**Supplementary Data**

*Bayesian Methodology*

The general expression for the posterior probability distribution given in Eq. 6 for population-level differences in $R_0$ was adjusted to reflect model specifics for each candidate model (Table 2). For example, Bayes' theorem took the following form for the population-level difference model (Model 3, Table 2):

$$\begin{aligned} P(R_{0,EA/NA}, R_i, \sigma^2_{R_i,EA/NA}, \alpha, \gamma, \mu_{EA/NA}, a_{EA/NA}, b_{EA/NA}, \sigma^2_{EA/NA}|\text{data}) \propto \pi(R_{0,EA/NA})\,\pi(R_i) \times \\ \pi(\sigma^2_{R_i,EA/NA})\,\pi(\alpha)\,\pi(\gamma)\,\pi(\mu_{EA/NA}|a_{EA/NA}, b_{EA/NA})\,\pi(a_{EA/NA}), \pi(b_{EA/NA}), \pi(\sigma^2_{EA/NA}) \times \\ \mathcal{L}(R_{0,EA/NA}, R_i, \sigma^2_{R_i,EA/NA}, \alpha, \gamma, \mu_{EA/NA}, a_{EA/NA}, b_{EA/NA}, \sigma^2_{EA/NA}|\text{data})\ . \end{aligned} \tag{S1}$$

Here the notation $EA/NA$ indicates whether the population was of European/African ($EA$) or Native-American ($NA$) descent. This population-level model accounts for overall differences in $R_0$ and $\sigma^2$ for Native-American and European/African populations, as is true of all models examined. It also accounts for population-level differences in reproductive rates of spread $R_i$. The population-level model was constructed for both the homogeneous (Eq. 1 – 4) and heterogeneous (Eq. 8 – 9) SEIR models.

For the homogeneous SEIR model incorporating population-level differences, the prior distributions are given as follows:

$$\begin{aligned} R_{i,EA} \sim \mathbf{LN}(R_{0,EA}, \sigma^2_{R_i,EA}) \quad &, \quad R_{i,NA} \sim \mathbf{LN}(R_{0,NA}, \sigma^2_{R_i,NA}), \\ R_{0,EA} \sim \mathbf{LN}(m_{EA}, \sigma^2_{R_0,EA}) \quad &, \quad R_{0,NA} \sim \mathbf{LN}(m_{NA}, \sigma^2_{R_0,NA}), \\ 1/\sigma^2_{R_i,EA} \sim \mathbf{\Gamma}(mean = 1, var = 0.1) \quad &, \quad 1/\sigma^2_{R_i,NA} \sim \mathbf{\Gamma}(mean = 68.5, var = 13.7) \\ \sigma^2_{EA/NA} \quad &\sim \quad \mathbf{U}(0, 1e6), \\ \alpha \quad &= \quad \frac{1}{\alpha_1 + \alpha_2}, \\ \alpha_1 \sim \mathbf{\Gamma}(mean = 2.6, var = 0.6) \quad &, \quad \alpha_2 \sim \mathbf{\Gamma}(mean = 12.3, var = 6.1), \\ 1/\gamma \quad &\sim \quad \mathbf{\Gamma}(mean = 16.0, var = 8.0), \\ \mu_{EA} \sim \mathbf{Be}(a_{EA}, b_{EA}) \quad &, \quad \mu_{NA} \sim \mathbf{Be}(a_{NA}, b_{NA}), \end{aligned} \tag{S2}$$

$$\frac{a_{EA}}{a_{EA}+b_{EA}} \sim \mathbf{Be}(3,17) \quad , \quad \frac{a_{NA}}{a_{NA}+b_{NA}} \sim \mathbf{Be}(4,4.5),$$
$$a_{EA} \sim \mathbf{\Gamma}(mean=30, var=150) \quad , \quad a_{NA} \sim \mathbf{\Gamma}(mean=4, var=4).$$

Here the hyperparameters on the fatality rate are $a_{EA}$, $b_{EA}$, $a_{NA}$, and $b_{NA}$, which differed between European/African and Native-American populations. **LN**, **U**, **Exp**, $\mathbf{\Gamma}$, and **Be** denote log-normal, uniform, exponential, gamma, and beta distributions, respectively.

We chose an informative prior on the inverse of the variance (i.e., the precision), $1/\sigma^2_{R_{i,EA/NA}}$, due to potential problems with traditional "non-informative" gamma distributed priors given relatively small sample sizes (Gelman 2006). However, this choice did not have a large effect on the posterior estimates of the $R_0$ or $R_i$. For populations of European/African descent, the prior median along with 95% central interval of the variance was 1.03 (0.586, 2.087) and the posterior was 0.88 (0.520, 1.600). Thus for these populations, there was overlap between the prior and the posterior. For the Native American populations, the prior median along with 95% central interval for the variance was 0.01 (0.013,0.016) and the posterior was 0.50 (0.341, 1.635). While the prior was informative, especially in the lower range of the variance, the prior distribution was relatively flat across other parts of the likelihood surface. Also note that these priors had little effect on the posterior distribution of $R_i$ across European/African and Native American populations as shown by the marginal posterior distributions of $R_i$ values between populations in Figs. 4 and 5.

Additionally, we performed a sensitivity analysis by using a more diffuse exponentially distributed prior for the inverse variance with the rate parameter equal to one. The median (95% central interval) for the prior on $\sigma^2_{R_{i,EA/NA}}$ was 1.44 (0.272, 39.594). The associated posterior estimates for $\sigma^2_{R_{i,EA}}$ were 0.46 (0.180, 1.633) and for $\sigma^2_{R_{i,NA}}$ were 0.61 (0.245, 1.901). Note that the posterior estimates using this prior overlap with the estimates using an informative prior. As with the informative prior, this more diffuse prior had little to no effect on the median and 95% credible intervals of $R_0$ or any $R_i$. It seems that the original informative prior has little impact on the reported results and conclusions.

For the homogeneous SEIR model incorporating spatial correlation, the prior for $R_{i,NA}$ from equation S2 needs to be slightly modified and additional structure added. These changes

can be expressed mathematically as,

$$\begin{aligned} R_{i,NA} &\sim \mathbf{MVLN}(R_{0,NA}, \sigma^2_{R_{i,NA}} \mathbf{\Psi}) \,, \\ \mathbf{\Psi} &\sim \mathbf{Exp}(-d\,\mathcal{D}) \,, \\ d &\sim \mathbf{Exp}(500) \,, \end{aligned} \tag{S3}$$

where $\mathbf{\Psi}$ represents the $8 \times 8$ dimensional variance-covariance matrix which specifies the pairwise spatial correlations between $R_i$ parameters for each pair of Native-American populations which decays exponentially with distance. $\mathcal{D}$ is a normalized geographic distance matrix between these populations (Fig. 1). **MVLN** denotes a multivariate log-normal distribution.

*Markov chain Monte Carlo (MCMC) sampler*

The posterior distribution for non-linear compartmental models is usually not available in closed form. One way to approximate the posterior probability density for the model parameters is via sampling-based methods, such as Markov chain Monte Carlo (MCMC). To use this approach, we evaluated the likelihood using an automatic second- and third-order pair step-size Runge-Kutta-Fehlberg integration method algorithm (Matlab; Mathworks, Natick, MA). The state variables in the SEIR model change smoothly over time (Anderson and May 1991) and modern integration methods perform well with such equations (Kreyszig 1993). Due to the complexity of the likelihood, however, the posterior distribution can only be formulated up to the proportionality constant and must be evaluated point-wise. We therefore used the Monte Carlo approach to obtain the approximate posterior distribution by generating samples from the posterior. To that end, we followed standard practice in using an MCMC algorithm that replaces each multi-dimensional random draw with a sequence of single-parameter or lower-dimensional draws (O'Neill and Roberts 1999). We ran five separate chains that differed in their initial starting conditions over 110,000 iterations. We assessed convergence using standard criteria (Brooks and Gelman 1998, Geweke 1992) after the chains had been trimmed of 10,000 iterations and the autocorrelation of the draws had been assessed. After confirming convergence (Fig. S4), we combined the chains for each of the models to generate a sample from the posterior distribution (Table 2).

*Unexplained Variation*

Socio-demographic differences between populations are almost certainly part of the reason why the model with population-specific basic reproductive rates $R_i$ describes the process more adequately than the model with a single, average basic reproductive rate. These differences, however, are responsible for only some of the variability among epidemics, and it is reasonable to expect that there will still be some unexplained variability present.

The main reason for this lies in the fact that epidemics are stochastic processes, and will differ from instance to instance - from population to population, from year to year, etc, beyond what we can attribute to effects of measured variables. This stochasticity of epidemics is one of the reasons why we see the variation between epidemics in different populations, and that variation is especially visible in smaller populations. Accordingly, it is not just the "explainable" differences that are driving the variability in the basic reproductive rates; some variability will always be present even after we account for the explicit differences, simply due to the stochastic nature of epidemics.

Variation between different instances of epidemics, even those resulting from the same contagion, is natural. This is particularly true when epidemics are occurring in different communities, such as missions, where the spread of contagion is guided by different demographics (e.g., age structure), environmental factors (e.g., availability of food and water) or social-mixing differences (e.g., differences in social norms or hierarchies). Consequently, having a flexible model that allows epidemic characteristics to vary from population to population is paramount. In particular, assuming that a single $R_0$ would be sufficient to describe transmission in all populations would be naïve.

The lack of fit between the model and the data thus cannot be assumed to be due solely to observation error. The total unexplained error includes the observation error, but also any error due to the fact that the model is an approximation of reality, regardless of whether it stems from us ignoring process stochasticity, which includes the assumption that the model is deterministic.

*Epidemics in Small Populations*

To determine whether deterministic models provided a good approximation, we compared the fit of a stochastic SEIR model to its deterministic counterpart in each population. To do this, we ran both the stochastic and deterministic models using the best-fit parameter estimates from the model that assumed inter-population differences in the reproductive rate of spread $R_i$ (Model 3). The stochastic model used the SEIR framework in combination with the Gillespie Direct Algorithm (Gillespie 1977, Keeling and Rohani 2008). The Gillespie algorithm assumes that only one event occurs at a time and so it steps through individual events in an epidemic as it progresses through the population. Given this assumption, only one individual can change their status during a single-time step in the SEIR model. The population is limited to three potential events: 1) an individual moves from susceptible to exposed; 2) an individual moves from exposed to infected; or, 3) an individual moves from infected to recovered. The overall probability of an event is the sum of the probabilities of each of the three types of events.

As in the deterministic model, we first began with a single infected individual. We then drew an exponentially distributed number to generate the time of the occurrence of the next event. Next, we drew a uniform random number to determine which of the three possible events occurred. The simulation was continued until either the end time of the epidemic data set or no more exposed or infected individuals existed in the population. We ran the stochastic model for all populations including the Santo Domingo de la Frontera, a Baja population which had the smallest population size of all the epidemics (Table 1). To summarize the stochastic runs, we estimated the log of the integrated likelihood (Berger et al. 1999) by taking the log of the average of the likelihoods (Ross 1997). To investigate the adequacy of this relatively simple approach to estimating integrated likelihoods, we repeated the above calculations with an increasing number of simulation runs. The stability of our results with increased simulations, as well as the unchanged conclusion in Table S2, seem to indicate that the simple method of estimating the integrated likelihood was adequate for the data at hand. However, if larger discrepancies were observed, alternative and more computationally intensive methods (Andrieu et al. 2010, Ionides et al. 2006) would be preferred.

The simulations and their 95% probability intervals show that the deterministic model fits the data well (Table S2 and Fig. S5), but that the uncertainty is far greater in the stochastic model, as expected. Note that the x- and y-axes in Fig. S5 differ due to population-level differences in mortality numbers and epidemic length. The pointwise median estimate clearly follows the deterministic output for all populations. This is not surprising given that stochastic models for disease outbreaks are recommended for populations under 100 individuals (Daley and Gani 1999). Additionally, the close match of the log of the average likelihoods for the deterministic and stochastic model lends further support (Table S2) to the use of the deterministic model in our analysis. The Bayesian analysis, however, could be undertaken using a stochastic SEIR model as well, and the uncertainty in such a model would be expected to be even greater than the one communicated in the Results presented in this paper.

**Supplementary Data References**


Anderson, R.M., May, R.M., 1991. Infectious diseases of humans: Dynamics and control. Oxford University Press.

Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle Markov chain Monte Carlo methods. Journal of the Royal Statistical Society Series B-statistical Methodology 72, 269–342.

Berger, J.O., Liseo, B., Wolpert, R.L., 1999. Integrated likelihood methods for eliminating nuisance parameters. Statistical Science 14, 1–22.

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7, 434–455.

Daley, D.J., Gani, J., 1999. Epidemic modelling: An introduction. Cambridge University Press, Cambridge.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 3, 515533.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Oxford University Press, Oxford. pp. 169–193.

Gillespie, J.H., 1977. Natural-selection for variances in offspring numbers - new evolutionary principle. American Naturalist 111, 1010–1014.

Ionides, E.L., Bretó, C., King, A.A., 2006. Inference for nonlinear dynamical systems. Proceedings of the National Academy of Sciences of the United States of America 103, 18438–18443.

Keeling, M.J., Rohani, P., 2008. Modeling infectious diseases in humans and animals. Princeton University Press.

Kreyszig, E., 1993. Advanced engineering mathematics. Wiley.

O'Neill, P.D., Roberts, G.O., 1999. Bayesian inference for partially observed stochastic epidemics. Journal of the Royal Statistical Society: Series A (Statistics in Society) 162, 121–129.

Ross, S., 1997. Introduction to probability models. Academic Press.

## Tables

Table S1: Effects of $R_0$ priors on the $R_0$ posterior distributions with 95% Credible Intervals for the best-fit model of population-level differences in $R_0$.

| Distribution | Prior Intervals | Posterior Intervals |
|---|---|---|
| Original | 4 (0.7, 23.7) | European/African – 6.7 (3.17, 14.43) |
| | | Native-American – 8.1 (4.11, 15.50) |
| Optimistic | 1.8 (0.13, 24.2) | European/African – 6.6 (3.00, 14.53) |
| | | Native-American – 8.1 (4.06, 14.32) |
| Pessimistic | 7 (1.9, 26.3) | European/African – 7.7 (3.96, 17.13) |
| | | Native-American – 8.5 (4.93, 14.96) |

Table S2: Comparison of the log-likelihood for the best- fit deterministic and the log of the average fit of the stochastic model. Numbers in parentheses are standard errors of the mean (SE). The percent change column shows the difference in likelihood estimates when comparing the deterministic and the stochastic model. Note that in general the stochastic and deterministic models give very similar results, and that increasing the number of stochastic realizations ("runs") had almost no effect on the likelihood score.

| Epidemic | Population Size | Deterministic Estimate | 1,000 runs Stochastic Estimate (SE) | 10,000 runs Stochastic Estimate (SE) | 25,000 runs Stochastic Estimate (SE) | % change † |
|---|---|---|---|---|---|---|
| Santo Domingo de la Frontera, Baja | 119 | 20.72 | 19.63 (0.095) | 19.62 (0.030) | 19.61 (0.019) | 5 % |
| Rosalia, Baja Sur | 133 | 12.05 | 11.23 (0.080) | 11.22 (0.025) | 11.20 (0.016) | 7 % |
| Saint Lawrence, NM | 215 | 13.01 | 12.32 (0.067) | 12.34 (0.021) | 12.35 (0.013) | 5 % |
| San Francisco Xavier, Chihuahua | 261 | 15.76 | 14.94 (0.084) | 14.92 (0.027) | 14.93 (0.017) | 5 % |
| Pojoaque, NM | 270 | 9.14 | 8.52 (0.032) | 8.48 (0.011) | 8.49 (0.007) | 7 % |
| San Buenaventura, NM | 501 | 15.36 | 14.34 (0.055) | 14.29 (0.018) | 14.30 (0.011) | 7 % |
| Santo Domingo, NM | 578 | 13.19 | 12.41 (0.090) | 12.42 (0.028) | 12.43 (0.018) | 6 % |
| Santa Clara, NM | 627 | 17.99 | 17.16 (0.085) | 17.19 (0.027) | 17.17 (0.017) | 5 % |
| Burford | 1,520 | 24.59 | 22.81 (0.158) | 22.78 (0.050) | 22.81 (0.032) | 7 % |
| Warrington | 7,000 | 63.94 | 63.58 (0.126) | 63.61 (0.040) | 63.59 (0.025) | 1 % |
| Boston | 10,565 | 47.13 | 45.19 (0.165) | 45.12 (0.052) | 45.16 (0.033) | 4 % |
| Chester | 12,009 | 68.81 | 67.00 (0.052) | 67.02 (0.016) | 67.03 (0.010) | 3 % |
| Mauritius | 37,110 | 73.95 | 73.93 (0.036) | 72.18 (0.011) | 72.21 (0.007) | 2 % |

† Percent change based on the stochastic estimates from 1,000 runs.

**Figure S1**

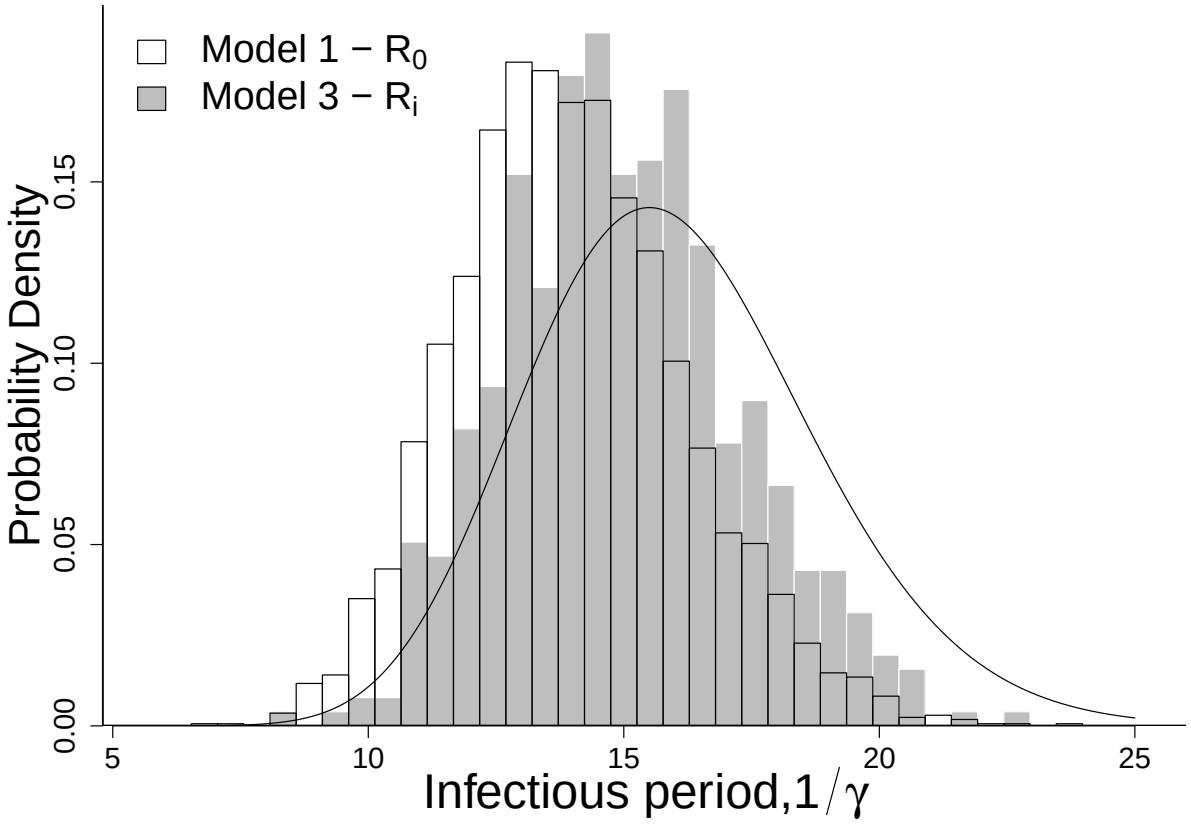


Figure S1: An example of the effects of tight priors on the posterior distribution. A comparison of the mean infectious period $1/\gamma$ distributions given that all epidemics have the same $R_0$ (Model 1) with a model that assumes that each epidemic has its own value of $R_i$ (Model 3). The solid line in the figure represents the prior.

## Figure S2

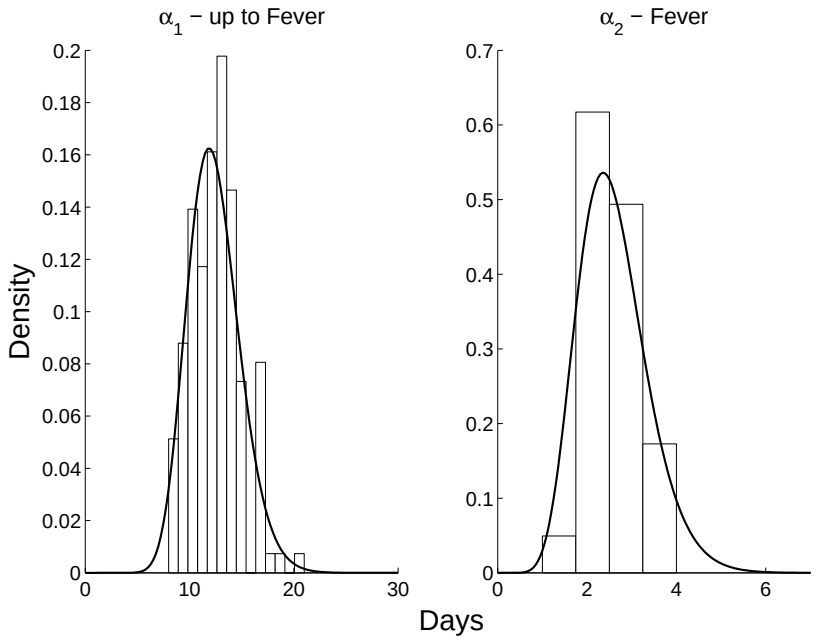


Figure S2: Histogram of the period up to the fever associated with smallpox ($\alpha_1$) and the period of the fever ($\alpha_2$) along with the Gamma distribution that best-fits the data cited in the text.

**Figure S3**

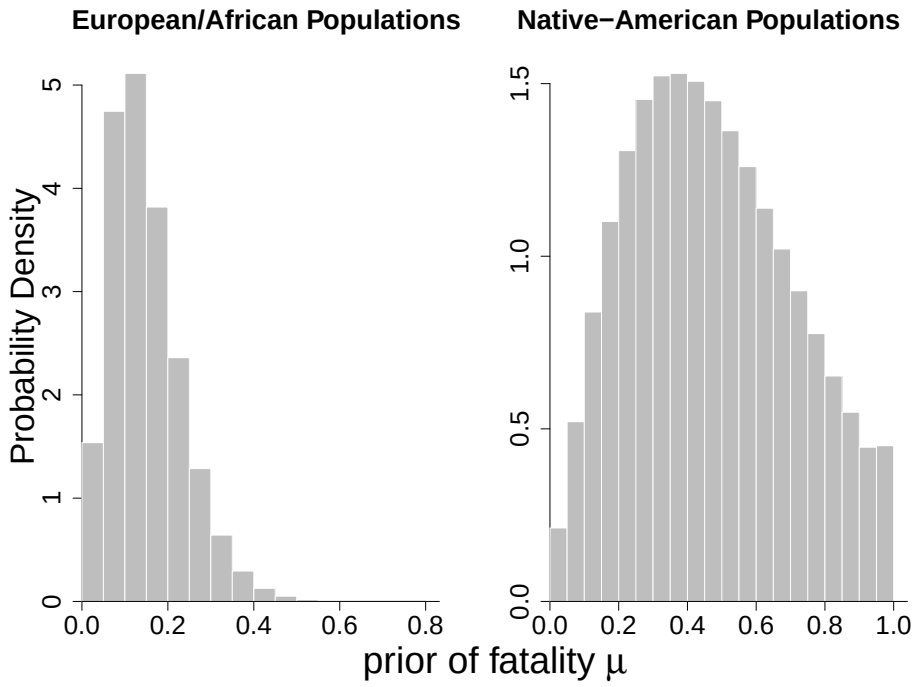


Figure S3: Histogram of the prior of mortality using the hyperprior structure for the European/African and Native-American populations.

## Figure S4

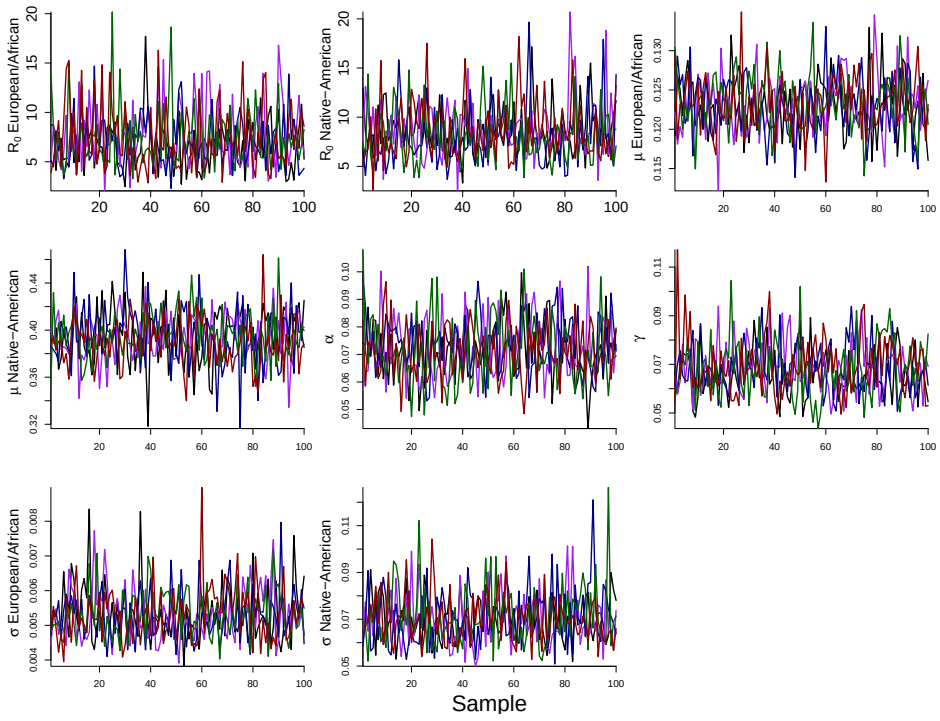


Figure S4: Convergence plots for eight of the main parameters in inter-population disease transmission model (Model 3, in the main text). The colors in each plot represent the five separate chains run to assess model convergence. The plots contain all iterations used to estimate the posterior distributions of the model parameters. Each chain was trimmed by discarding the first 10,000 iterations and thinned by only using every 1,000th draw.

**Figure S5**

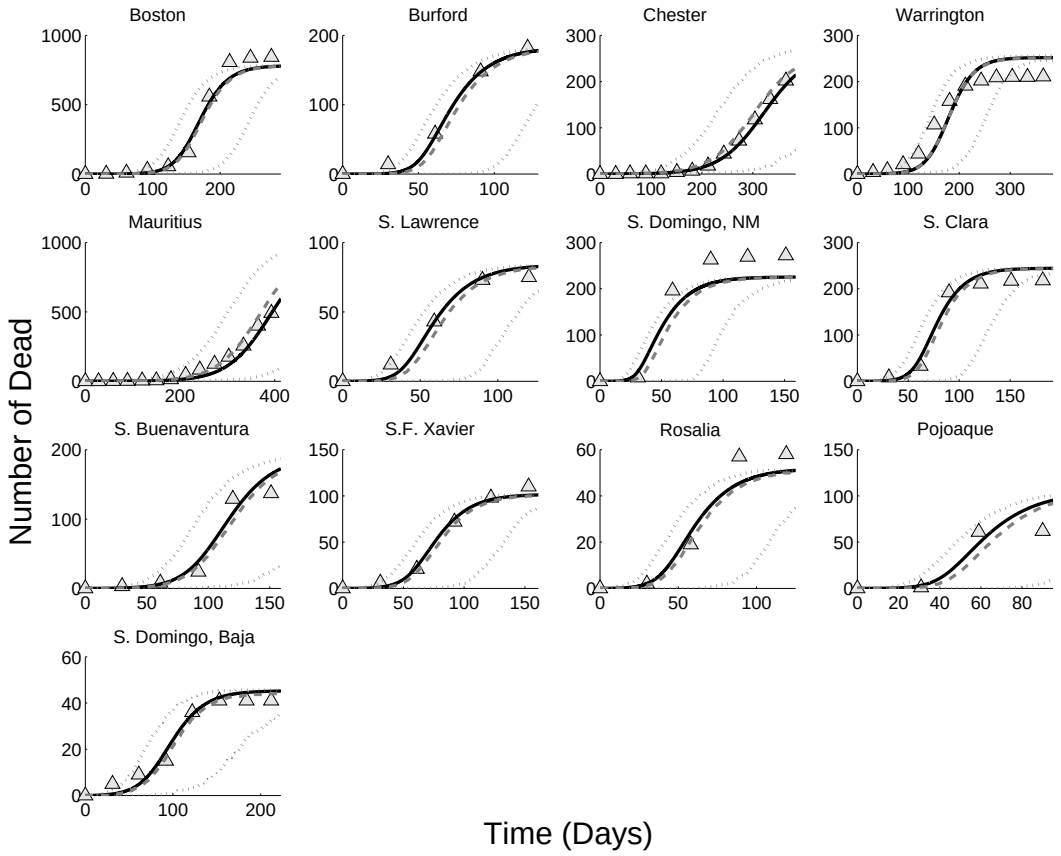


Figure S5: Comparison of the stochastic SEIR model to the deterministic SEIR model for all populations used in the analysis. The black line represents the deterministic output. The gray dashed line is the pointwise median value of the 1,000 iterations of the stochastic model. The gray dotted lines are the 2.5% and 97.5% pointwise bounds of the stochastic model. The gray triangles are the data for each epidemic. Both the stochastic and deterministic models fit the time series well, especially during the exponential phase of the epidemic when mortality is rising rapidly. Note the different scales on the y-axis due to differences in population and epidemic size in each of the outbreaks. The scale also differs on the x-axis due to differences in epidemic length for each population.