Classification: Biological Sciences: Ecology; Physical Sciences: Statistics

Estimation of multiple transmission rates for epidemics in heterogeneous populations

Alex R. Cook[1,2], Wilfred Otten[2,3], Glenn Marion[4], Gavin J. Gibson[1], Christopher A. Gilligan[2]

[1]Department of Actuarial Mathematics & Statistics, Heriot-Watt University, Riccarton, Edinburgh EH14 4AS, UK. [2]Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK. [3]Current address: SIMBIOS, University of Abertay Dundee, Dundee DD1 1HG, UK. [4]Biomathematics & Statistics Scotland, James Clerk Maxwell Building, The King's Buildings, Edinburgh EH9 3JZ, UK.

Corresponding author: Alex R. Cook
Department of Plant Sciences,
University of Cambridge,
Downing Street,
Cambridge, CB2 3EA. United Kingdom.
Tel: +44 1223 330229
Fax: +44 1223 333953
Email: a.r.cook@ma.hw.ac.uk

Manuscript information: 13 pages, 2 tables, 4 figures, supplementary materials include 1 further table.

Word and character counts: abstract: 249 words; number of characters in the paper: 43 660.

Contributions to the paper:
All authors were involved in modeling and methodological development and in co-writing the paper. ARC implemented the methodology and carried out the inference and analysis. CAG, GJG and WO identified the original hypotheses, designed the experiment and conceived the inferences with ARC. WO performed the experiment.

Abstract

One of the principal challenges in epidemiological modeling is to parameterize models with realistic estimates for transmission rates, in order to analyze strategies for control and to predict disease outcomes. Using a combination of replicated experiments, Bayesian statistical inference and stochastic modeling we introduce and illustrate a strategy to estimate transmission parameters for the spread of infection through a two-phase mosaic, comprising favorable and unfavorable hosts. We focus on epidemics with local dispersal and formulate a spatially explicit, stochastic set of transition probabilities using a percolation paradigm for an *SI* (susceptible–infected) epidemiological model. The *SI* percolation model is further generalized to allow for multiple sources of infection including external inoculum and host-to-host infection. We fit the model using Bayesian inference and Markov chain Monte Carlo simulation to successive snapshots of damping-off disease spreading through replicated plant populations that differ in relative proportions of favorable and unfavorable hosts and with time-varying rates of transmission. Epidemiologically plausible parametric forms for these transmission rates are compared using the deviance information criterion. Our results show that there are four transmission rates for a two-phase system corresponding to each combination of infected donor and susceptible recipient. Knowing the number and magnitudes of the transmission rates allows the dominant pathways for transmission in a heterogeneous population to be identified. Finally we show how failure to allow for multiple transmission rates can over or underestimate the rate of spread of epidemics in heterogeneous environments, which could lead to marked failure or inefficiency of control strategies.

1 Introduction

One of the principal challenges in epidemiological modeling is to parameterize models with realistic estimates for transmission rates, in order to analyze strategies for control and to predict disease outcomes. Whereas the durations of infectious and latent periods can often be estimated from observation of individuals challenged with inoculum, the probabilities, and the associated rates, for transmission of infection between infected and susceptible individuals are notoriously difficult to measure or estimate (1, 2). The problem is especially acute in spatially structured, heterogeneous host populations, in which hosts differ in susceptibility and infectivity. The magnitudes of the transmission rates typically change over space, according to the nature of the infected donor and the susceptible recipient, and may also change over time (2). In human diseases, infectivity and susceptibility may be affected by genetic, physiological or social differences (3, 4, 5), as well as by immune and vaccination history. Examples also occur in animal epidemiology, with transmission of infection of a common pathogen within and between species at the landscape scale, such as foot and mouth disease in sheep and cattle (6) or in prophylactic treatment of some herds or animals but not others. Analogous examples occur at two scales in cropping systems, either as mixtures of crop species within fields (7, 8) (where the plant is the unit of epidemiological interest), or as a mosaic of crops, such as wheat and barley with differing susceptibility to a common pathogen, arranged within a landscape (where the field is the unit of epidemiological interest) (9, 10).

It is now widely acknowledged that host heterogeneity can endow systems with a far more complex range of dynamics, relating to species and population persistence (9, 11, 12) or evolutionary processes (13), than would be exhibited in homogeneous settings. Even for a two-phase system, there may be as many as four distinct rates arising from each combination of transmission between donor (infected) and recipient (susceptible) individual, and as few as one, if hosts respond homogeneously to infection. Together with the spatial distribution or network structure of favorable and unfavorable hosts in the population, multiple transmission rates determine the spatial and temporal evolution of the epidemic in ways that are quite different from homogeneous transmission. Preferential spread on one component for example, changes the evolution of the contact structure between infected and susceptible sites during the course of the epidemic (Fig. 1). To be able to predict such behavior depends on being able to discriminate differences in transmission rates. Using a combination of replicated experiments, statistical estimation and stochastic modeling we show below how failure to account for heterogeneity in transmission rates by assuming a common (average) transmission rate can seriously misinterpret the dynamics of the epidemic.

< Figure 1 near here >

Specifically, we consider spatially heterogeneous epidemic systems in which pathogen spread occurs through a landscape comprising favorable sites (such as a susceptible or untreated host) and less favorable sites (such as partially-resistant or treated hosts). We focus on processes and models with a strong network structure with short-range interactions. We develop an innovative statistical methodology for the estimation of transmission parameters for a stochastic model of epidemics in heterogeneous populations and test it on successive spatial maps of epidemics from replicated microcosms of mixed populations of radish and mustard seedlings (henceforth labeled favorable, F, and unfavorable, U, respectively) exposed to the fungal plant pathogen *Rhizoctonia solani* Kühn. Soil-borne pathogens are important determinants in the dynamics of plant populations in natural environments (14) and in epidemics in agricultural environments (9, 13). The model is formulated as a stochastic, spatially-explicit set of transition probabilities using a percolation paradigm for an *SI* (susceptible–infected) epidemiological model that has previously been shown to be appropriate for epidemics with short-range, nearest-neighbor contacts on a lattice (15, 16). The *SI* percolation model is

generalized to allow for multiple sources of infection including primary (from external inoculum) and secondary (host to host) infection. In common with many real-life systems such as measles (17), the infection process is subject to temporal forcing, which is accounted for by differential time-varying rates of infection as the two species become resistant to damping-off disease (see Methods). Parameter estimation is effected by fitting the model to observations of disease spread through time and space in replicated epidemics using a combination of Bayesian inference and Markov chain Monte Carlo simulation, adapted from Gibson et al. (2).

Our specific objectives are:

- to develop methods to fit percolation-based, spatio-temporal models for the spread of epidemics that evolve in heterogeneous environments with local dispersal and time-varying transmission rates within and between species;
- to demonstrate methods for formal comparison of models and to test for significant differences amongst transmission rates;
- to use parameter estimates to analyze and predict the effects of different degrees of spatial heterogeneity on disease dynamics;
- to assess whether or not temporal forcing of transmission rates can be detected with the infrequent sampling typical of large scale natural systems.

We also address the effect of large-scale (typified by the percentage of area covered by favorable sites) and small-scale heterogeneity (typified by the degree of clustering of favorable sites) on the connectivity within and between favorable and unfavorable sites, and how connectivity and transmission rates determine disease levels, by controlling the way an epidemic invades its environment.

2 Results

We first identified the parametric forms for time-varying transmission rates (see Methods) from which we conclude that the rate of primary infection declines exponentially with time and the rate of secondary infection rises and falls over time, described by a three-parameter Weibull function. The nuisance parameter used to account for so-called tertiary infection arising from the occasional non-nearest transmission was constant throughout the epidemics. Models were compared using the deviance information criterion (DIC, see Methods and Supplementary Information). Although the particular form of time-dependency for the transmission rates (Fig. 2) reflects the characteristics of the specific system we used to test our method, the methodology introduced below to estimate multiple transmission rates can easily be generalized to any epidemic with or without temporal forcing (see Discussion).

Model fitting and comparison of transmission rates

Bayesian methods were used to estimate parameters and to compare transmission rates for the fully parameterized, spatially explicit, stochastic model for time-varying primary and secondary infection rates. We show that not only was there no evidence to support a common rate amongst and between species but that the rates of secondary infection differed according to which species was the donor and which the recipient (Table 1; Fig. 1). It follows that there are two transmission rates for primary infection, one per host species, but that the differential transmission of secondary infection needs to be explained by four distinct rates depending upon which species is the donor and which is the recipient. The full joint posterior distribution is used to calculate the rates shown in Fig. 2. The corresponding point estimates for parameters together with 95% credible intervals are given in the Supplementary Information. Both parameters governing primary infection differed between species ($\Pr(a_U > a_F \mid \text{data})$

< 0.0001, $\Pr(r_U > r_F \mid \text{data}) < 0.01$). Since both species were challenged by the same source of inoculum, this indicates that although the unfavorable (mustard) sites were initially more susceptible to inoculum, they became resistant more quickly than the favorable (radish) sites (Fig. 2). Interpretation of the various parameters for secondary infection is more complicated, since each set of three parameters contributes jointly to the shape of the secondary transmission rate (Fig. 2). All four within- and between-species transmission rates displayed similar temporal dynamics: the rate of transmission initially increased, followed by a decrease. The absolute values were however significantly different for all four transmission rates (Fig. 2, Table 1).

The largest of the secondary transmission rates unsurprisingly occurred between two favorable sites, followed by the transmission of infection from unfavorable to favorable sites. The rate at which unfavorable sites became infected was lower, in particular for transmission between two unfavorable sites. Overall, however, there were appreciable rates for between-species (*UF, FU*) transmission of infection with differing magnitudes that could not have been predicted from the within species (*UU, FF*) rates (*cf.* Fig. 2).

< Table 1 and Figures 2&3 near here >

We plot the predicted and measured daily distribution of new infections as a measure of goodness-of-fit (Fig. 3). It is striking that with a single set of parameters, the prediction of the number of new infections for each day agrees well with the measured data for all population structures, representing a wide range of heterogeneity. Although the predicted distributions follow the central trend of the observations, the amount of variability in the number of new infections is under-estimated, with more observations falling outside the credible bounds than would be expected. This may indicate environmental differences between replicates that we have not modeled and which could in principle be remedied by taking a hierarchical Bayesian approach (e.g. 18, 19).

Sensitivity of results to frequency of observations

In order to assess whether or not the temporal change in transmission rates could still have been identified with less intensive sampling, we discard some of the data and repeat the model fitting routine. The inferred posterior mean and 95% credible intervals for the transmission rates, taking observations to be at times 4, 8 and 12 rather than daily until day 13, are plotted in Fig. 2 in gray. The correspondence between the analyses based on the full and reduced data sets is remarkably close. The increase in the variability of the estimates from the reduced data set is small compared with the putative decrease in sampling costs.

Identification of dominant pathways for transmission of infection

Estimation of the set of transmission rates for an epidemic in a heterogeneous population of hosts allows us to identify the dominant pathways for infection upon which future efforts to control and manage disease should be targeted. We derive the following posterior distributions: the number of hosts in a mixed population that became infected by primary infection and by infection via secondary infection from the same species or class (e.g. favorable–favorable or unfavorable–unfavorable) and between species (e.g. favorable–unfavorable or unfavorable–favorable) (Table 2). In heterogeneous populations, a population with 75% favorable sites (Fig. 4a) provides a well-connected network of neighboring favorable sites that is preferentially exploited by the epidemic. Unfavorable sites become infected predominantly from neighboring favorable sites and they contribute little to transmission of infection. In contrast, for a population with 50% favorable sites (Fig. 4d) the connectivity of the favorable network is reduced and the contribution of unfavorable sites to the spread of epidemics increased significantly (Table 2). Only by fitting spatial models is it possible to obtain estimates of the most likely pathways for transmission of infection.

< Table 2 near here >

The effect of heterogeneity in population structure

Model inferences were used with the estimated parameters to analyze the effect of heterogeneity at two spatial scales: at the large scale, exemplified by the area of the population covered by favorable sites, and at the small scale, exemplified by the clustering of favorable sites within the population. The intuitive expectation here is that an increase in the proportion or clustering of a sub-population through which disease spreads will lead to an increase in infection as the number of contacts to the same host type increases and a better connected network for transmission of infection is formed (Fig. 4). The proportion of infected hosts in the favorable sub-population did indeed increase, irrespective of transmission *between* sub-populations (Fig. 4b,c). However, the combined effect of the four transmission rates for secondary infection and the evolution of contacts between favorable and unfavorable sites resulted in counter-intuitive responses in the unfavorable sub-population. Here, an increase in the proportion of the unfavorable population resulted in either more or less infection, depending upon the relative magnitude of the transmission rates between the two sub-populations (Fig. 4e). A similar dichotomous response resulted from reducing the amount of local clustering (Fig. 4f). We conclude that the effect of a better connected network in the unfavorable sub-population is off-set against a decrease in the number of contacts with the favorable sub-population within which disease can spread faster. Additional simulations (not shown) showed that this pattern depends on neither the level of primary infection nor the time-dependency of the transmission rates, but that it does depend on the relative values of the transmission rates.

< Figure 4 near here >

3 Discussion

Transmission rates for epidemics are notoriously difficult to quantify (20) and resort is often made to indirect methods of estimation. Our approach introduces and tests a framework for direct estimation of transmission rates from spatio-temporal snapshots of disease spread through heterogeneous populations. By integrating experimentation, modeling and parameter estimation using Bayesian estimation coupled with Markov chain Monte Carlo simulation, the framework allows identification and analysis of the processes that underlie the spread of disease in heterogeneous populations. We have shown that the following are possible: (i) estimation of multiple transmission rates from spatio-temporal data of disease in heterogeneous (two-phase) environments, (ii) formal comparison of models and tests for significant differences amongst transmission rates, (iii) identification of the main sources and pathways of infection, (iv) analysis, using parameter estimates, of how epidemics evolve in response to changes to the heterogeneity of their environment. While the framework was tested for a full dataset of 13 successive spatio-temporal maps for which the observational time-scale was well matched to the biological time-scale of the system, we have also shown that the method is robust to drastic reduction in the number (three) of snap-shots. Small numbers of snap-shots are much more typical of epidemics in natural systems (21, 22) and the robustness of the methods to these small samples supports the generality of the statistical framework to non-model systems.

Knowing the number and magnitude of transmission rates enabled us to identify the dominant pathways for transmission of infection (Fig. 4). Failure to allow for multiple transmission rates could grossly under- or over-estimate the rate of spread of disease, especially in inferring the effects of changing the proportion of favorable or unfavorable sites associated with disease control strategies in a population. For example, at the landscape scale, we may consider vaccinating hosts spatially to control diseases such as foot and mouth (23); the choice of whether to vaccinate all susceptible animals or just

cattle, say—without prior knowledge of the resulting transmission rates for the epidemic—could lead to marked failure or inefficiency in prophylactic use.

We developed our methods for a generic model for epidemics (9) where hosts are classified as either susceptible (S) or infected (I) in order to illustrate the approach. Although the model does not include any hidden classes, such as latent infections (1), extra classes can, in principle, be accommodated. The model and approach also readily generalize to multiple phases or host types. If h_j, the heterogeneity covariate of host j, remains categorical, the extension is obvious, although the number of parameters increases quadratically with the number of phases. If h_j represents a continually varying trait, then some functional form (24) must be imposed for the rate $\beta[h_i,h_s](t)$ of infection from i to s: for example, that $\beta[h_i,h_s](t)$ is proportional to h_s and constant with time (25).

The plant pathogen experimental microcosm system used here is well suited to testing methodological advances (1, 2). It is repeatable yet introduces stochasticity to replicated epidemics in unpredictable ways that reflect well the uncertainty of biological systems not always captured by computer simulation. Epidemics are short with completion in 15–20 days and allow repeated observations usually at daily intervals of evolving replicated epidemics. Even in the model system, complexities arise. The rates of primary infection decay exponentially with time; the rise and fall in the rates of secondary infection are consistent with changes in infectivity as plants grow and become stronger donors, offset by increasing resistance as plants age (26). These results are consistent with analyses of epidemics in analogous homogeneous systems, and biological interpretations associated with changing dynamics of host susceptibilities are detailed elsewhere (26). The important feature to note here is that time-dependency of transmission rates is often not known *a priori*, yet plays a crucial role in the dynamics of epidemics, especially when it leads to rapid quenching of disease spread (27). Our method enables formal comparison of models to identify appropriate functions for time-dependency of transmission rates.

Time-varying infection rates are not limited to plant diseases. Rates may change as a result of policy (e.g. culling of livestock (28), the imposition of travel restrictions (29)), social reaction to the threat of infections (29), environmental changes (30) or seasonal variation (17). Host-age (31) and viral load (32) can also have an effect on transmissibility of infection. The detection of changes depends on sufficient information being present in the available data, but is not as dependent on frequent sampling as might be expected: repeating the analysis in this paper but omitting 75% of the observation times yields very similar results (Fig. 2).

Although the focus of this paper has been on crop mixtures, marked differences in transmissibility of infection within and amongst classes of hosts are also important determinants of the outcome of disease outbreaks in human and other animal populations. Typical examples include species differences in transmission of the foot and mouth pathogen (28) and polymorphism and recombinational hotspots in susceptibility to malaria (3). Sexual orientation, behaviour and partner choice impose heterogeneities in relation to sexually-transmitted diseases (5, 4) as do age (31, 4) and sex (33, 4) for a range of other diseases. Until now, it has been very difficult to parameterize models to take account of such heterogeneities, despite their implicit importance in the dissemination and control of disease, as for example in the recent outbreak of SARS (29).

The main feature of the framework introduced here is that it allows for analysis of epidemics in heterogeneous environments whilst accounting for two crucial underlying aspects, namely the contact between sites in the population, and real estimates of multiple transmission rates that operate in heterogeneous systems. We have shown that both factors affect the dynamics of epidemics and hence

the effectiveness of disease-control strategies. Such strategies could include spatially-explicit control by shielding susceptible hosts, fields or even farms for example by the spatial deployment of resistant varieties, or a local deployment of a chemical or biological control agent (8, 10, 34). The effectiveness is largely determined by the relative magnitude of the transmission rates. If transmission rates between favorable and unfavorable sites are intermediate or high, we have shown that the levels of disease in sites that are less favorable for spread are dominated by the disease pressure from the favorable sites in the population (Fig. 3). Hence the heterogeneity of the population determines the underlying landscape within which contacts between infected and susceptible sites, either favorable or unfavorable, are subsequently dynamically generated by the pathogen as it explores specific contacts preferentially. The latter is mainly determined by the relative magnitude of the transmission rates. Knowledge of multiple transmission rates as estimated in this paper is therefore essential in addressing epidemiologically important questions such as: '*what is the minimum spatial coverage of a vaccination treatment required to reduce the risk of invasion?*' (23); or '*can we safely introduce a susceptible crop or cropping system (e.g. organic farms) without enhancing the risk of invasive spread at the regional scale?*' (10). These questions can only be addressed within spatial models. Our method ensures that answers to such questions are statistically sound, and fully integrated with experimental trials or field data.

4 Methods

Model structure

The model of Gibson et al. (2) can be readily generalized to represent a mixed-species population. The heterogeneity of the population is described by assigning the covariate h_j to each member j of the population, where h_j which takes the values 0, 1 for favorable or unfavorable sites for disease transmission. Let $I_j(t) = 1$ if j is infected by time t and 0 if still susceptible. In common with most stochastic epidemiological models (1, 35, 36), we assume that $\Pr\left(I_s(t+\mathrm{d}t)=1 \middle| I_s(t)=0\right)=\phi_s(t)\mathrm{d}t$ as $\mathrm{d}t\to 0$, where $\phi_s(t)$ is the rate of infection of s. In the specific model used here, $\phi_s(t)$ is comprised of terms representing the rate of *primary* infection from inoculum at time t if s is inoculated, denoted $\alpha[h_s](t)$, and the rate of *secondary* infection from each infected neighbor i, denoted $\beta[h_i,h_s](t)$. We restrict the transmission of primary and secondary infection to nearest neighboring sites only, to accommodate the limited dispersal commonly found for soil-borne pathogens and for which data for model testing were available. A nuisance parameter, termed the rate of *tertiary* infection, $\gamma[h_s](t)$, is introduced to allow for a small proportion of non-nearest neighbor transmission. Using the indicator function $\mathbf{1}\{A\} = 1$ if A is true and 0 otherwise, this can be written:

$$\phi_s(t)=\alpha[h_s](t)\mathbf{1}\{s\in X\}+\left(\sum_i \beta[h_i,h_s](t)\mathbf{1}\{I_i(t)=1\}\mathbf{1}\{i\in N_s\}\right)+\gamma[h_s](t)$$

where X and N_s are the sets of inoculated hosts and nearest-neighbors of s, respectively. The total rate of infection $\phi_s(t)$ of a host s therefore depends upon the time-varying transmission rates as well as on localized conditions (presence of inoculum, neighboring infectious hosts) which evolve with time and are different for each host (figure 1). A range of functional forms were tested for the transmission rates with the following emerging with strong support from the deviance information criterion (see below):

$$\alpha[h_s](t)=a[h_s]\exp\left(-r[h_s]t\right),$$

$$\beta[h_i,h_s](t)=b[h_i,h_s]\kappa[h_i,h_s]\left\{\left(t/\mu[h_i,h_s]\right)^{\kappa[h_i,h_s]}\right\}\exp\left\{-\left(t/\mu[h_i,h_s]\right)^{\kappa[h_i,h_s]}\right\}/t\ ,$$

$$\gamma[h_s](t)=\varepsilon[h_s]\ ,$$

in which the rate for primary infection decays exponentially with time, the rate for secondary infection changes non-monotonically in accordance with a Weibull function and the rate for background

infection is constant. The Weibull function is selected here as a function allowing the expression of rise-and-fall dynamics with analytical integrals.

Bayesian model fitting

We fitted the model using Markov chain Monte Carlo techniques and data augmentation to draw a sample from the joint posterior distribution of model parameters and unobserved infection times and sources of infection for each of the circa 10 000 hosts in the system. The algorithm is quite involved, so details are provided in the web appendix. The deviance information criterion (37, 38) was used to discriminate between competing models.

Experimental data

Our methods are applied to epidemics in heterogeneous plant populations in replicated microcosm experiments. Details of the experiments can be found in Otten *et al.* (39). In summary, dynamics of damping-off epidemics were recorded in populations comprising 414 seedlings of a favorable (radish, *Raphanus sativus* L., *Cherry Belle*) or an unfavorable species (mustard, *Sinapsis alba* L.), planted in a square lattice. At the densities used, spread of disease occurs predominantly between nearest neighbors. Populations comprised either 100% favorable, 100% unfavorable, a mixture with 75% favorable and 25% unfavorable, or a 50:50% mixture, with up to six replicates per treatment. The host species at each point on the lattice was randomly selected, and in each tray 32 randomly selected plants were challenged by inoculum of the soil-borne fungal pathogen *R. solani*. The position of damped-off plants was recorded daily for 13 days following emergence; to assess the dependency of the method on high-resolution temporal data, we also consider a censoring of the data with observations at 4d intervals only. The model was fitted to all replicates jointly.

Acknowledgements

ARC, GJG and CAG thank the Biotechnology and Biological Sciences Research Council (BBSRC) for financing the research project (grant number BB/C007263/1). The experimental data were collected by WO in a previous joint BBSRC project with CAG and GJG. CAG also gratefully acknowledges the support of a BBSRC professorial fellowship and GM acknowledges support from the Scottish Executive. We thank two anonymous referees whose comments have improved the manuscript.

References

1. Gibson GJ, Kleczkowski A & Gilligan CA (2004) *Proc Natl Acad Sci USA* 101:12120–12124.
2. Gibson GJ, Otten W, Filipe JAN, Cook A, Marion G & Gilligan CA (2006) *Statistics and Computing* 16:391–402.
3. Kwiatkowski DP (2005) *Am J Hum Genet* 77:171–190.
4. Quinn TC & Overbaugh J (2005) *Science* 308:1582–1583.
5. Gregson S, Nyamukapa CA, Garnett GP, Mason PR, Zhuwau T, Caraël M, Chandiwana SK & Anderson RM (2002) *The Lancet* 359:1896–1903.
6. Ferguson NM, Donnelly CA & Anderson RM (2001) *Nature* 413:542–548.
7. Finckh MR, Gacek ES, Goyeau H, Lannou C, Merz U, Mundt CC, Munk L, Nadziak J, Newton AC, de Vallavieille-Pope C*, et al.* (2000) *Agronomie* 20:813–837.
8. Garrett KA & Mundt CC (1999) *Phytopathology* 89:984–990.
9. Gilligan CA (2002) *Adv Bot Res* 38:1–64.
10. Gilligan CA (2008) *Phil Trans R Soc London Ser B* (DOI:10.1098/rstb.2007.2181).
11. Levin SA (2000) *Ecosystems* 3:498–506.

12. Ettema CH & Wardle DA (2002) *Trends Ecol Evol* 17:177–183.
13. Shea K, Thrall PH & Burdon JJ (2000) *Ecol Lett* 3:150–158.
14. Packer A & Clay K (2000) *Nature* 404:278–281.
15. Bailey DJ, Otten W & Gilligan CA (2000) *New Phytol* 146:535–544.
16. Sander LM, Warren CP & Sokolov IM (2003) *Physica A* 325:1–8.
17. Bjørnstad ON, Finkenstädt BF & Grenfell BT (2002) *Ecol Monographs* 72:169–184.
18. Gross K, Ives AR & Nordheim EV (2005) *Ecology* 86:740–752.
19. Link WA & Sauer JR (2002) *Ecology* 83:2832–2840.
20. Dwyer G & Elkinton JS (1993) *J Anim Ecol* 62:1–11.
21. Gibson GJ (1997) *Appl Stat.* 46:215–233.
22. Keeling M, Brooks SP & Gilligan CA (2004) *Proc Natl Acad Sci USA* 101:9155–9160.
23. Keeling MJ, Woolhouse ME, May RM, Davies G, Grenfell BT (2003) *Nature* 421:136–142.
24. Cook A, Marion G, Butler A & Gibson G (2007) *Bull Math Bio* 69:2005–2025.
25. O'Neill PD & Becker NG (2001) *Biostatistics* 2:99–108.
26. Otten W, Filipe JAN, Bailey DJ & Gilligan CA (2003) *Ecology* 84:3232–3239.
27. Kleczkowski A, Bailey DJ & Gilligan CA (1996) *Proc R Soc London Ser B* 263:777–783.
28. Chis Ster I & Ferguson NM (2007) PLoS ONE 2:e502.
29. Riley S, Fraser C, Donnelly CA, Ghani AC, Abu-Raddad LJ, Hedley AJ, Leung GM, Ho LM, Lam TH, Thach TQ, *et al.* (2003) Science 300:1961–1966.
30. Gottwald TR, Sun X, Riley T, Graham JH, Ferrandino F & Taylor EL (2002) *Phytopathology* 92:361–377.
31. Ferguson NM, Donnelly CA & Anderson RM (1999) *Phil Trans R Soc London Ser B* 354:757–768.
32. Gray RH, Wawer MJ, Brookmeyer R, Sewankambo NK, Serwadda D, Wabwire-Mangen F, Lutalo T, Li X, vanCott T, Quinn TC & the Raka Project Team (2001) *The Lancet* 357:1149–1153.
33. Madkan VK, Giancola AA, Sra KK & Tyring SK (2006) *Arch Dermatol* 142:365–370.
34. Brophy LS & Mundt CC (1991) *Agric Ecosyst Environ* 35:1–12.
35. Renshaw E (1991) *Modelling Biological Populations in Space and Time* (Cambridge Univ Press, Cambridge, UK).
36. Höhle M, Jørgensen E & O'Neill PD (2005) *Appl Stat* 54:349–366.
37. Celeux G, Forbes F, Robert CP & Titterington DM (2006) *Bayesian Analysis* 1:651–674.
38. Spiegelhalter DJ, Best NG, Carlin BR & van der Linde A (2002) *J R Stat Soc B* 64:583–616.
39. Otten W, Filipe JAN & Gilligan CA (2005) *Ecology* 86:1948–1957.

Tables:

Table 1: Difference in deviance information criteria (DIC) relative to the best-fitting model, used to confirm absence of common parameters for secondary transmission rates, associated with the donor, the recipient, or both. The estimated effective number of parameters is denoted $\hat{p}$. For brevity the time dependence of β is omitted from the table.

Model		DIC	$\hat{p}$
Full		0	21.0
Donors same	$\beta[FF]=\beta[UF]$; $\beta[FU]=\beta[UU]$	25	15.7
Recipients same	$\beta[FF]=\beta[FU]$; $\beta[UF]=\beta[UU]$	482	11.5
Cross same	$\beta[FU]=\beta[UF]$	50	17.3
All rates same	$\beta[FF]=\beta[FU]=\beta[UF]=\beta[UU]$	921	7.4

Table 2: Identification of the most likely pathways and sources of transmission of infection in mixed populations, estimated by fitting the model to replicated epidemics of mixed populations with different ratios of favorable (*F*) and unfavorable (*U*) plants. The values show the posterior mean (expressed as % of *F* or *U* present in the population) of plants that became infected by primary infection, secondary infection from favorable plants, or secondary infection from an unfavorable plant. Tertiary infection accounted for the infection of a further 11–15% of favorable and 4–5% of unfavorable plants. Standard deviations are given in parentheses.

	100% *F*	75% *F*		50% *F*		100% *U*
Species	*F*	*F*	*U*	*F*	*U*	*U*
Primary infection	4.6 (0.2)	4.5 (0.3)	5.8 (0.3)	4.2 (0.3)	5.8 (0.2)	5.7 (0.2)
Secondary transmission from *F*	58.4 (0.5)	45.4 (0.7)	27.1 (0.9)	25.7 (0.7)	14.5 (0.6)	–
Secondary transmission from *U*	–	5.0 (0.5)	2.0 (0.5)	13.1 (0.6)	9.3 (0.5)	7.9 (0.3)

Figures:

Figure 1: Dynamics of infection in a heterogeneous host population comprising unfavorable (*U*, gray squares) and favorable (*F*, white) sites. The state of an epidemic is shown for two time points, with solid circles indicating infected and hollow circles susceptible hosts. The arrows represent potential transmission routes between infected and neighboring susceptible sites. The figure shows how the localized conditions change with time, and how the disease load depends on the relative strength of each of the four possible transmission rates $\beta[.,.]$. In the left picture, the total infective challenge on *F* is $4\beta[F,F]$ and $3\beta[F,U]$ on *U*. As more hosts become infected (right), this changes to $3\beta[F,F]+\beta[U,F]$ on *F* and $2\beta[F,U]+2\beta[U,U]$ on *U*. Even though the number of transmission pathways has increased from seven to eight, the combined effective challenge may be greater or less than in the left-hand picture depending upon the relative strength of each of the four transmission rates.

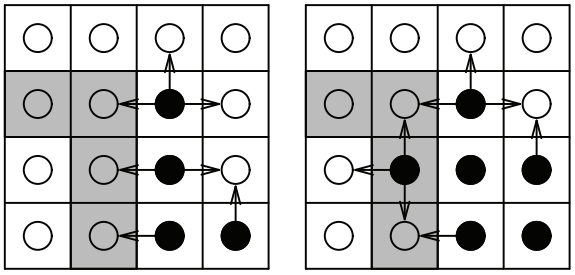


Figure 2: posterior means (solid curves) and 95% credible intervals (dashed) for primary ($\alpha[F](t)$, $\alpha[U](t)$) and each of the four secondary transmission rates ($\beta[FF](t)$, $\beta[FU](t)$, $\beta[UF](t)$, $\beta[UU](t)$) against time. In black are results using the full spatio-temporal data (with observations at times indicated by black circles in the bottom-right plot), in gray are the corresponding results having discarded all but three observations at times 4, 8 and 12.

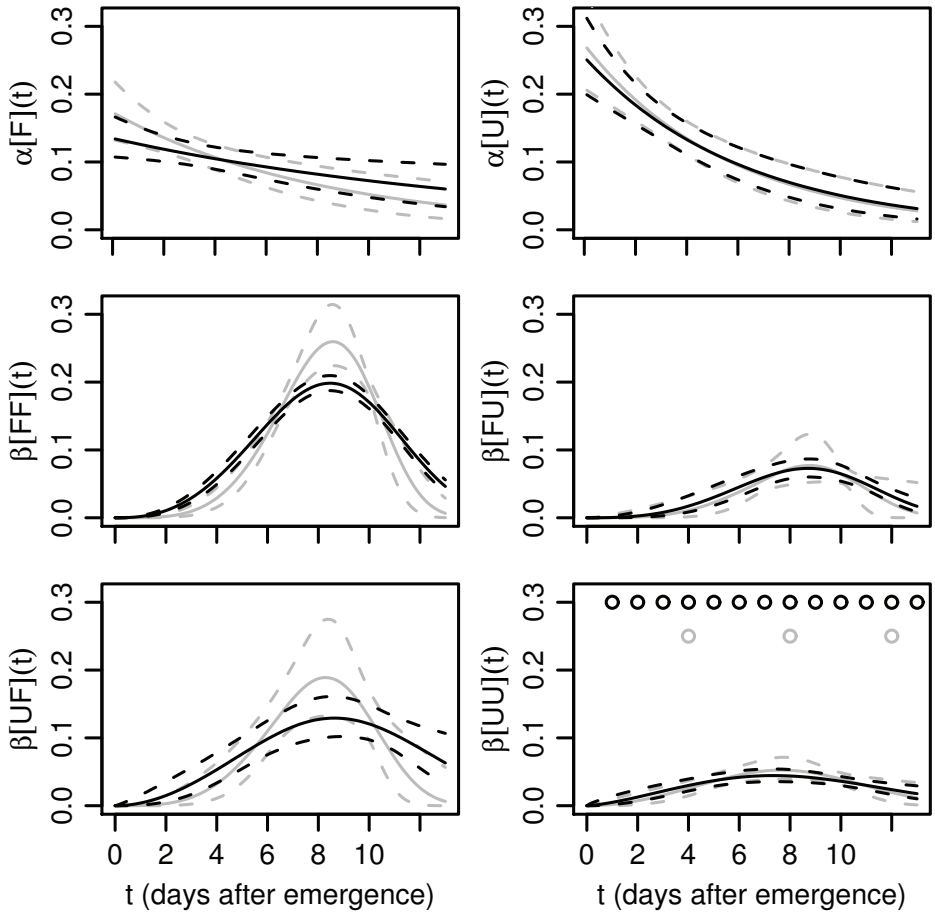


Figure 3: Posterior predictive new infections of daily increments (shaded region corresponding to credible intervals) with observations (points, area of symbols are proportional to the number of observations) for damping-off epidemics in mixed populations comprising 100%, 75%, 50% or 0% favorable hosts. Predictions take the form of mixture distributions with each component conditional on the previous spatial observation of disease presence in an experimental replicate. Predictions take account of both parametric uncertainty and population stochasticity. Denote by $I_F(t)$ and $I_U(t)$ the number of infected favorable and unfavorable hosts at time t, respectively.

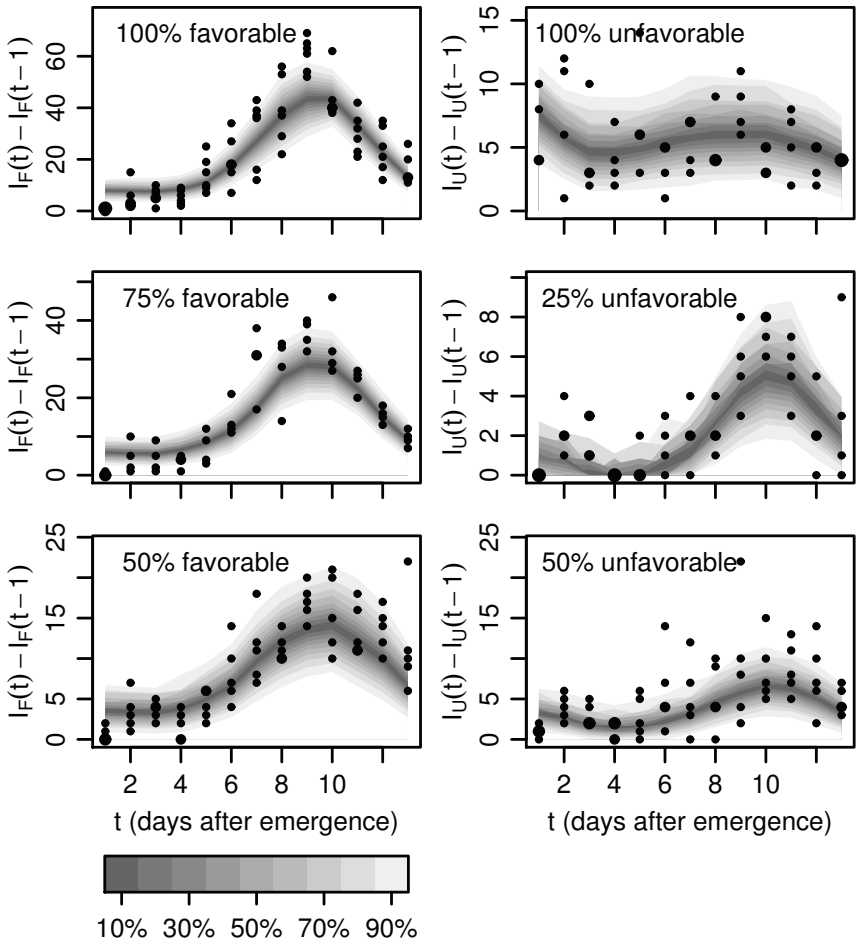
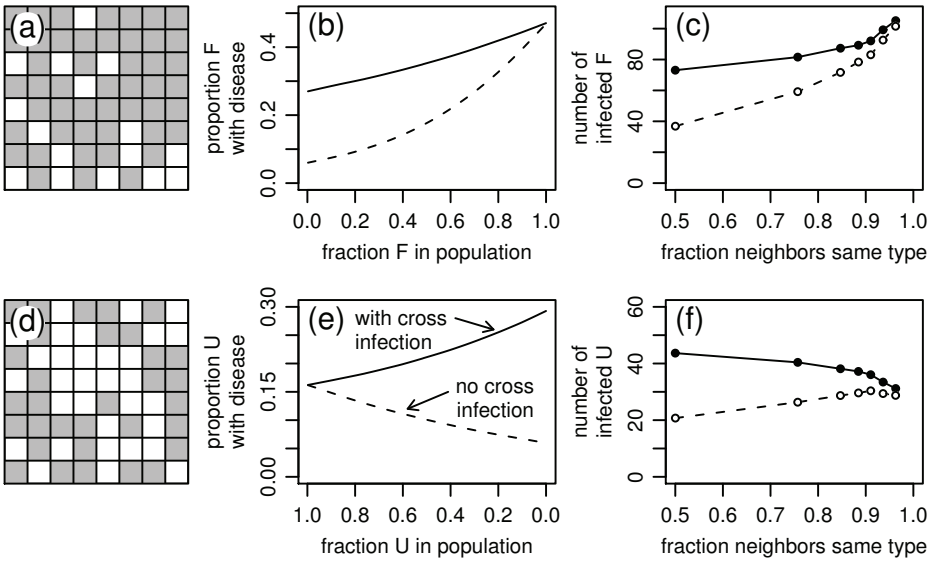


Figure 4: Nature and effects of the spatial structure of heterogeneity. When 75% of the population are favorable (a, gray squares), they form a well-connected network. At 50%, both favorable and unfavorable (d, white squares) sites form isolated patches. The effect of the fraction of sites in the population that are favorable (b) or unfavorable (e) and the proportion of neighbors of the same type (c&f), a measure of clustering, on levels of disease, with (continuous lines, solid circles) or without (dashed lines, hollow circles) inclusion of transmission of disease between favorable and unfavorable sites in the population.



ONLINE SUPPORTING INFORMATION

Classification: Biological Sciences: Ecology; Physical Sciences: Statistics

Estimation of multiple transmission rates for epidemics in heterogeneous populations: supplementary material

Alex R. Cook[1,2], Wilfred Otten[2,3], Glenn Marion[4], Gavin J. Gibson[1], Christopher A. Gilligan[2]

[1]Department of Actuarial Mathematics & Statistics, Heriot-Watt University, Riccarton, Edinburgh EH14 4AS, UK. [2]Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK. [3]Current address: SIMBIOS, University of Abertay Dundee, Dundee DD1 1HG, UK. [4]Biomathematics & Statistics Scotland, James Clerk Maxwell Building, The King's Buildings, Edinburgh EH9 3JZ, UK.

Corresponding author: Alex R. Cook
Department of Plant Sciences,
University of Cambridge,
Downing Street,
Cambridge, CB2 3EA. United Kingdom.
Tel: +44 1223 330229
Fax: +44 1223 333953
Email: a.r.cook@ma.hw.ac.uk

Appendix 1: Additional details of methods

Bayesian model fitting

Following Gibson et al. (1), we fit the model by deriving the form for the likelihood function of the parameter vector (denoted by $\boldsymbol{\theta}$) conditioned on the sets of non-observed infection times ($\boldsymbol{t} = \{\tau_s = \min t : I_s(t) = 1\}$) and sources of infection ($\boldsymbol{\sigma}$). The posterior distribution for $\boldsymbol{\theta}$ given the observed replicate data (D) is $f(\boldsymbol{\theta}|D) \propto f(\boldsymbol{\theta}) \iint f(\boldsymbol{t}, \boldsymbol{\sigma}|\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\sigma}\mathrm{d}\boldsymbol{t}$ with the region of integration being limited to the set of ($\boldsymbol{t}$, $\boldsymbol{\sigma}$) consistent with D. Following (2), the $f(\boldsymbol{t}, \boldsymbol{\sigma} \mid \boldsymbol{\theta})$ term is defined by:

$$f(\boldsymbol{t}, \boldsymbol{\sigma}|\boldsymbol{\theta}) = \prod_{s:I_s(t_{end})=1} \lambda_s(\tau_s) \exp\left\{-\int_0^{\tau_s} \phi_s(u)du\right\} \prod_{s:I_s(t_{end})=0} \exp\left\{-\int_0^{t_{end}} \phi_s(u)du\right\}$$

where λ_s is the rate of infection exerted on s by the infecting source, σ_s. This follows from the definition of a Markov process (2). A detailed derivation for a related model with homogeneous hosts can be found in Gibson *et al.* (1).

It is preferable to use functional forms for $\phi_s(t)$ that have analytical integrals so that the survival terms above can readily be evaluated. Examples of such forms include the constant, exponentially declining and Weibull rates used in this paper.

The prior, $f(\boldsymbol{\theta})$, is a subjective description of our beliefs regarding the parameters ($\boldsymbol{\theta}$). We take relatively broad, independent priors, namely exponential with mean 10 for all parameters except those for tertiary infection, which are exponential with mean 0.1.

To evaluate the 10 000-dimensional posterior, we use Markov chain Monte Carlo techniques (see (3) for a review) to draw a sample from the joint distribution of ($\boldsymbol{\theta}$, $\boldsymbol{t}$, $\boldsymbol{\sigma}$) that then provides an estimate of the posterior distribution of $\boldsymbol{\theta}$ from which summary statistics were made (4, 5). Summary statistics are tabulated in Web Table 1. Proposed changes to the parameters are made through a Gaussian proposal kernel and accepted according to the Metropolis-Hastings algorithm; the kernel was centered on the current value and had variance chosen arbitrarily based on trial runs of the routine, with a view to making the chain mix well. Proposals to infection times were also accepted by dint of Metropolis-Hastings, with a uniform proposal distribution on the permissible range of infection times consistent with the set of sources of infection and the observed data. Proposals to the source of infection were made by a Gibbs step. A random scan algorithm was used to select parameters and hosts to propose changes to at a ratio of 120 parameter proposals to 10 000 time and source proposals.

Convergence of the chain was assessed by inspection of parameter and log-posterior trace plots, and by running a further chain with different initial conditions for ($\boldsymbol{\theta}$, $\boldsymbol{t}$, $\boldsymbol{\sigma}$). 25 000 iterations of the routine were performed and stored, each comprising 120 parameter and 10 000 time and source proposals. A further 1000 such iterations were discarded as burn-in. Every 1000 iterations took around an hour to perform on a modern desktop computer.

We infer the distribution of relevant functions of the parameters, such as $\beta[h_i,h_s](t)$, from the posterior sample. Posterior distributions of sources of infection were similarly obtained to find average proportions of infection by transmission pathway.

The sample was also used to generate a posterior predictive distribution of the stochastic process through Monte Carlo simulation to explore hypothetical scenarios such as the effects of varying heterogeneity. Additionally, the predictive distribution of the daily increment in infection conditional on the existing (observed) pattern of disease was used to assess the goodness-of-fit of the model.

Model selection

The deviance information criterion (DIC) was used to select the best model from a set of competing

models with epidemiologically-plausible time-varying functional forms for primary and secondary infection; it was also used to test for differences in secondary infection rates within and between species. The definition of the DIC used here is given by (6) (DIC_8), and is obtained by first sampling the posterior before using the parameter means to estimate the augmented variables and hence the effective number of parameters of the model. A rule of thumb is to give consideration to models with a DIC within 2 of the model with the lowest DIC score (7); models with a DIC between 2 and 10 of the best have less support, and those with a difference greater than 10 of the best have essentially no support.

Further tests for differences in parameters were carried out using Bayesian hypothesis probabilities, by estimating Pr(hypothesis is true | data). The continuous independent priors used mean that, automatically, the posterior probability of parameters being equal is zero. We therefore consider the posterior probability that one parameter is *greater than* another: should this probability be close to zero or one it indicates there be a difference; should it not be near these extrema, we have insufficient evidence to declare which is greater. Note the difference in interpretation with classical hypothesis testing. As with classical hypothesis testing, the decision of what constitutes a "significant" result is arbitrary.

Analysis of environmental heterogeneity

Using the posterior distributions for the parameters we analyze the effects of heterogeneity in the population structure at two spatial scales on the epidemic dynamics with and without spread between species. Large-scale heterogeneity is investigated by changing the proportion of favorable and unfavorable sites in the population, with random allocation of individuals throughout the population. Small-scale heterogeneity is characterized by a clustering parameter, K, defined as the proportion of neighboring pairs of hosts of the same type; this is varied by increasing the size of regular homogeneous clusters, each randomly allocated to host type such that a 50%:50% mixture is maintained. In all cases, 10 000 simulations of each scenario and spatial arrangement are used to generate the posterior predictive distribution of final disease levels, at an arbitrarily defined time, t = 20, after which further spread of the epidemic was negligible due to the decline in the transmission rates.

Robustness of method

We assessed the ability of the approach to detect temporal-changes to transmission rates for less frequently sampled systems by reducing the temporal detail of the data. This was done by discarding observations, from thirteen at daily intervals to three at times 4, 8 and 12. The effect is that the bounds on the infection time of each plant are broadened; both model and method remain the same. Posteriors are obtained and the temporal evolution of the primary and secondary plotted in Fig. 2 in the main paper.

This choice of observation times was made arbitrarily. It has been shown elsewhere how a small number of observation times can be chosen optimally for non-spatial, homogeneous systems, resulting in a substantial reduction in sampling costs with little loss of observation relative to intensively-sampled populations (8). There is considerable scope and potential benefits for further work in this field.

References in supplementary material

1. Gibson GJ, Otten W, Filipe JAN, Cook A, Marion G & Gilligan CA (2006) *Statistics & Computing* 16:391–402.
2. Cox DR & Isham V (1980) *Point Processes* (Chapman & Hall, London).

3. Gilks WR, Richardson S & Spiegelhalter DJ (eds) (1996) *Markov chain Monte Carlo in Practice: Interdisciplinary Statistics* (Chapman & Hall, London).
4. Gibson GJ & Renshaw E (2001) *Statistics & Computing* 11:347–358.
5. O'Neill PD & Roberts GO (1999) *J R Stat Soc A* 162:121–129.
6. Celeux G, Forbes F, Robert CP & Titterington DM (2006) *Bayesian Analysis* 1:651–674.
7. Spiegelhalter DJ, Best NG, Carlin BR & van der Linde A (2002) *J R Stat Soc B* 64:583–616.
8. Cook AR, Gibson GJ & Gilligan CA (2007) *Biometrics* to appear.

Appendix 2: Parameter estimates

Web Table 1: Posterior means and 95% credible intervals for the model parameters, fitted to 19 replicates of populations comprising different ratios of hosts favorable (*F*) and unfavorable (*U*) for damping-off infection caused by the soil-borne fungal plant pathogen, *R. solani*. The time dependency of the transmission rates is given by a 2-parameter exponential for primary infection, and a 3-parameter Weibull function for secondary infection.

Parameter		Posterior	
		mean	95% CI
Primary transmission	$a[F]$	0.16	(0.11 , 0.17)
	$a[U]$	0.26	(0.20 , 0.32)
	$r[F]$	0.06	(0.01 , 0.12)
	$r[U]$	0.16	(0.11 , 0.22)
Secondary transmission	$b[FF]$	1.4	(1.3 , 1.4)
	$b[FU]$	0.5	(0.4 , 0.6)
	$b[UF]$	1.2	(0.9 , 1.6)
	$b[UU]$	0.4	(0.4 , 0.6)
	$\kappa[FF]$	3.5	(3.3 , 3.8)
	$\kappa[FU]$	3.8	(2.9 , 4.6)
	$\kappa[UF]$	2.9	(2.1 , 3.6)
	$\kappa[UU]$	2.4	(1.7 , 2.9)
	$\mu[FF]$	9.3	(9.1 , 9.5)
	$\mu[FU]$	9.5	(9.0 ,10.0)
	$\mu[UF]$	10.2	(9.0 ,13.1)
	$\mu[UU]$	9.4	(8.3 ,12.1)
Tertiary transmission	$\varepsilon[F]$	0.012	(0.011, 0.014)
	$\varepsilon[U]$	0.004	(0.003, 0.005)