



Large-Scale Spatial-Transmission Models of Infectious Disease

Steven Riley, *et al.*
Science **316**, 1298 (2007);
DOI: 10.1126/science.1134695

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 1, 2007):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/316/5829/1298>

This article **cites 34 articles**, 13 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/316/5829/1298#otherarticles>

This article appears in the following **subject collections**:

Epidemiology

<http://www.sciencemag.org/cgi/collection/epidemiology>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

Large-Scale Spatial-Transmission Models of Infectious Disease

Steven Riley

During transmission of seasonal endemic diseases such as measles and influenza, spatial waves of infection have been observed between large distant populations. Also, during the initial stages of an outbreak of a new or reemerging pathogen, disease incidence tends to occur in spatial clusters, which makes containment possible if you can predict the subsequent spread of disease. Spatial models are being used with increasing frequency to help characterize these large-scale patterns and to evaluate the impact of interventions. Here, I review several recent studies on four diseases that show the benefits of different methodologies: measles (patch models), foot-and-mouth disease (distance-transmission models), pandemic influenza (multigroup models), and smallpox (network models). This review highlights the importance of the household in spatial studies of human diseases, such as smallpox and influenza. It also demonstrates the need to develop a simple model of household demographics, so that these large-scale models can be extended to the investigation of long-time scale human pathogens, such as tuberculosis and HIV.

Outbreaks of directly transmitted infectious diseases of humans have the potential to become pandemics, causing extensive morbidity and mortality (1, 2). Even when widespread transmission is averted, a multicountry outbreak can have a disproportionately large negative economic impact at a regional level (3). Similarly, in domestic animals, the need to maintain disease-free geographical areas requires extreme measures to control economically important pathogens (4). It would be unusual for naturally occurring disease emergence or nondeliberate disease importation to occur simultaneously at many locations. Therefore, infectious individuals are necessarily clustered in space during the initial phase of sustained transmission. If containment is possible, this initial clustering provides an opportunity to make effective use of limited resources for intervention.

Many types of host heterogeneity influence host/pathogen interactions at the scale of the individual, such as genetics, age, sexual activity, location, and typical movement patterns. Sometimes, these factors are important at larger scales and sometimes they are not (5). Consider heterogeneities in the location and movement of hosts during outbreaks of directly transmitted infectious diseases of humans. The vast majority of hosts spend a substantial proportion of their time at a single location; that is, at home or in the herd. When cases arise, their location is often reported rapidly. Furthermore, for many populations, accurate census data with which to estimate spatial population distributions are available. For humans, an accurate estimate of population density is available for the entire

Earth, up to a resolution of 1 arc sec (6). Typical movement patterns of hosts are also important. For directly transmitted pathogens, an infectious individual must be close to a susceptible individual for infection to occur. Therefore, the probability that an infectious individual from one home location infects a specific susceptible individual from another is influenced greatly by the journeys made away from home by both infectious and susceptible individuals during the period of infectiousness. For many populations, data sets from which typical movement patterns can be accurately inferred are now available (7–10).

Spatial models of infectious disease transmission provide the only plausible experimental system in which knowledge of the location of hosts and their typical movement patterns can be combined with a quantitative description of the infection process and disease natural history to investigate observed patterns and to evaluate alternative intervention options. As such, the use of these models will increase as spatially heterogeneous interventions are considered more frequently and as spatially resolved incidence data are made available for more pathogens (the best currently available data sets cover notifiable childhood diseases in developed countries). However, one problem is that these models can be seductive to policy-makers. Realistic population densities permit results to be presented as maps and movies, and although these formats are useful to describe spatiotemporal incidence patterns, their visual impact conveys credibility that may not be justified. Rather, the underlying structure of each transmission model should be appropriate for the infection process, the potential interventions, and most important, the specific hypotheses that are under consideration.

In this review, I highlight results from different diseases (Fig. 1) that use different

methodologies (Fig. 2). I did not include more analytically sophisticated approaches (11), because they have only rarely been applied to the dynamics of infectious disease in realistically structured populations of animals (4, 12).

Measles and Patch Models

Measles is a disease that requires government notification in England and Wales, for which district-level data are available for each bi-weekly period from 1948 onward. Two intriguing epidemiological features of this rich data set have been explained, with the help of patch models (or spatial metapopulation models) (Fig. 2): (i) the longer-than-expected period of fade-out after the start of mass vaccination in 1968 (13, 14) and (ii) the apparent waves of infection observed before and after vaccination (15). In these studies, large amounts of detailed data and sophisticated descriptive statistics (e.g., wavelets) were used in conjunction with very simple illustrative patch models to demonstrate the plausibility of dynamical explanations. For example, the spatial hierarchy of transmission observed in (15) was illustrated with the use of 16 patches of two types (urban or semiurban) arranged in a line. A refined version of this methodology has been used to show that state-scale spatial patterns of excess human mortality attributed to pneumonia and influenza in the United States are consistent with human travel patterns (10). This latter study highlights the computational efficiency of patch models over individual-based approaches; that is, it was possible to include the transmission model within the inferential framework because solutions could be obtained so efficiently.

Patch models have also been used to investigate aspects of global disease spread (8, 9, 16), largely motivated by the 2002–2003 outbreak of severe acute respiratory syndrome (SARS). The availability of disease incidence data at the national level and the accuracy with which contemporary global travel is described by airline-ticket data (17) make this topic well suited to the use of computationally efficient patch models. For example, the observed size of country outbreaks of SARS during 2003 is largely consistent with model outcomes (9), although mainland China was not included, and temporal variation in the underlying rate of transmission was assumed to be the same in every country as that observed in Hong Kong and Singapore.

Similar studies have investigated the potential impact of reduced travel on the rate of international disease spread during an influenza pandemic (8, 16). A theoretical model of 100 identical populations suggests that even the most stringent border controls (>99% effective) would delay global spread of a novel influenza strain by only a few weeks (16). This result is supported by a study of 52 globally connected cities (8), which was validated using travel and

Department of Community Medicine and School of Public Health, Faculty of Medicine, University of Hong Kong, Hong Kong, Special Administrative Region, People's Republic of China. E-mail: steven.riley@hku.hk

disease data from the 1968–1969 pandemic (18, 19).

Foot-and-Mouth Disease and Distance Transmission

The 2001 foot-and-mouth disease (FMD) outbreak in the United Kingdom resulted in the slaughter of 4.2 million animals and produced a severe negative economic impact in the affected districts (20). During the outbreak and afterward, various stakeholder groups disagreed over the relative merits of spatial culling and vaccination strategies. Largely because updated case data were made available to researchers early in the course of the outbreak, statistical and mathematical models were developed rapidly enough to be of use in policy formulation while the outbreak was underway (4, 12, 21, 22).

The availability of spatially resolved farm census data, with unique identifiers that were consistent with the case database, encouraged one research team to formulate a spatially explicit distance-transmission model (Fig. 2) of FMD, with farms as the individual units of infection (21). The infection kernel for FMD in the United Kingdom in 2001 was estimated directly from case data (4, 21) and used in a distance-transmission model (Fig. 2) to assess the marginal benefits of refinements to culling strategies, such as reducing the average delay from report to culling of index premises (IPs) and additional presumptive culling of contiguous premises (21). Initial results using this spatially explicit approach helped to validate those from an earlier pair-correlation model (4, 12), in which properties of an implicit spatial network were approximated with ordinary differential

equations. This series of studies (4, 12, 21, 22), conducted while the outbreak was in progress, demonstrates the potential benefits to policy-makers of multiple research teams addressing similar questions using alternative methodologies.

The precise location of individual farms in distance-transmission models of FMD facilitated the investigation of highly structured spatial interventions that could not be considered in pair-correlation models; for instance, vaccination of a barrier region to protect large pools of susceptible farms (21), prophylactic vaccination of high-risk farms (23), reactive ring vaccination (23), predictive vaccination of nearby farms thought to be most susceptible to infection (23), and vaccination of an annulus around infectious farms (24). However, the logistical implications of constantly revising a large vaccination queue would be substantial. Therefore, a more straightforward strategy was specified: to prioritize farms for vaccination based only on their proximity to IPs reported in the previous 10 days or to dangerous contacts of those IPs. This latter strategy was shown to be the most effective in terms of reducing the expected number of animals that would be culled (24).

Pandemic Containment and Groups

In the event of sustained human-to-human transmission of a novel strain of influenza, public health interventions should ideally attempt to contain the outbreak at the source. The transmission of influenza between individual human hosts occurs over much smaller distances than does the transmission of FMD between farms. Therefore, spatial multigroup models have an advantage over pure distance-transmission-models

when used to investigate influenza containment because they explicitly represent the actual locations (such as households, schools, and workplaces) in which interventions will be used to reduce transmission (Fig. 2). The household is particularly important because relatively static groups of hosts (families) spend prolonged periods together in this setting.

The basic reproductive number R_0 is used to quantify the transmissibility of an infectious disease and is defined as the average number of secondary cases generated by one typically infectious individual in an otherwise susceptible population (25). Two studies of influenza based on models that include a multigroup component suggest that the spatially targeted use of antivirals in rural Asia, in addition to other control measures, has a greater than 90% probability of containing pandemics with basic reproductive numbers $R_0 < 1.9$ (vaccination not included) (26) or $R_0 < 2.4$ (partial vaccination coverage included) (27). In addition to recent estimates of R_0 for influenza that fall in this range (26, 28), these findings led to the adoption of containment as an objective of the World Health Organization (29).

The results from these independent studies are difficult to compare, owing to the many interventions considered and to the principal objective of both, which was to demonstrate feasibility, rather than optimality. However, the two models simulated quite different spatial infection processes, as reflected by their assumed force of infection (FOI) (the hazard of infection experienced by susceptible individuals). In (26), although most individuals belonged to a household and to either a school or a workplace, only two-

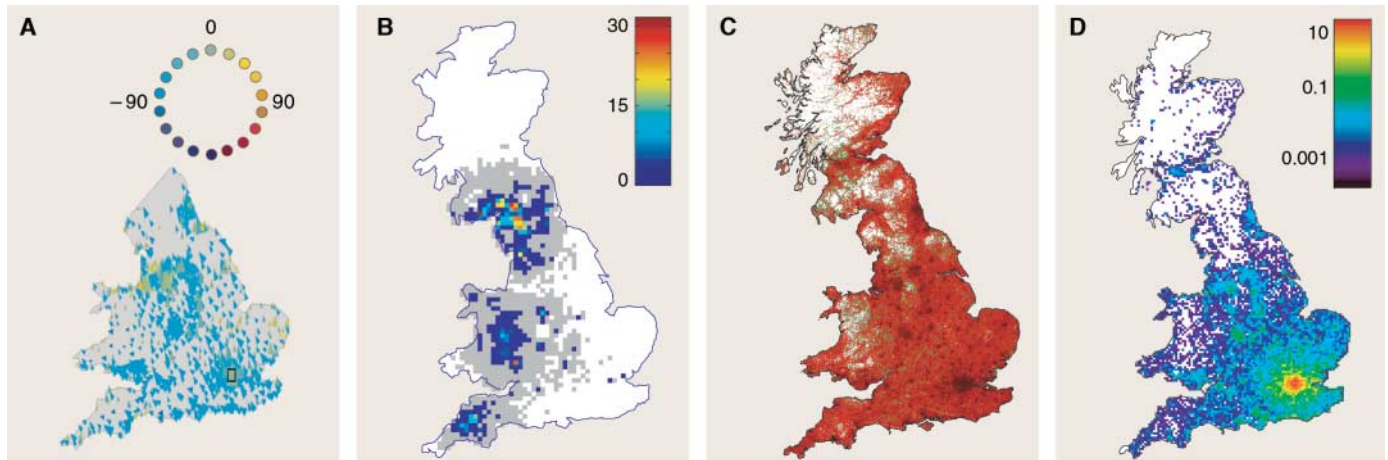


Fig. 1. Patterns of disease transmission in the United Kingdom. (A) Wavelet analysis of prevaccination measles epidemics in 954 locations in England and Wales [reproduced with permission from (15)] shows how London (black box) drove the epidemics in most of the country, with the exception of the Manchester-Liverpool urban concentration (upper left). (B) Simulations show the three high-risk areas [adapted with permission from (21)] for FMD in the United Kingdom in 2001. The key indicates the average number of cases in a 10-by-10-km square from 100 model realizations. (C) A novel strain of influenza [reproduced with permission from (34)] would spread rapidly

though the United Kingdom during a global pandemic. Only 75 days after the arrival of the first cases from overseas, the intensity of red color shows the relative concentration of infectious individuals, and green indicates that the epidemic is already over in some small communities. (D) In contrast, even under a pessimistic transmission scenario, 75 days after 10 initial seeds became infectious with smallpox in London [adapted with permission from (31)], there would have been relatively few cases, and the degree of spatial correlation would still be striking [same key as in (B) but with 5-by-5-km squares].

thirds of the disease transmission occurred in these settings. The remaining third occurred randomly, as a function of the distance from infectious individuals. In contrast, individuals in the model used in (27) belonged to multiple groups, in addition to households, schools, and workplaces. The inclusion of these extra settings was an alternative to the distance-transmission component used in (26). Although both approaches require parameters for which there is little or no supporting empirical data, the model used in (26) could be viewed as more parsimonious because its conclusions are less sensitive to these uninformed parameters, and it made good use of the data that were available, such as spatially resolved population density and travel survey data.

Model design choices such as these can have a substantial impact on predictions of intervention efficacy. For example, in (27), targeted antiviral prophylaxis implicitly assumes that household clusters and small neighborhoods that are important for transmission could be easily identified, resulting in a relatively low upper bound of 1 million courses of treatment required for containment. Conversely, the spatial recruitment of households into the intervention processes proposed in (26) implicitly assumed a weaker correlation between antiviral distribution and transmission, resulting in the more conserv-

ative estimate that 3 million courses of treatment would be required to achieve containment.

Smallpox and Networks

Patch, distance kernel, and multigroup models can all be considered as special cases of spatial-network models, in which nodes represent individual hosts and arcs represent potentially infectious links. As networks, all of the model/pathogen combinations described above would have high average numbers of arcs per node (i.e., a large neighborhood size) relative to their basic reproductive number R_0 . However, for smallpox, intimate contact was almost always required for transmission to occur (30). Therefore, it is necessary to represent a substantial proportion of smallpox transmission as occurring over a static network with a relatively small neighborhood size. This approach was used (31) to show that for the United Kingdom, the additional benefits of geographically targeted regional vaccination would not outweigh the adverse effects of vaccinating many low-risk individuals. Specifically, contact tracing with isolation and vaccination alone would probably result in fewer deaths from a small initial cluster of cases in London than would occur if geographically targeted regional vaccination was used in addition to such a policy.

Perhaps the most innovative modeling approach to emerge from smallpox epidemiology (32) is the derivation of static contact networks from individual-based second-by-second microsimulation. In (33), all of the people, locations, and journeys in the city of Portland, Oregon, were simulated explicitly. Simple rules were then used to construct static contact networks from dynamic networks of individuals and locations. For example, if two individuals were present in the same location for more than an hour, it was assumed that a social contact existed between the two. Because the intensity of contact was assumed to be similar in all locations, implicitly, people's behavior in supermarkets (with respect to disease transmission) was assumed to be the same as in the home. This uniformity of contact intensity is unrealistic and must have resulted in overly connected social networks. However, when good data are available on the relative transmissibility of respiratory pathogens in different social settings, the derivation of large spatial contact networks from microsimulations will provide a natural refinement of the distance-transmission approach described above.

Current Challenges

Some of the individual-based spatial models described above include age classes and house-

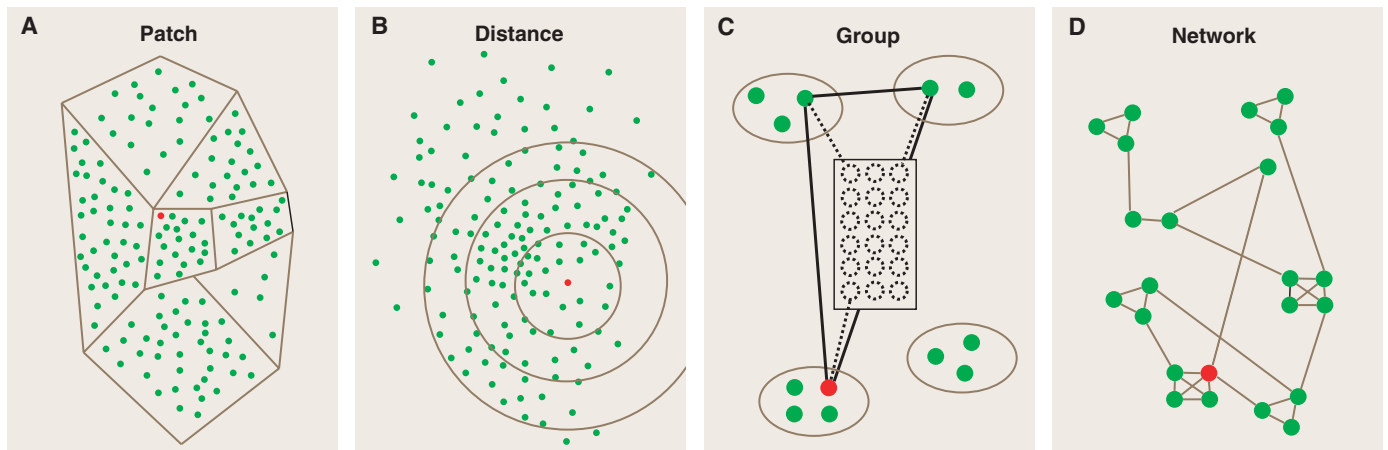


Fig. 2. Four common abstractions for the spatial transmission of infectious diseases. Differences between these approaches are best understood in terms of the FOI, which is location-specific in spatially explicit models. Red dots represent infectious individuals. (A) For patch transmission, all members of the same patch (residents of a town, for example) receive the same FOI, which is a function of the distance from their home patch to other patches and of the prevalence of infection in all patches. (B) Distance transmission is explicitly individual-based; that is, each farm is assigned a precise location. It is assumed that any given infectious individual can infect all susceptible individuals within range. The pairwise probability of infection is usually a monotonically decreasing function of distance, and the absolute FOI experienced by each susceptible individual because of a single infectious individual is low. (C) In a pure multigroup model, the FOI is determined entirely by group membership. For example, if an infectious individual shares a household with a susceptible individual (ovals), there is a high probability of transmission occurring between the two. However, if an infectious individual does not share a group with a particular susceptible individual, transmission

between the two cannot occur. Spatial patterns of spread are determined by the locations of households and workplaces/schools (rectangle) and by the typical distribution of journeys between them. Dashed lines indicate group membership and solid lines indicate potentially infectious links between individuals. (D) Network transmission is similar to group transmission in that the FOI experienced by susceptible individuals is zero, unless they share an arc with an infectious individual. For directly transmitted respiratory pathogens, network transmission can be thought of as a refinement of an implicit group structure, in which it is assumed that not all members of a group are equally well connected; e.g., all colleagues at a workplace are not contacts. More than one component of transmission is included in some models. In general, computational requirements increase from (A) to (D). Patch models can be implemented effectively on a typical desktop computer because they do not explicitly represent individuals. For population sizes greater than 10 million, individual-based models have been implemented on clusters of large-memory personal computers (26, 31, 34). Detailed microsimulation models (33) have not yet been implemented at scales larger than a city.

hold structure (26, 33–35). However, none of them include demographic processes so that the birth, death, and aging of individuals are consistent with the formation and dissolution of households over time. Although these processes would introduce substantial additional complexity, spatially resolved census data and projections are available for many populations, which would allow the small number of parameters required by the additional demographic processes to be estimated independently of the transmission model. Demographic projections have already been incorporated with good effect into nonspatial models, such as those for HIV (36) and pertussis (37). I suggest that several human-pathogen interactions that exhibit spatial clustering at large scales and that are often transmitted within the household could be investigated with such an extension to current individual-based spatial models. For example, it may be possible to refine predictions of spatial patterns of combined tuberculosis and HIV transmission in developing countries (38) and hence to consider the use of spatially heterogeneous household-based intervention strategies at different distance scales. The sensitivity of results from such models to the demographic processes that were used could be illustrated clearly with appropriate sensitivity analyses.

If one uses the network paradigm to critique models of directly transmitted infectious diseases, there is a desperate need for empirical evidence to inform basic choices of topology, such as average neighborhood size. For example, the model structures used in many recent studies of pandemic influenza (26, 27, 34, 35, 39) assume implicitly that neighborhood size outside the household is large; i.e., that a substantial proportion of the small number of secondary cases (basic reproductive number $R_0 < 2$, including household transmission) arises with low probability from brief contacts with many people. Although this assumption is appropriate because

it is conservative with respect to the efficacy of group-based interventions, there is very little supporting empirical evidence. For example, I am not aware of any studies that have been able to infer reliable estimates (for any pathogen) of the infectivity and susceptibility of individuals in the home, relative to the workplace or school. Techniques that allow highly specific non-invasive sampling for bacteria and viruses, coupled with intense empirical studies based in households, workplaces, and schools, could dramatically reduce these gaps in our knowledge. Detailed microsimulation models (33) could be used to obtain estimates of spatial-network topologies for different pathogens.

References and Notes

- D. R. Olson, L. Simonsen, P. J. Edelson, S. S. Morse, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11059 (2005).
- T. C. Quinn, *Lancet* **348**, 99 (1996).
- The World Bank, <http://siteresources.worldbank.org/INTEAPHALFYEARLYUPDATE/Resources/EAP-Brief-avian-flu.pdf> (accessed 28 November 2006) (2005).
- N. M. Ferguson, C. A. Donnelly, R. M. Anderson, *Science* **292**, 1155 (2001); published online 12 April 2001 (10.1126/science.1061020).
- S. A. Levin, B. Grenfell, A. Hastings, A. S. Perelson, *Science* **275**, 334 (1997).
- J. E. Dobson, E. A. Bright, P. R. Coleman, R. C. Durfee, B. A. Worley, *Photogramm. Eng. Remote Sens.* **66**, 849 (2000).
- D. Brockmann, L. Hufnagel, T. Geisel, *Nature* **439**, 462 (2006).
- B. S. Cooper, R. J. Pitman, W. J. Edmunds, N. J. Gay, *PLoS Med.* **3**, e212 (2006).
- L. Hufnagel, D. Brockmann, T. Geisel, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15124 (2004).
- C. Viboud *et al.*, *Science* **312**, 447 (2006); published online 29 March 2006 (10.1126/science.1125237).
- S. A. Levin, R. Durrett, *Philos. Trans. R. Soc. London Ser. B* **351**, 1615 (1996).
- N. M. Ferguson, C. A. Donnelly, R. M. Anderson, *Nature* **413**, 542 (2001).
- B. M. Bolker, B. T. Grenfell, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12648 (1996).
- M. J. Keeling, B. T. Grenfell, *Science* **275**, 65 (1997).
- B. T. Grenfell, O. N. Bjornstad, J. Kappey, *Nature* **414**, 716 (2001).
- T. D. Hollingsworth, N. M. Ferguson, R. M. Anderson, *Nat. Med.* **12**, 497 (2006).
- International Airline Travel Association, www.iata.com (accessed 23 November 2006).
- I. M. Longini, P. E. M. Fine, S. B. Thacker, *Am. J. Epidemiol.* **123**, 383 (1986).
- L. A. Rvachev, I. M. Longini, *Math. Biosci.* **75**, 3 (1985).
- Department for Environment Food and Rural Affairs (UK), www.defra.gov.uk/footandmouth (accessed 23 November 2006).
- M. J. Keeling *et al.*, *Science* **294**, 813 (2001); published online 3 October 2001 (10.1126/science.1065973).
- M. Woolhouse *et al.*, *Nature* **411**, 258 (2001).
- M. J. Keeling, M. E. Woolhouse, R. M. May, G. Davies, B. T. Grenfell, *Nature* **421**, 136 (2003).
- M. J. Tildesley *et al.*, *Nature* **440**, 83 (2006).
- J. A. P. Heesterbeek, *Acta Biotheor.* **50**, 189 (2002).
- N. M. Ferguson *et al.*, *Nature* **437**, 209 (2005).
- I. M. Longini Jr. *et al.*, *Science* **309**, 1083 (2005); published online 3 August 2005 (10.1126/science.1115717).
- C. E. Mills, J. M. Robins, M. Lipsitch, *Nature* **432**, 904 (2004).
- World Health Organization, www.who.int/csr/resources/publications/influenza/WHO_CDS_CSR_GIP_2005_5/en/index.html (accessed 28 November 2006) (2005).
- F. Fenner, D. A. Henderson, L. Arita, Z. Jezek, I. D. Ladnyi, *Smallpox and Its Eradication* (World Health Organization, Geneva, Switzerland, 1988).
- S. Riley, N. M. Ferguson, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12637 (2006).
- N. M. Ferguson *et al.*, *Nature* **425**, 681 (2003).
- S. Eubank *et al.*, *Nature* **429**, 180 (2004).
- N. M. Ferguson *et al.*, *Nature* **442**, 448 (2006).
- T. C. Germann, K. Kadau, I. M. Longini Jr., C. A. Macken, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5935 (2006).
- R. M. Anderson, R. M. May, M. C. Boily, G. P. Garnett, J. T. Rowley, *Nature* **352**, 581 (1991).
- H. W. Hethcote, *Math. Biosci.* **145**, 89 (1997).
- B. G. Williams, R. Granich, L. S. Chauhan, N. S. Dharmshaktu, C. Dye, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9619 (2005).
- J. T. Wu, S. Riley, C. Fraser, G. M. Leung, *PLoS Med.* **3**, e361 (2006).
- I thank the Research Fund for the Control of Infectious Diseases of the Health, Welfare and Food Bureau of the Hong Kong Special Administrative Region Government for funding and three anonymous reviewers, R. Anderson, B. Cowling, N. Ferguson, C. Fraser, B. Grenfell, M. Heys, G. Leung, P. Riley, and J. Wu for comments on the manuscript.

10.1126/science.1134695