SUPPORTING ONLINE MATERIAL

MATERIALS AND METHODS

ANALYTIC DETERMINATION OF R_0

The analytic determination for R_0 was made by considering the proportion infected reported in (*1*) for each platoon containing an index case and each contemporary platoon with at least one serologically defined case:

$$R_0 = \frac{-\ln\left(\frac{1-i_\infty}{(N-1)/N}\right)}{i_\infty}$$

For each of these platoons the variance in the estimate of the proportion of cases was calculated using the normal approximation to the binomial distribution (*2*):

$$\mathrm{VAR}(i_\infty)=\sigma_i^2=\frac{i_\infty(1-i_\infty)}{N}$$

where i_∞ is the final proportion infected and N is the size of the platoon.

For each grouping the point estimate for R_0 was calculated assuming an underlying deterministic epidemic model and the variance for each estimate R_0 was found using the delta method ((*2, 3*)). This gave the following formula for the variance of R_0:

$$\mathrm{VAR}(R_0)=\left(\frac{\partial}{\partial i_\infty}R_0\right)^2\sigma_i^2=\left(\frac{1}{i_\infty^2}\right)\left(\frac{i_\infty}{1-i_\infty}+\ln(1-i_\infty)-\ln(N-1)+\ln(N)\right)^2\sigma_i^2$$

The overall estimate and confidence interval were then found by combining these estimates using inverse variance weighting (i.e. Woolf's method) (*4*) :

$$R_0 = \frac{1}{\sum_1^n \frac{1}{\mathrm{VAR}(R_0^k)}} \sum_{k=1}^{n} \frac{1}{\mathrm{VAR}(R_0^k)} R_0^k$$

where R_0^k is the estimate for R_0 in the kth platoon. Hence, the overall variance is:

$$\mathrm{VAR}(R_0) = \frac{1}{\sum_1^n \frac{1}{\mathrm{VAR}(R_0^k)}}$$

MODEL DEFINITION

Our model is a compartmental model with four sections, susceptibles (S), exposed with latent infection (E), infectious (I), and recovered (R). The model is stochastic and discrete: whole numbers of individuals are moved between compartments based on a random process. Figure 1 shows the path of the individuals through the model compartments. The distribution for the latent periods was selected based on information reported in ((*5*)) and the distribution of the infectious period was selected based on the distribution reported in ((*6*)). The algorithm for the model is as follows:

RunModel($\beta, w_\eta, w_\beta, w_\gamma, n_\mu, n_\sigma, \Delta t, pop$)

$S \leftarrow \{s_1 \cdots s_{pop-1}\}$
$E \leftarrow \varnothing$
$I \leftarrow \{i_1\}$
$R \leftarrow \varnothing$
$T_{S,E} \leftarrow \varnothing$
$T_{E,I} \leftarrow \varnothing$

while $I \neq \varnothing$ and $E \neq \varnothing$

for $\forall x \in I$

pick n from $\text{Binomial}\left(|S|, \Delta t \,\beta / pop\right)$

move n individuals from S to $T_{S,E}$

for $\forall x \in E$

$t_x^E \leftarrow t_x^E - \Delta t$

if $t_x^E \leq 0$ then move x from E to $T_{E,I}$

for $\forall x \in I$

$t_x^I \leftarrow t_x^I - \Delta t$

if $t_x^I \leq 0$ then move x from I to R

for $\forall x \in T_{S,E}$

pick $t_x^E \sim \text{Weibull}\left(w_\eta, w_\beta, w_\gamma\right)$

move x from $T_{S,E}$ to E

for $\forall x \in T_{E,I}$

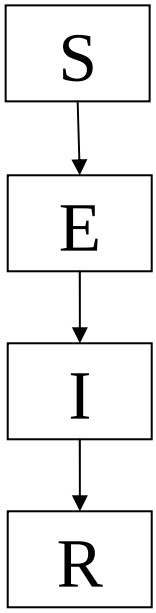
pick $t_x^I \sim \text{logNormal}\left(n_\mu, n_\sigma\right)$

move x from $T_{E,I}$ to I

Where the variables are defined as follows:

- β is the transmissibility
- w_η is the Weibull distribution scale parameter for the latent period.
- w_β is the Weibull distribution shape parameter for the latent period.
- w_γ is the Weibull distribution offset for the latent period.
- n_μ is the mean of the log-normal distribution for the infectious period.
- n_σ is the standard error of the log-normal distribution for the infectious period.

- Δt is the time-step
- *pop* is the population size



Individuals remain in *S* until infected by an $i \in I$, at which point they are moved to *E*.

Individuals remain in *E* for $t \sim \text{Weibull}(\eta, \beta, \gamma)$ days and then move to *I*.

Individuals remain in *I* for $t \sim \text{logNormal}(\mu, \sigma)$ days and then move to *R*.

Once in *R* individuals remain there permanently.

FIGURE 1: Path through the model compartments.

PARAMETERIZATION

Each run of the model was parameterized by R_0 and mean serial interval, with the rest of the model parameters derived from these. Based on the lack of data on when people actually start to transmit influenza and the results of the sensitivity analysis described below we decided to randomly divide the serial interval between latent and infectious periods. For each model run a random division the serial interval between latent and infectious periods was selected. Specifically, the mean infectious period was selected from a uniform distribution in the range $\{1...t_{si} - 0.5\}$ where t_{si} is the desired serial

interval. Based on this selection the values for β and scale parameter (w_η) for the distribution of the latent period are selected. The variance of the infectious period ($n_\sigma = 0.2264$) and the shape and offset of the latent period ($w_\beta = 2.21$ and $w_\gamma = 0.5$) are held constant to values based on ((*6*)) and ((*5*)). The full algorithm is as follows:

ParameterizeModel(R_0, t_{si})

pick $n_\mu \sim \text{Uniform}(1, t_{si} - 1)$

$\beta = R_0 e^{n_\mu}$

$$w_\eta = \frac{t_{si} - w_\gamma - e^{-n_\mu}}{\Gamma\left(1/w_\beta - 1\right)} \quad (*)$$

Where equation (*) is based on the following derivation.

For a constant β over the infectious period:

$$t_{si} = E(\text{latent period}) + E(\text{infectious period}) = w_\gamma + w_\eta \Gamma\left(1/\beta + 1\right) + e^{-n_\mu} \Rightarrow$$

$$w_\eta \Gamma\left(1/\beta + 1\right) = t_{si} - w_\gamma - e^{-n_\mu} \Rightarrow w_\eta = \frac{t_{si} - w_\gamma - e^{-n_\mu}}{\Gamma\left(1/\beta + 1\right)}$$

IMPLEMENTATION AND BATCHED RUNS

The model described above was implemented using the MATLAB mathematical computing environment. Experiments were performed by doing 1,000 batched runs at each parameter setting, for both sensitivity analyses the main analysis.

The "main" simulations were those where we swept the R_0 parameter space between 0.2 and 3.0 at increments of 0.1 and the serial interval space between 1.6 and 10 at increments of 0.1. This sweep was performed both assuming mixing at the platoon (50 person) and company (200 person) levels. In the main simulations the division of the serial interval between infectious and latent periods was chosen as described above, with a new selection for that division made on each run. The main simulations represent 2,352,000 runs of the model over 2,352 different settings for R_0 and serial interval.

Sensitivity Analysis

There were three major sensitivity analyses performed on the model in addition to the sweeping of the parameter space of R_0 and serial interval. We examined the effect of varying Δt, the division of serial interval into infectious and exposed periods, and the effect of changing the size of the mixing population. We found that results for both final epidemic size and epidemic length were insensitive to changes in Δt for values lower than 0.25 (Figure 2), which was selected as the time step for all model runs.

Figure 2a. Sensitivity of epidemic length to Δt

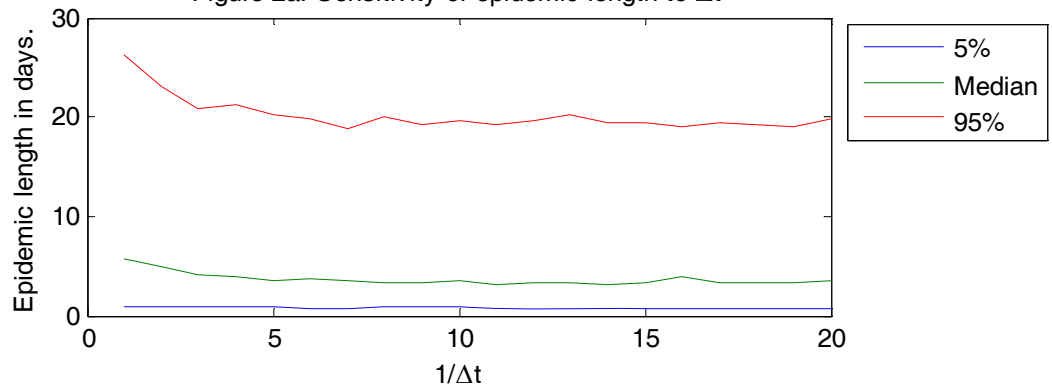
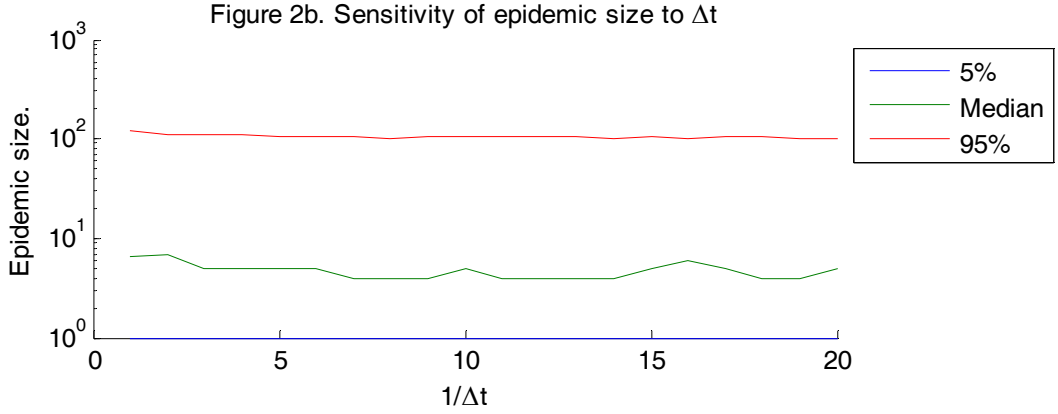


Figure 2b. Sensitivity of epidemic size to Δt



Our analysis shows that the model is mildly sensitive to the division between exposed and infectious periods (Figure 3), with an increase in both epidemic length and epidemic size as more of the serial interval is devoted to the infectious period. This is in part due to the fact that the joint distribution of the infectious and exposed periods is not precisely maintained by our method of splitting the interval (see *Parameterization* section). In order to avoid biasing our results based on the split between infectious and exposed periods we randomly selected the split as described in the *Parameterization* section above.

Figure 3a: median epidemic length in a poulation of 200, R_0=1.5

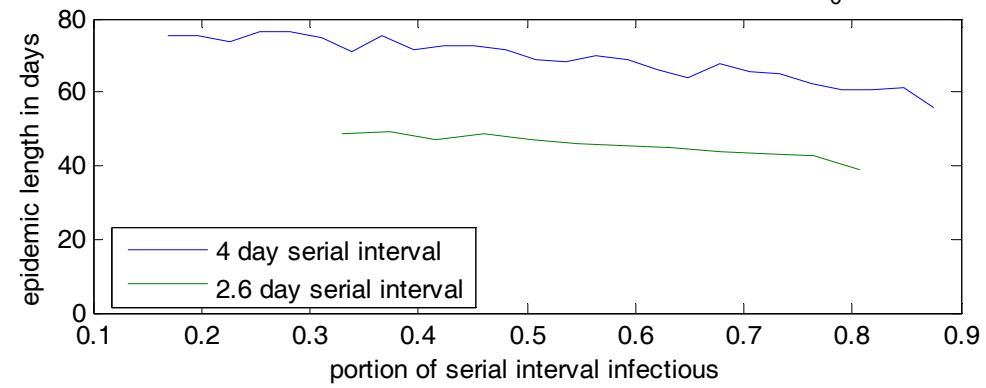
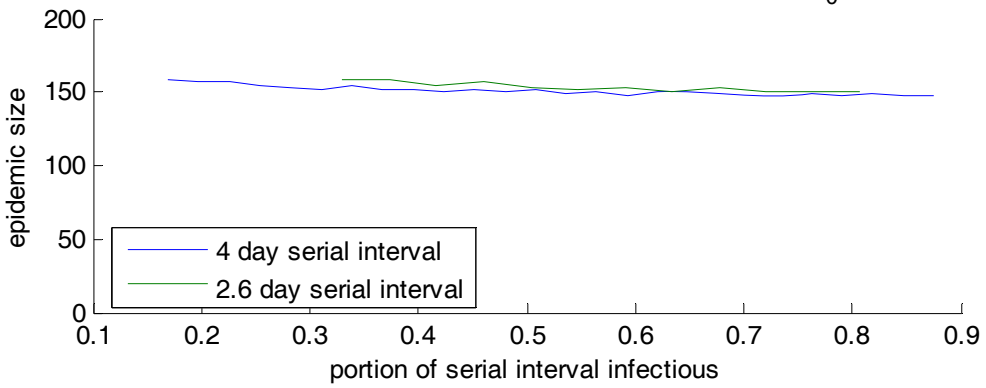


Figure 3b: median epidemic size in a population of 200, R_0=1.5



Analysis of sensitivity to size of the mixing community shows a large effect (Figure 4), as expected. Due to evidence presented in ((*1*)) and the fact that few parameter settings will produce epidemics in the observed range with company level mixing (Figure 5), we assumed platoon level mixing in all our final simulations.

Figure 4a. Difference in median epidemic lengths between comapnies and platoons. (days)

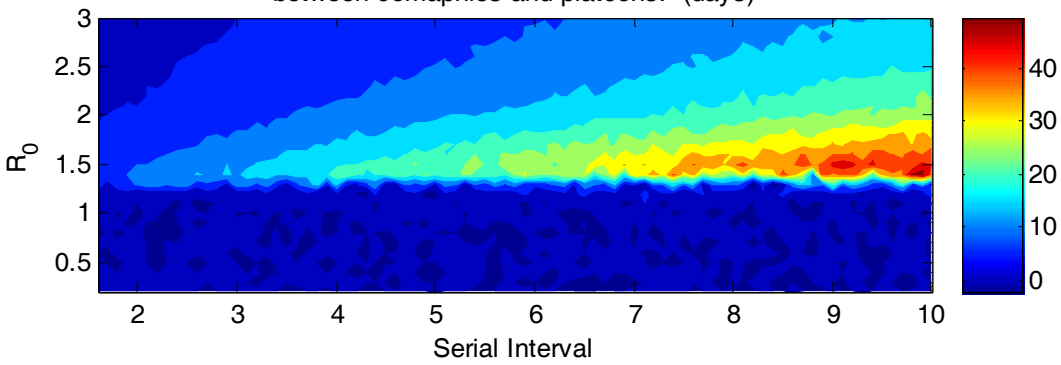


Figure 4b. Difference in median proportion infected between companies and platoons.

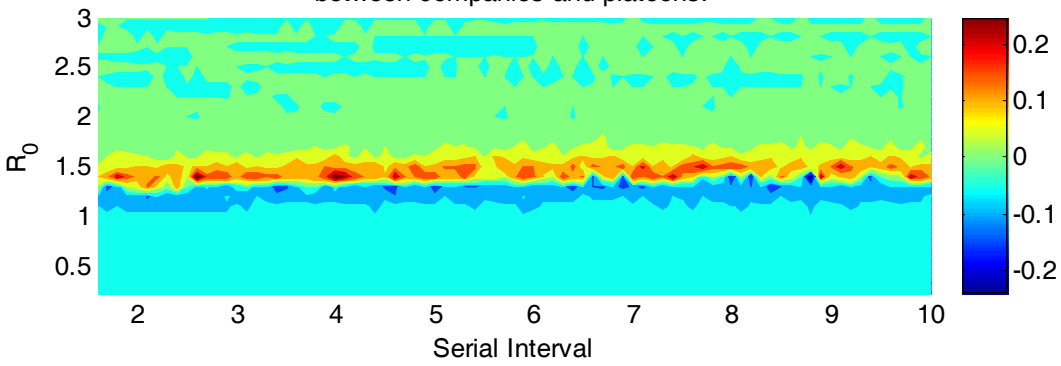
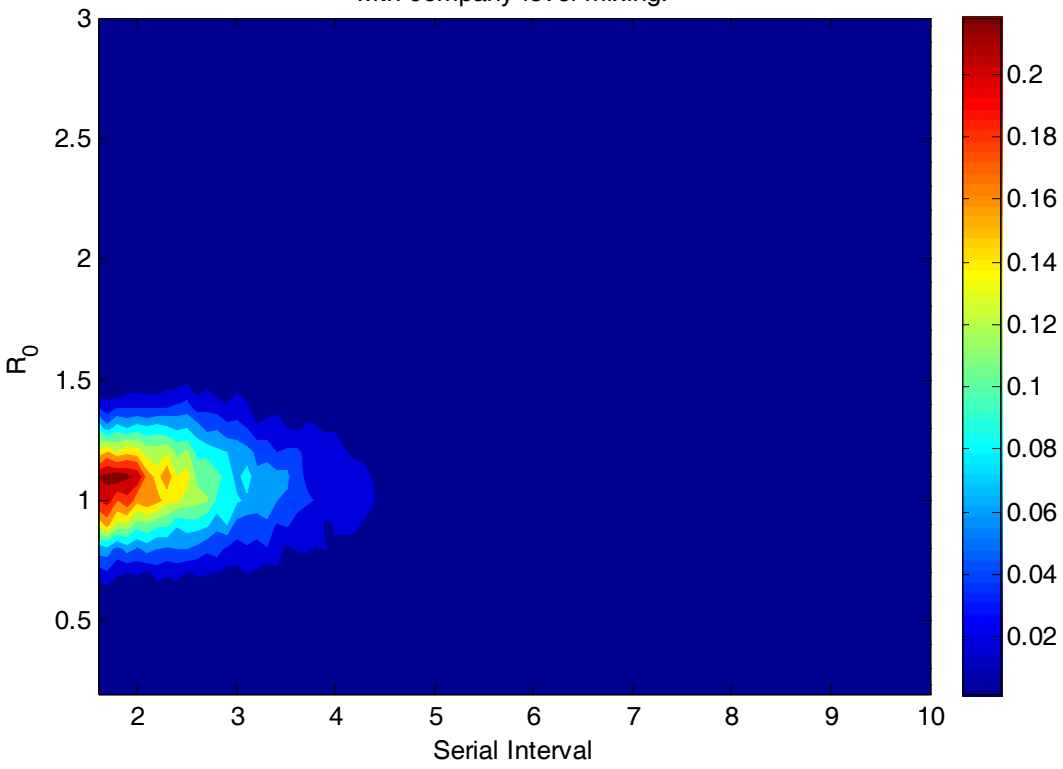


Figure 5: portion of epidemics within the observed range with company level mixing.



Reference List

1. J. C. Gaydos *et al.*, *Journal of Infectious Diseases* **136**, S363 (1977).

2. J. A. Rice, *Mathematical Statistics and Data Analysis* (Duxbury Press, ed. Second, 1995).

3. H. W. Hethcote, *Siam Review* **42**, 599 (2000).

4. K. J. Rothman, S. Greenland, *Modern Epidemiology* (Lippincott Williams and Wilkins, ed. 2nd Ed., 1998).

5. N. M. Ferguson *et al.*, unpublished data.

6. I. M. Longini, Jr., M. E. Halloran, A. Nizam, Y. Yang, *Am. J. Epidemiol.* **159**, 623 (2004).