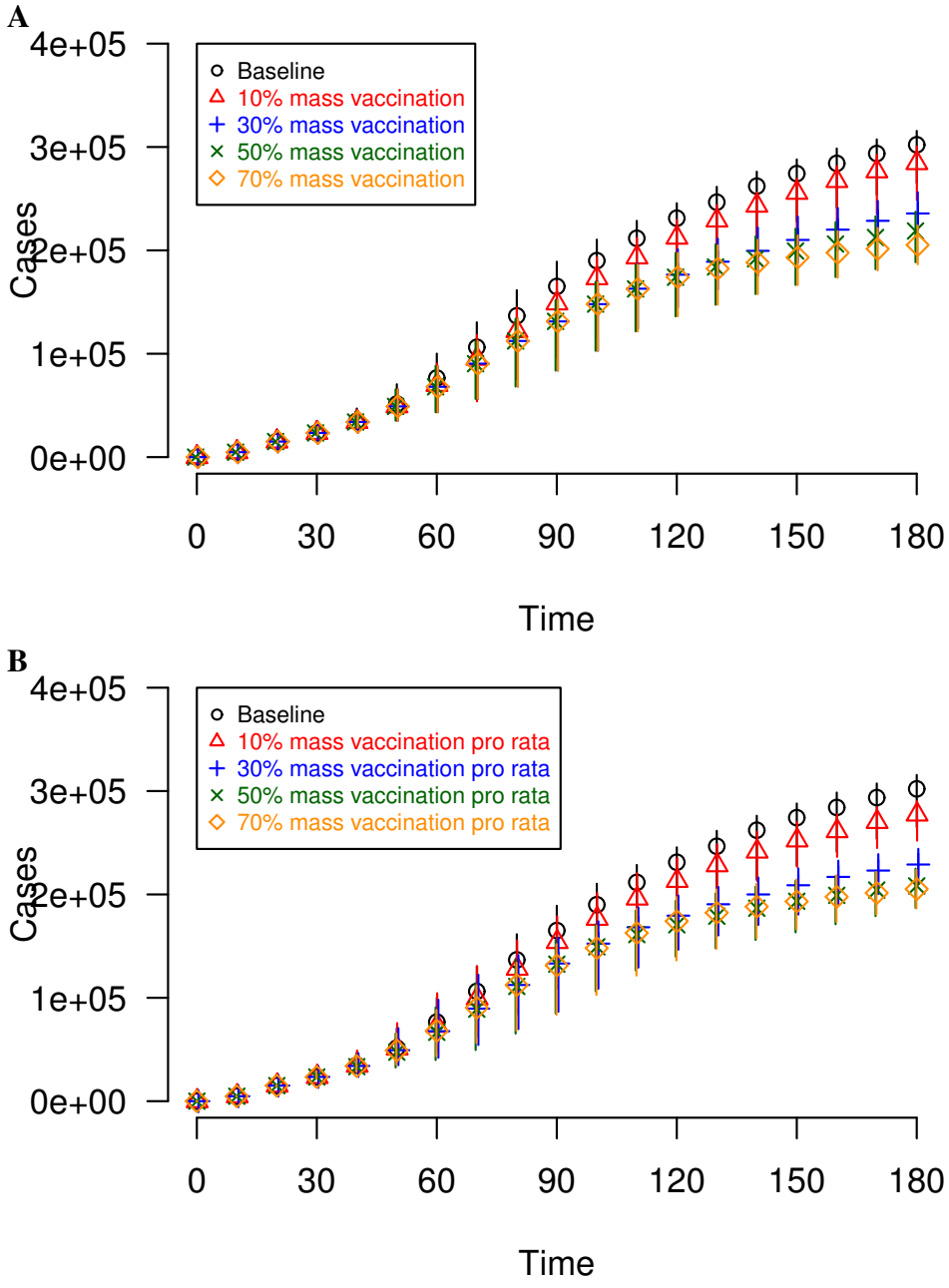

Figure S6: Cumulative incidence of cholera in simulations of reactive mass vaccination. Vaccinations start on day 21, with 50,000 courses available per day. The model was run 50 times for each scenario. Points represent median numbers of cases, lines indicate range of simulation results. **(A)** Reactive mass vaccination, covering 70% of limited regions as described in the main text. **(B)** A variant of reactive mass vaccination in which vaccine is distributed *pro rata*, i.e., the same fraction of individuals is vaccinated throughout the country rather than attempting to reach 70% of residents of each cell.

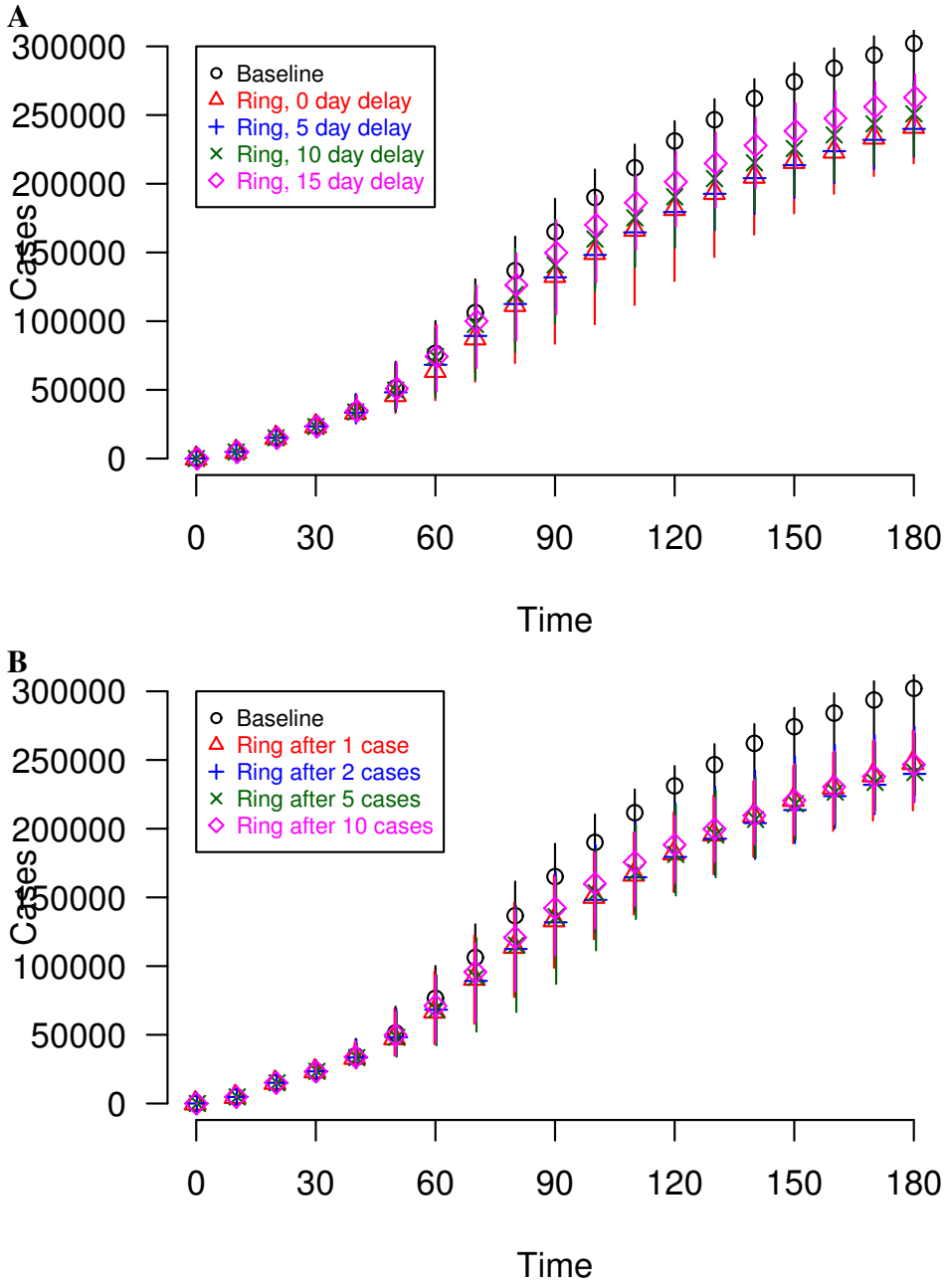

Figure S7: Cumulative incidence of cholera in simulations of reactive ring vaccination with enough vaccine for 30% of the country's population. The median number of cases from 50 simulations is plotted, with vertical lines to indicate the minimum and maximum values from the simulations. **(A)** The effect of delays in ring vaccination, assuming cells are prioritized for vaccination n days after two cases occur in the cell. **(B)** Using different thresholds (numbers of cases) to begin ring vaccination. We assumed that cells were prioritized to be vaccinated 5 days after 1, 2, 5, or 10 cases appeared in the cell.

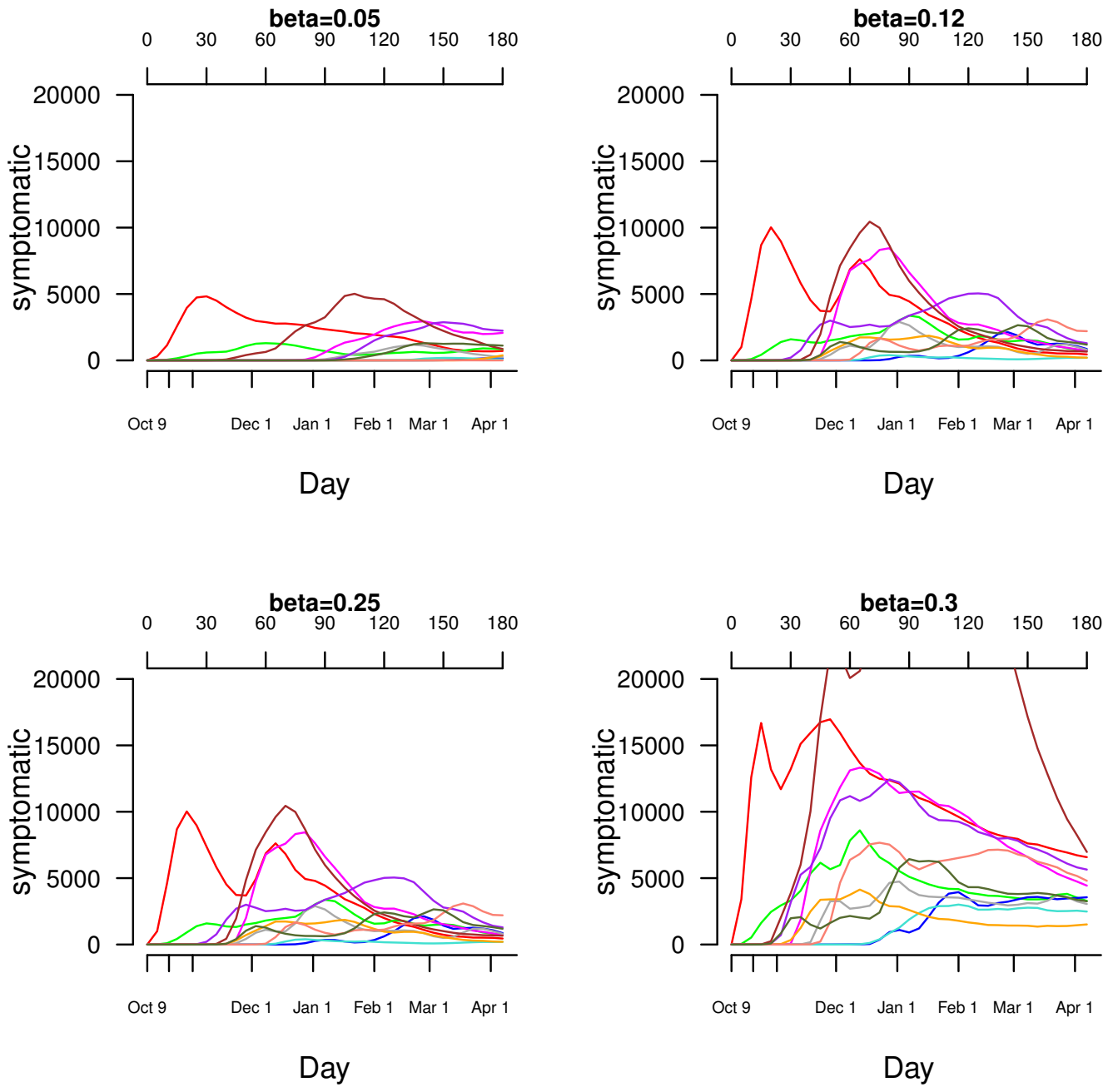

Figure S8: Effect of changing β in the model. The default is $\beta = 0.12$. Each curve represents a single department, using the color code from Fig. 2A.

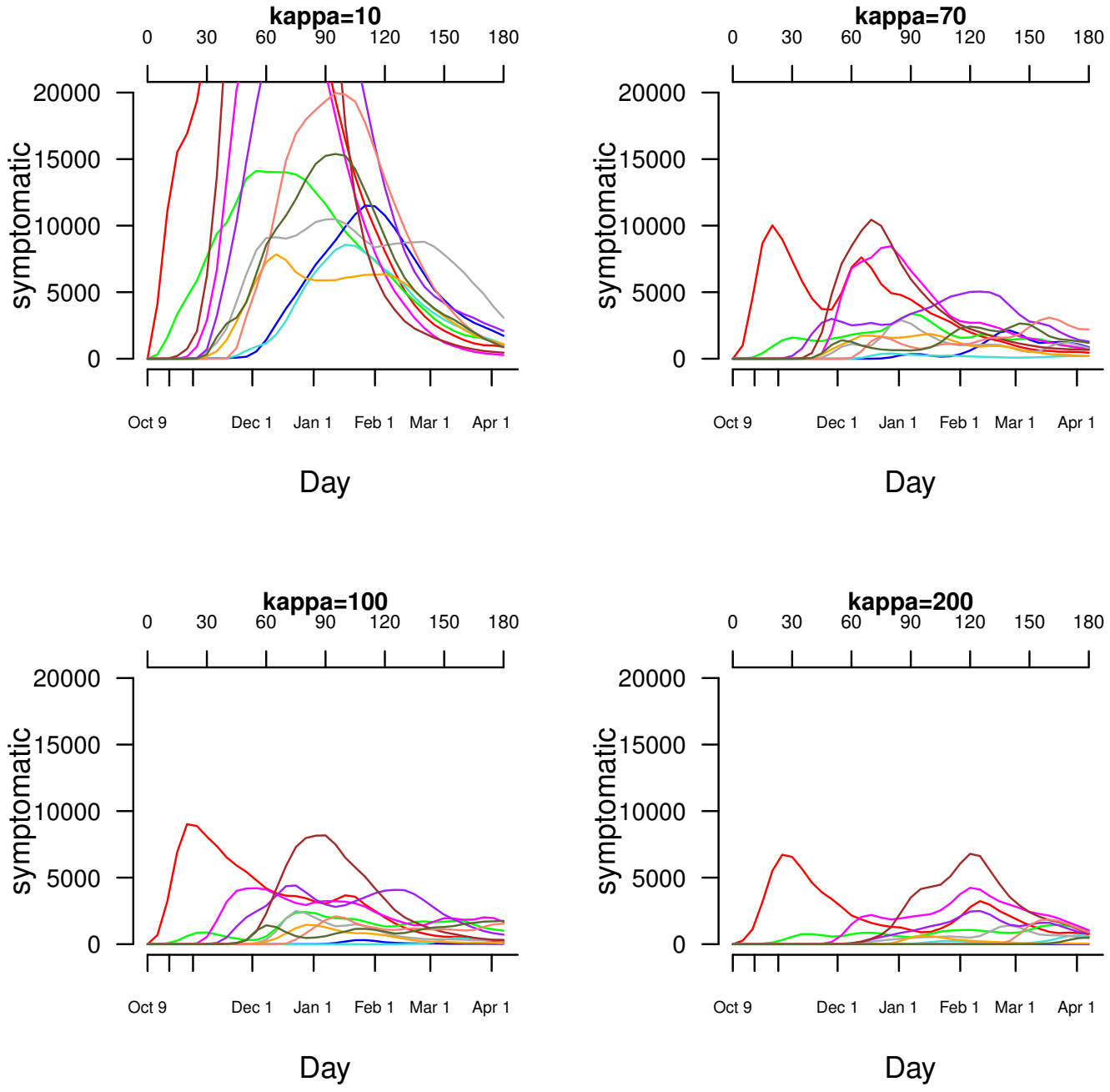

Figure S9: Effect of changing κ in the model. The default is $\kappa = 70$. Each curve represents a single department, using the color code from Fig. 2A.

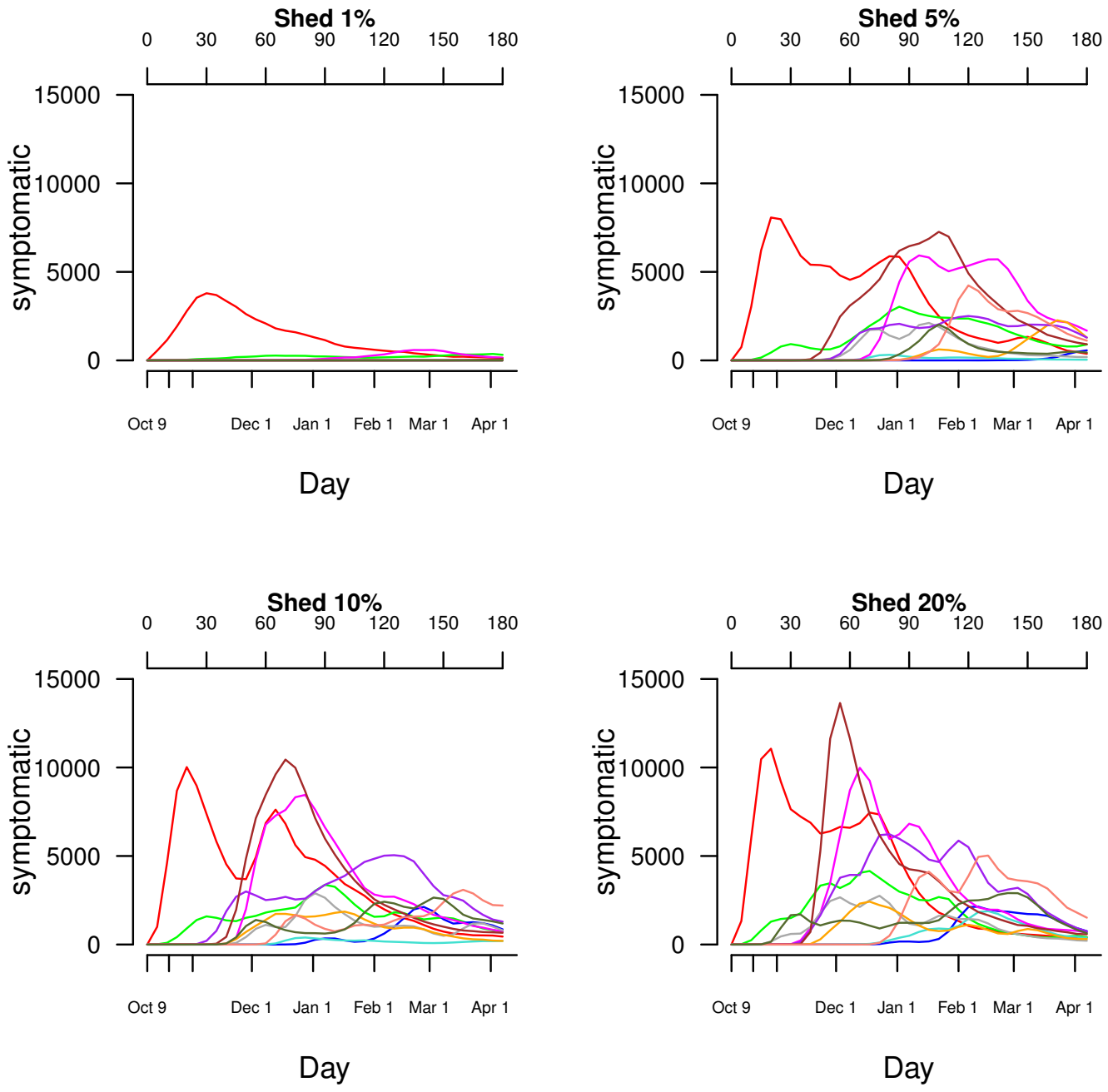

Figure S10: Effect of changing the fraction of cholera shed into the river. The default is 0.1 (10%). Each curve represents a single department, using the color code from Fig. 2A.

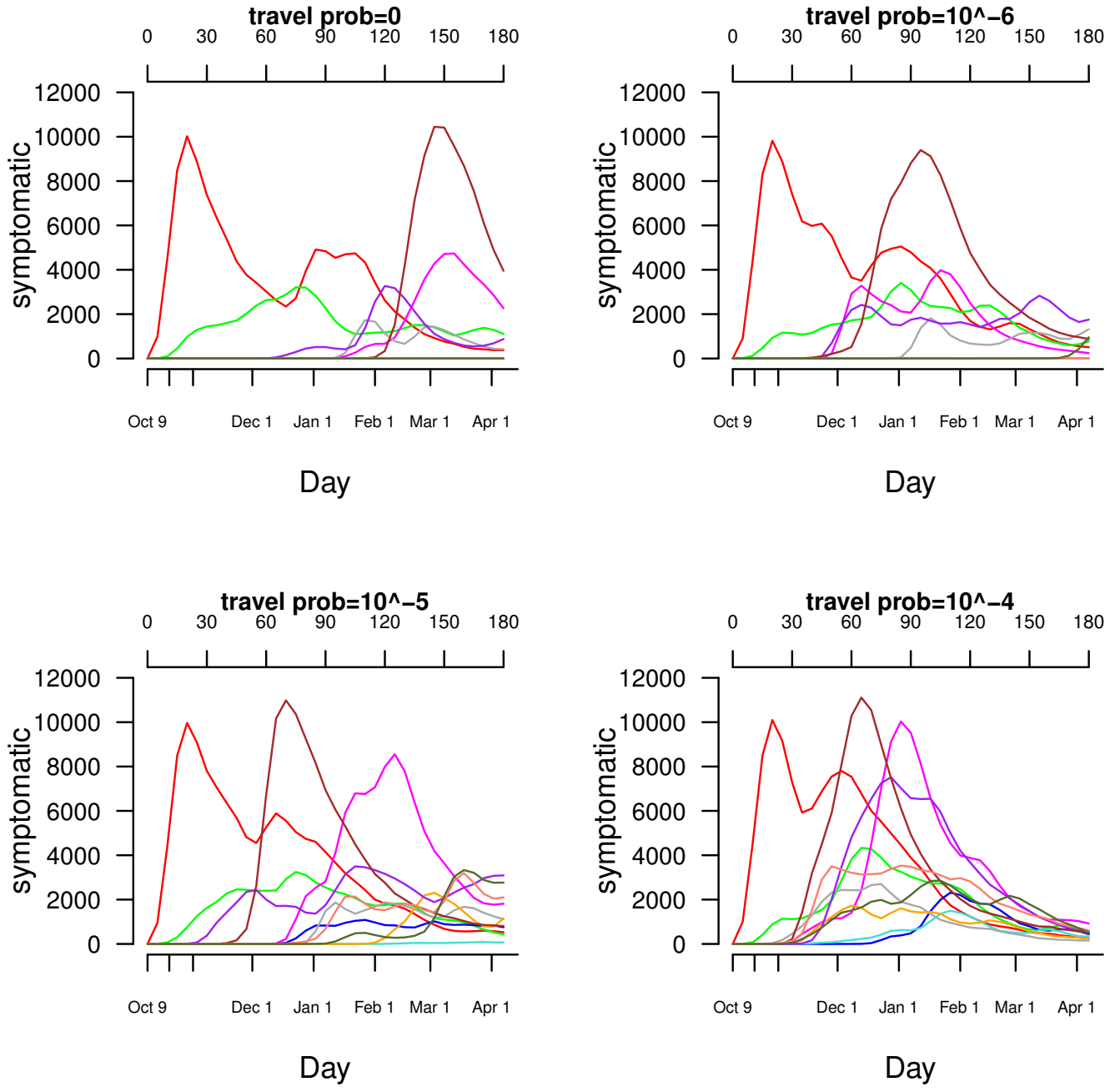

Figure S11: Effect of changing the daily probability of long-distance travel. The default is 5×10^{-6}. Note that for all cases, highway travel occurs with probability of 0. Each curve represents a single department, using the color code from Fig. 2A.

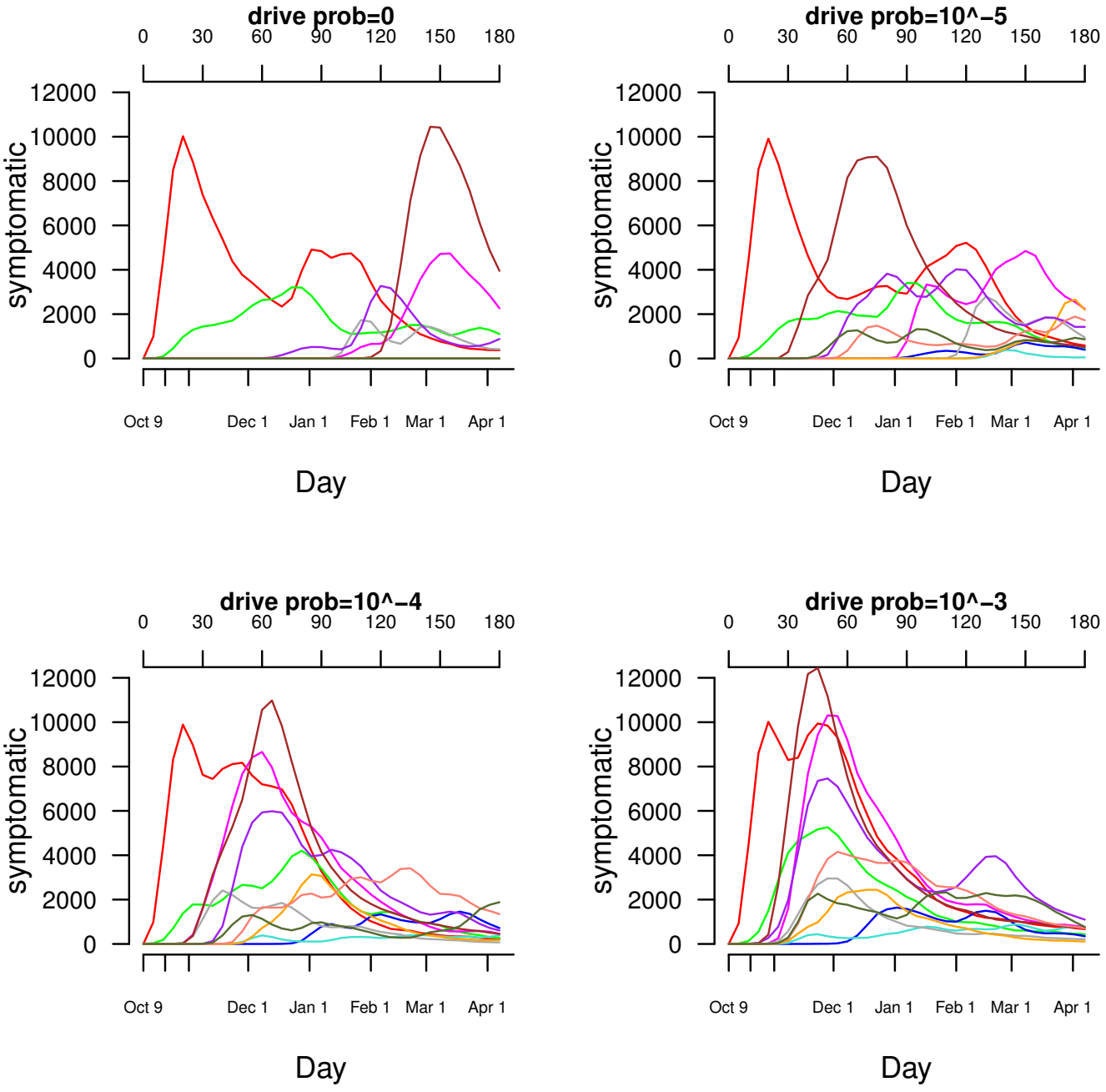

Figure S12: Effect of changing the daily probability of driving along a highway. The default is 5×10^{-5}. Note that for all cases, long-distance travel occurs with probability of 0. Each curve represents a single department, using the color code from Fig. 2A.

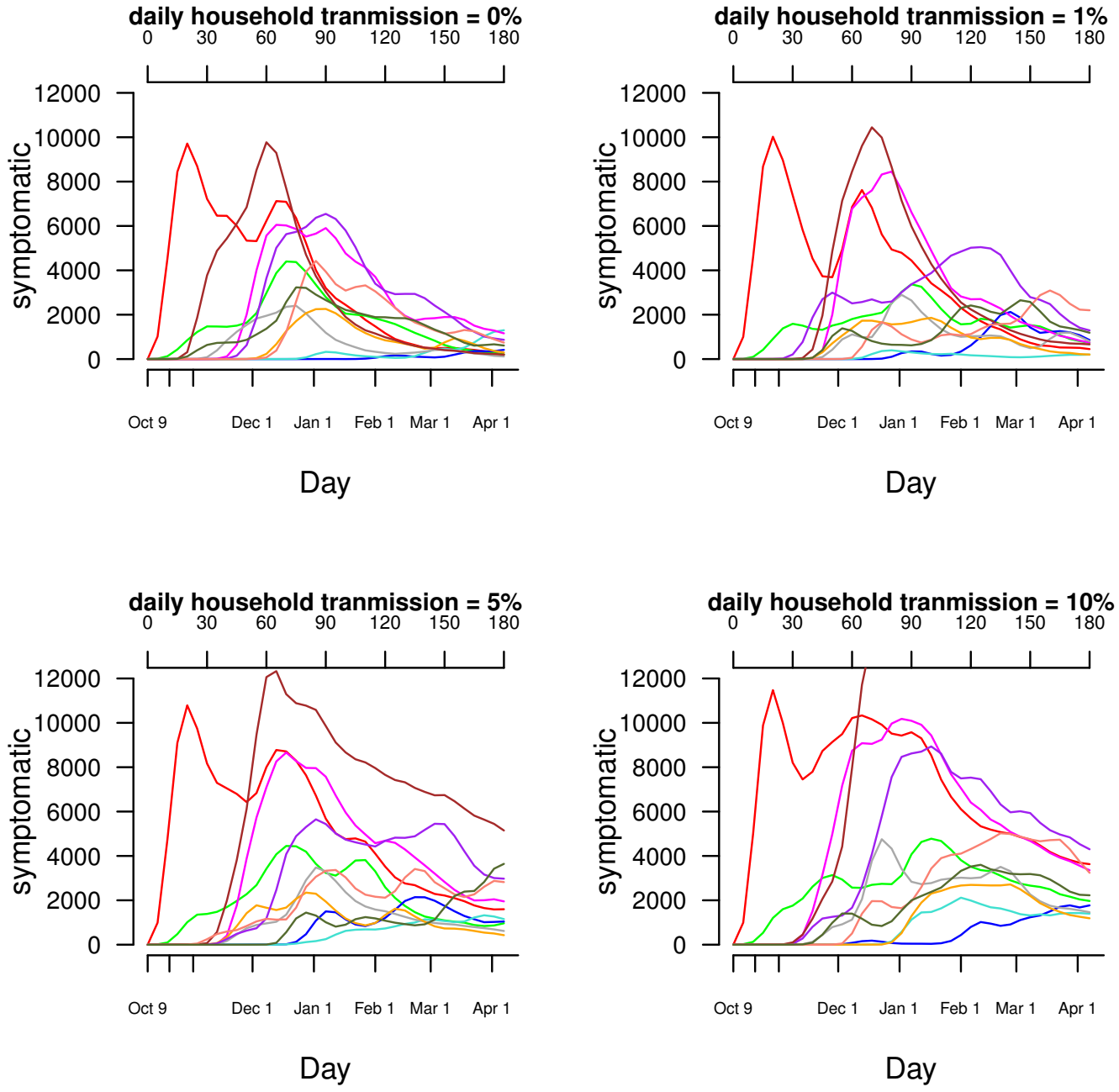

Figure S13: Effect of changing the daily within-household transmission probability. The default is 1% per day if the infectious person is symptomatic. Note that in the model, household composition is the same in all departments. Each curve represents a single department, using the color code from Fig. 2A.

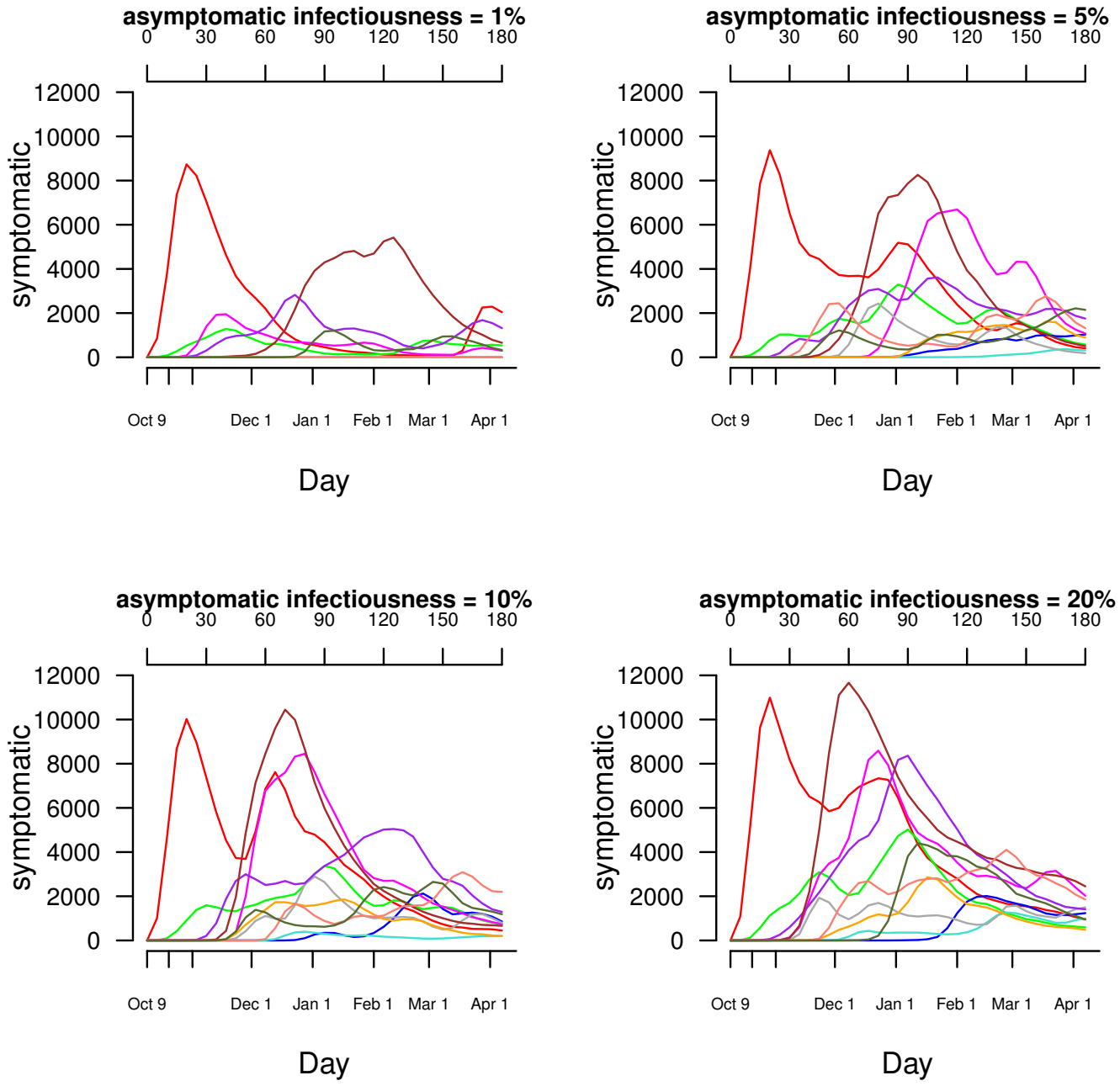

Figure S14: Effect of changing m, the infectiousness of asymptomatic individuals relative to symptomatic individuals. The default is 10%.

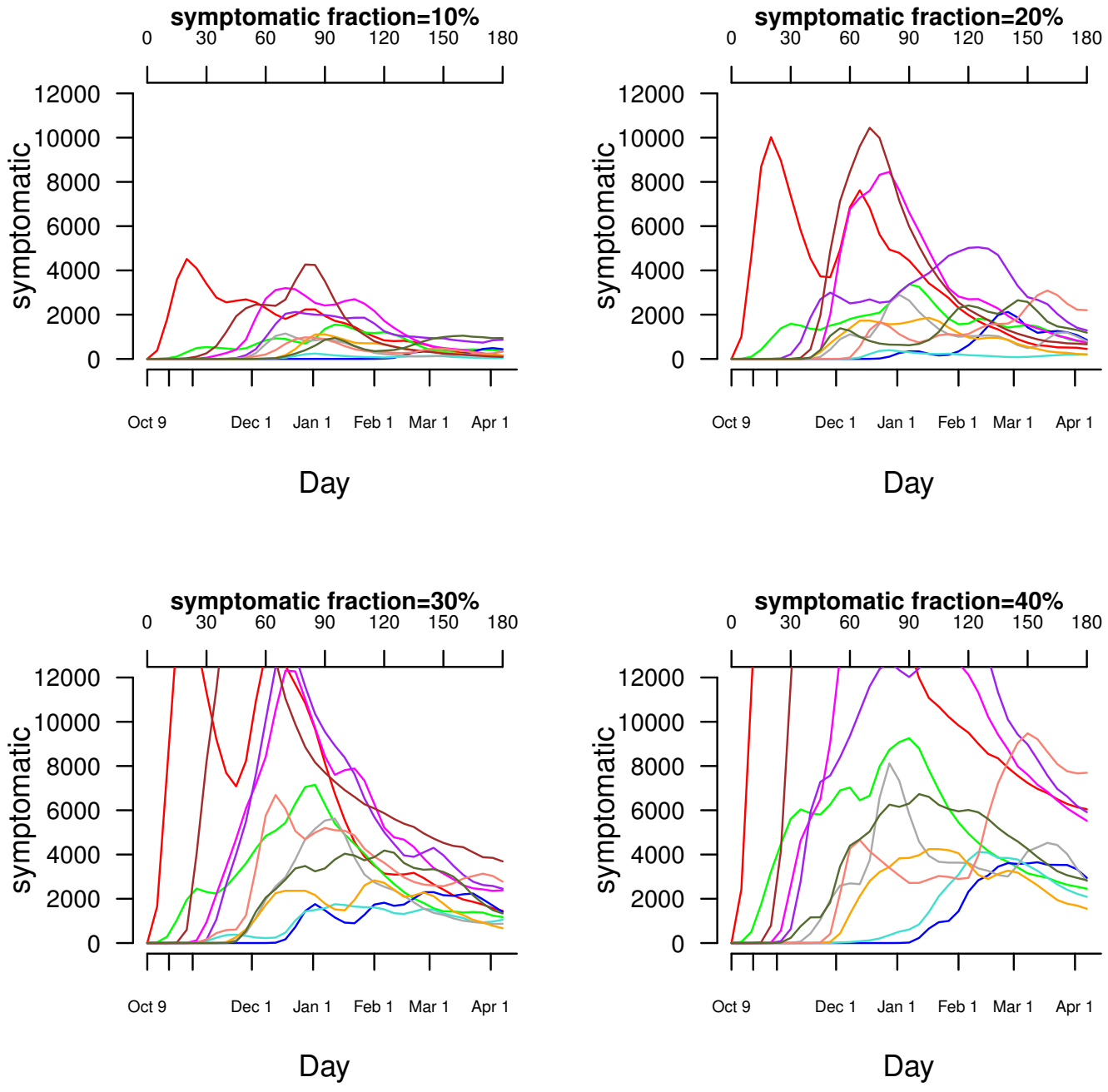

Figure S15: Effect of changing the fraction of infected individuals who become symptomatic. The default is 20%.

References

[1] Longini, I. M., Jr. *et al.* Controlling endemic cholera with oral vaccines. *PLoS Med* **4**, e336 (2007).

[2] Codeço, C. T. Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir. *BMC Infect Dis* **1**, 1 (2001).

[3] Merrell, D. S. *et al.* Host-induced epidemic spread of the cholera bacterium. *Nature* **417**, 642–5 (2002).

[4] Halloran, M. E., Longini, I. M., Jr. & Struchiner, C. J. *Design and Analysis of Vaccine Studies* (Springer, New York, 2010).

[5] Black, R. E. *et al.* Protective efficacy in humans of killed whole-vibrio oral cholera vaccine with and without the B subunit of cholera toxin. *Infect Immun* **55**, 1116–20 (1987).

[6] Ferguson, N. M. *et al.* Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437**, 209–14 (2005).

[7] Viboud, C. *et al.* Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312**, 447–51 (2006).

[8] Weil, A. A. *et al.* Clinical outcomes in household contacts of patients with cholera in Bangladesh. *Clin Infect Dis* **49**, 1473–9 (2009).

[9] Sack, D. A., Sack, R. B., Nair, G. B. & Siddique, A. K. Cholera. *Lancet* **363**, 223–33 (2004).

[10] Nelson, E. J., Harris, J. B., Morris, J. G., Jr., Calderwood, S. B. & Camilli, A. Cholera transmission: the host, pathogen and bacteriophage dynamic. *Nat Rev Microbiol* **7**, 693–702 (2009).

[11] Hartley, D. M., Morris, J. G., Jr. & Smith, D. L. Hyperinfectivity: a critical element in the ability of *v. cholerae* to cause epidemics. *PLoS Med* **3**, e7 (2006).

[12] l'Institut Haïtien de Statistique et d'Informatique (IHSI). Population totale, population de 18 ans et plus ménages et densités estimeés en 2009. Available from `http://www.ihsi.ht/pdf/projection/POPTOTAL&MENAGDENS_ESTIM2009.pdf` (2009).