Towards Real Time Epidemiology: Data Assimilation, Modeling and Anomaly Detection of Health Surveillance Data Streams

Luís M.A. Bettencourt[1], Ruy M. Ribeiro[1], Gerardo Chowell[1], Timothy Lant[2], and Carlos Castillo-Chavez[3]

[1] Theoretical Division and Center for Non-Linear Studies,
Los Alamos National Laboratory,
MS B284, Los Alamos NM 87545, USA
[2] Decision Theater, Arizona State University,
PO Box 878409, Tempe AZ, 85287-8409, USA
[3] Department of Mathematics and Statistics, Arizona State University,
PO Box 871804, Tempe AZ, 85287-1804, USA

Abstract. An integrated quantitative approach to data assimilation, prediction and anomaly detection over real-time public health surveillance data streams is introduced. The importance of creating dynamical probabilistic models of disease dynamics capable of predicting future new cases from past and present disease incidence data is emphasized. Methods for real-time data assimilation, which rely on probabilistic formulations and on Bayes' theorem to translate between probability densities for new cases and for model parameters are developed. This formulation creates future outlook with quantified uncertainty, and leads to natural anomaly detection schemes that quantify and detect disease evolution or population structure changes. Finally, the implementation of these methods and accompanying intervention tools in real time public health situations is realized through their embedding in state of the art information technology and interactive visualization environments.

Keywords: real time epidemiology, data assimilation, Bayesian inference, anomaly detection, interactive visualization, surveillance.

1 Introduction and Motivation

Surveillance systems collect, analyze and report data continuously. Recent progress in information science and technology is increasingly enabling the collection of public health data worldwide in near real-time. In the United States, partly motivated by bio-security concerns, real-time bio-surveillance systems, that follow direct and indirect indicators of (epidemic) public health outbreaks have been developed nationwide, at the city level e.g. for New York City or Los Angeles, and at the state level such as in Michigan, Utah or Ohio. The monitoring of indirect data streams, pertaining e.g. to work or school absenteeism, emergency room and physician office calls, sales of certain over-the-counter medicines, known as syndromic surveillance [1,2,3,4] is also underway.

The use of a large number of real time data streams to infer the status and dynamics of the public health of a population presents enormous opportunities as well as significant scientific and technological challenges. A recent committee report from the Institute of Medicine of the National Academies concluded: "... that steps [must] be taken now to adapt or develop decision-aid models that can be readily linked to surveillance data to provide real-time feedback during an epidemic" [17]. Among the "Grand Challenges in Global Health" (http://www.gcgh.org/), articulated by a consortium of international organizations such as the Bill and Melinda Gates Foundation, The Wellcome Trust and the Canadian Institutes of Health Research, one finds the development of "... technologies that permit quantitative assessment of population health status". Naturally, the early identification and detection of emerging pathogens, pandemics or bioterrorist attacks call for the development and deployment of real-time syndromic surveillance systems.

Real time public health surveillance has as its primary mission pre-emptying the successful invasion or establishment of emerging infectious diseases. Its goals include the prevention or mitigation of epidemic growth, if possible by setting quantitative targets for intervention as events unfold and by providing a running time expectation for logistic allocations, including medicines, hospital beds and vaccines, if available. The success of real time public health surveillance depends on our ability to develop not only effective detection systems but also ways of evaluating the uncertainties associated with outbreaks, methods and models.

To a great degree, the reliability of a surveillance system hinges on its ability to make the extrapolations needed to predict the likely course of a public health emergency given the incoming streaming information. Hence, the need to develop statistical models of data assimilation that effectively estimate the changing parameters that characterize models of disease *dynamics*. In this paper a general procedure for data assimilation from real time data streams that feeds into epidemic models of communicable disease spread is developed. This procedure allows i) the estimation of probability densities for epidemiological parameters (such as disease transmissibility) and ii) the *prediction* of future observables with a quantified degree of confidence. This dynamical estimation/extrapolation environment lends itself naturally to a statistical anomaly detection scheme capable of identifying shifts in the public health status of a population, either via pathogen evolution, or as the result of changes in the population structure.

2 Data Assimilation and Probabilistic Prediction

Real time estimation of epidemiological parameters is conditioned by the types of data available as well as the target quantities for estimation (Figure 1). It is essential to produce running time statistics that incorporate past and present measurements as well as ways to assimilate these data to generate probabilistic predictions of new cases with quantified uncertainty because often the objective of real time surveillance is to detect change. This approach also allows the use of statistical estimation to partially bypass issues of non-stationarity in the data,
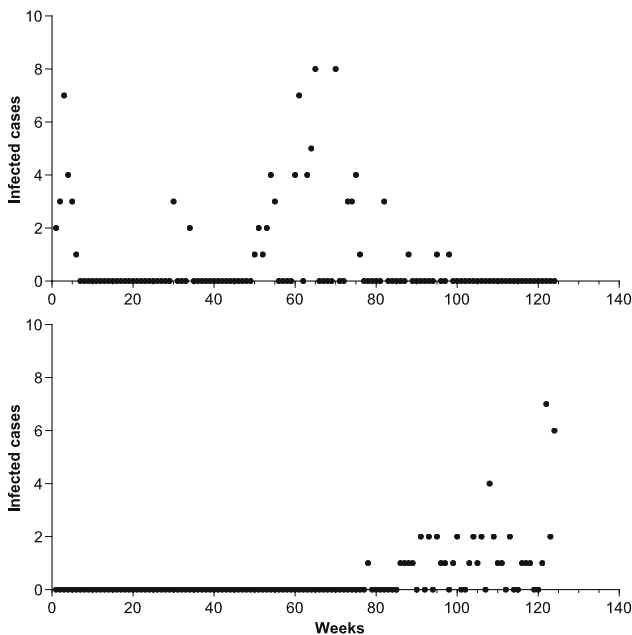
Fig. 1. Number of cases of H5N1 human influenza. Cases confirmed by the World Health Organization for Vietnam (top) and Indonesia (bottom), from January 2004 to June 2006. Although these are currently the two countries with most recorded cases, the incidence times series is very stochastic and intermittent (note the y-scale).

which typically hinder purely statistical (stationary) approaches to syndromic surveillance. Technically, this can be achieved by adopting a probabilistic approach to the *prediction* of new cases (or their proxy variables) at a given time t, $\Delta C(t)$, given past incidence time series and a model for the disease dynamics written in terms of a set of parameters Γ. The set Γ includes familiar parameters such as measures of transmissibility (e.g. the reproduction number R) and the duration of the infectious and incubation periods.

Here we focus on the estimation of a key quantity in epidemiology known as the reproduction number or ratio which quantifies the transmissibility of infectious diseases under identifiable initial conditions. In the context of novel infectious pathogens (e.g., the 1918-19 influenza pandemic, or the 2002-03 SARS epidemic) to which most of the population is completely susceptible, the *basic* reproduction number (denoted by R_0) quantifies the number of secondary cases generated by a primary case during its period of infectiousness within a *purely* susceptible population. In more practical situations, the population's actual susceptibility is affected by recurrent exposures to the infectious agent (e.g., dengue fever, influenza), vaccination campaigns (e.g., influenza, measles, mumps) and by "depletion" of susceptibles during the course of an outbreak. Then, the reproduction number denoted by R accounts for the residual immunity in the population. The relationship between R_0 and R *may* be modeled by $R = (1 - p)R_0$ where p

denotes the proportion of the population that is effectively protected to infection, assuming a well-mixed population [5,7,18,19,20]. In the advent of an epidemic, the timely estimation of the reproduction number with the minimal amount of available data crucial as it would allow public health authorities to determine the types and intensity of interventions necessary to achieve *fast* epidemic control. The level of critical vaccination coverage needed to diminish or eliminate various infectious diseases have also been determined from the magnitude of the estimated reproduction number [5,7,18,19,20].

A number of approaches have been used to estimate the reproduction number of infectious diseases including trajectory matching via least square fitting of epidemic models to time series epidemic data (e.g., [9,10,11]), methods that rely on the final epidemic size relation with the reproduction number (e.g., [13,15,6]), methods that make use of serological survey data (e.g., [14,12]), and recent probabilistic approaches that estimate the *effective* reproduction number over time (e.g., [8,21,22,23,24,16]).

The dynamical probabilistic approach described here [8,21,22,23,24,16] allows for the estimation of parameter density distributions via Bayes' theorem. Furthermore, as byproducts, the method leads to the automatic uncertainty quantification of predictions and a scheme for anomaly detection. The probabilistic prediction for new cases ($\Delta C(t)$) is written in terms of the probability distribution of given previous incidences ($\Delta C(t - \tau)$) and dynamical parameters, which we denote by

$$P\left[\Delta C(t) \leftarrow \Delta C(t - \tau)|\Gamma\right] = P\left[\Delta C(t)|\Delta C(t - \tau), \Gamma\right]. \tag{1}$$

This distribution encapsulates the dynamical prescriptions from different dynamical models of (epidemiological) disease dynamics. Below we show how this formulation can be specified in practice for different models, how it implies the full distribution for epidemiological parameters Γ, and how it leads naturally to predictions with quantified uncertainty and to anomaly detection schemes. Given probabilistic information on past cases and/or parameters, the probability for new cases is given as

$$P\left[\Delta C(t)\right] = \int d\Delta C(t - \tau)\ d\Gamma\ \ P\left[\Delta C(t)|\Delta C(t - \tau), \Gamma\right]\ P\left[\Delta C(t - \tau), \Gamma\right]. \tag{2}$$

If parameters or previous cases are known exactly, then the distributions under the integral on the right become δ-functions, enforcing specific (observed) values. This identity gives rise to practical schemes for parameter estimation and new case prediction, as we illustrate below.

The probabilistic prescription of new cases (1) implies in turn the distribution of model parameters Γ. The key to parameter estimation, given a probabilistic disease dynamics model, is the well known Bayes' theorem

$$P\left[\Gamma|\Delta C(t) \leftarrow \Delta C(t - \tau)\right] = \frac{P\left[\Delta C(t) \leftarrow \Delta C(t - \tau)|\Gamma\right]\ P\left[\Gamma\right]}{P\left[\Delta C(t) \leftarrow \Delta C(t - \tau)\right]} \tag{3}$$

where the denominator is a normalization factor.

Thus observation of time series for epidemiological quantities ΔC at consecutive times is equivalent to knowledge of the probability distribution for epidemiological parameters. The latter is given, in analogy to (2) as

$$P[\Gamma] = \int d\Delta C(t-\tau) \, d\Delta C(t) \, P[\Gamma|\Delta C(t), \Delta C(t-\tau)] \, P[\Delta C(t-\tau), \Delta C(t)].$$
(4)

This expression also allows for the inclusion of uncertainties in case counts, which may result e.g. from underreporting or other sampling biases.

3   Probabilistic Disease Models and Real Time Estimation of Epidemiological Parameters

In this section, it is shown how standard "mean-field" compartment models of communicable diseases prescribe $P[\Delta C(t)|\Delta C(t-\tau), \Gamma]$. Alternative models have been proposed to achieve the same goal, including the probabilistic reconstruction of putative chains of transmission [8,22], and stochastic ensemble methods [25]. The approach that we follow here has however the distinct advantage of simplicity. Furthermore, its implementation requires only a very modest computational effort. Other methods present certain advantages, however, such as the estimation of probable chains of transmission [8,21,22].

Mathematical models predicting the time evolution of the average number of infectious cases, deaths, etc. at a given time are among the most useful and most commonly used descriptions of contagion processes. Classical epidemiological models such as SIR (Susceptible-Infectious-Removed) or SEIR (Susceptible-Exposed-Infectious-Removed) are of this form. Each class or compartment counts the mean number of individuals in a specific epidemiological state and may refer additionally to a geographic location, age or risk group.

We have shown elsewhere [23] that in the absence of sources these models imply a relation between new case (or death) numbers at consecutive times, of the form

$$\langle \Delta C(t+\tau) \rangle = b(\Gamma)\Delta C(t),$$
(5)

where the $\langle ... \rangle$ denotes expectation and $b(\Gamma) = \exp[\tau\lambda(\Gamma)]$. For example in the case of the SIR and SEIR models we have

$$\lambda_{SIR} = \gamma(R-1); \qquad \lambda_{SEIR} = \frac{\kappa+\gamma}{2}\left[-1 + \sqrt{1 + \frac{\kappa\gamma}{(\kappa+\delta)^2}(R-1)}\right].$$
(6)

which are the leading positive eigenvalues characterizing the evolution of case numbers. Here κ^{-1} is the latency period, γ^{-1} is the duration of the infectious period, and R is the effective reproduction number, which is defined as the mean number of infected cases caused by an infectious individual. Other models will result in a different form of λ.

These average relations for future new cases can now be used to define a probabilistic model for new cases as

$$\Delta C(t+\tau) \sim P[\Delta C(t+\tau) \leftarrow \Delta C(t)|\Gamma] = P[b(\Gamma)\Delta C(t)],$$
(7)

where $\Delta C(t+\tau)$ is taken as a stochastic variable distributed according to the distribution $P[\langle A \rangle]$, with average $\langle A \rangle$. Note that the "mean field" models do not prescribe $\Delta C(t+\tau)$, only its average. Knowledge of higher correlations, such as the variance in case numbers, can help constrain the form of P among a class of discrete (for cases, deaths) probability density distribution functions. In the absence of additional information a Poisson distribution is the highest entropy (i.e. most general) form. Another common choice is a Negative Binomial, which allows for clumping effects, but requires the specification of a quantity related to the variance of $\Delta C(t+\tau)$.

Depending on knowledge of details of the contagion process, correlations among observables and availability of data different model choices in this class may be appropriate, although an SIR or SEIR forms are often adequate.

There are many circumstances when disease cases, or their proxy variables, do not change purely under contagion dynamics. These additional effects may result from multiple introductions, e.g. from an animal reservoir, as well as from unexpected patterns of human behavior. Introductions create cases in addition to those due to contagion, and may start epidemics if initially $R > 1$. To include these effects, we have developed models [23] that account for infections from a reservoir, in addition to the usual transmission from infectious individuals to others. This results in the modification of (5), due to additive terms,

$$\langle \Delta C(t+\tau) \rangle = \Delta B(t+\tau) + b(\Gamma)\left[\Delta C(t) - \Delta B(t) + \tau \gamma R \psi(t, \tau, \Gamma)\right], \qquad (8)$$

where $\psi(t, \tau, \Gamma) = \int_t^{t+\tau} dt' \exp\left[-\lambda \times (t' - t)\right] f_c \Delta B(t')$. Here $\Delta B(t)$ are the new cases due to introductions from the reservoir at time t, and f_c is the probability that such cases are transmissible between humans.

This expression shows that inference of disease transmissibility between humans requires statistical knowledge on the number of introductions. This can
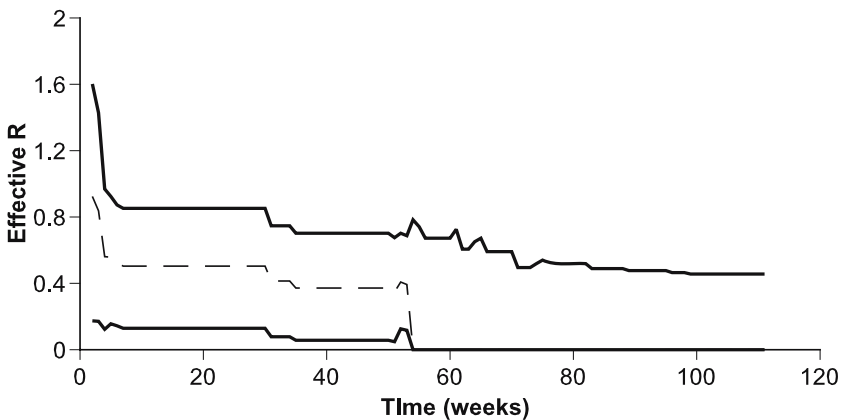

Fig. 2. Predicted effective reproduction number for H5N1 influenza in humans in Vietnam. This is an example of the application of the method described in the text, for details see [23].

usually be inferred from qualitative information and/or can be parameterized to create scenarios, under varying proportions of observable cases being attributable to contagion vs. introductions. We have recently performed such analysis to place bounds on the current transmissibility of H5N1 avian influenza, from data of the outbreaks in Southeast Asia (e.g. see Figure 2)[23].

4   Anomaly Detection

Anomaly detection is not a simple business because often alternative evolutionary changes may lead to similar outcomes. For example, in the case of communicable diseases these may lead to changes in susceptibility or infectiousness and both "anomalies" often result in comparable outcomes at the population level. Nevertheless, it is desirable to determine when the statistical predictions of epidemiological models fail to describe future data automatically. Inability to predict correctly may signal disease evolution (potentially increasing or decreasing transmissibility or lethality), or changes in the population structure due to population movements and/or unforeseen human behavior.

The probabilistic prediction of new cases (or deaths, etc) that we described above also lends itself to natural and simple schemes for anomaly detection. One approach consists in extracting a prediction for the interval of future cases (or deaths) at a chosen level of confidence (Figure 3). This prescription can be formalized in terms of a two-sided p-value significance test, where $p = \alpha/2$ is the probability that corresponds to a $1 - \alpha$ level of confidence in the predictions of the model. The model is rejected, and an anomaly is flagged, if case numbers are
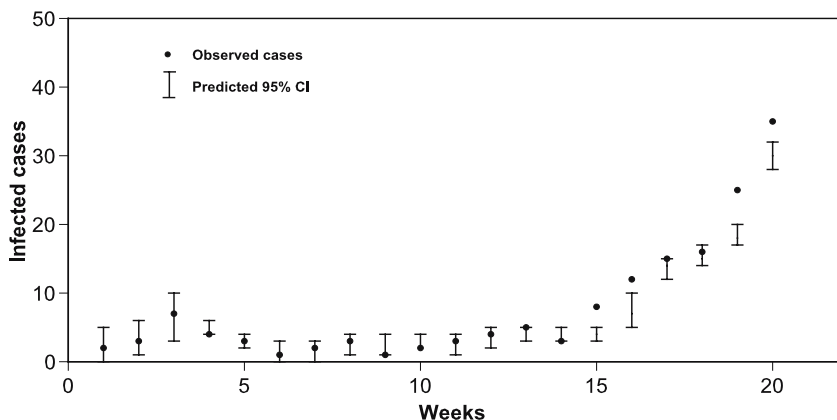

Fig. 3. Schematic illustration of the implementation of anomaly detection. The Bayesian method described in section 3 allows the prediction of the number of cases, with given confidence bounds. Whenever the number of observed cases falls outside that prediction an anomaly is detected and the need for further epidemiological investigation is flagged. Here such events start happening from week 15.

smaller than those corresponding to probability p, or are larger than those corresponding to $1 - p$. The significance α is typically chosen to be $\alpha = 0.05$, which corresponds to (dis)agreement with model predictions at 95% confidence level.

The automatic detection of anomalies should prompt investigations potentially for faults in surveillance, or disease evolution or changes in the population structure. If several measurable quantities are available for prediction and anomaly detection, several statistical tests can be combined into a collective test. The implementation of this aggregated test can take different forms depending on the independency and judged relevance of the individual tests and is the subject of meta-analysis [26]. The changes in disease dynamics (anomalies), may result from co-evolutionary processes, that is, from the interplay between pathogen evolution and changing social landscapes and, consequently, while their detection may be possible, the identification of the underlying mechanisms may not be straightforward without further investigation.

5   Visualization, Interventions and Table Top Exercises

We are currently using the methods described above to create a real-time public health situation awareness and intervention environment. Our objective is to embed real-time statistical inference methods and modeling algorithms in a state-of-the-art visualization and computing environment at the Decision Theater (DT) of Arizona State University. The technology under development will integrate two types of public health applications in a unified user environment: the first is the real-time analysis of geographically tagged public health data; the second is a platform to simulate disease outbreaks for public health planning. As a byproduct, we are also developing a platform for research and visual analysis of spatially-explicit disease transmission models.

The platform was prototyped with the Arizona Department of Health Services to study West Nile Virus in Maricopa County, Arizona, USA. The initial implementation was built using open-source software: a postgres/postgis database [28,27] for the storage and retrieval of spatial data, the statistical software application R [29], algorithms from the DCluster package, written in R [30], and a visualization and rendering application dubbed Minerva [31] developed at the Decision Theater to visualize spatial disease data and now being used in other application areas as well. Figure 4 shows an implementation of the system during a working session to study West Nile Virus in the Decision Theater with staff from the Arizona Department of Health Services.

The Minerva application is the front-end of the information system. This program performs four tasks in a common application: (i) it provides ODBC connectivity to postgres/postgis databases to retrieve data to render through standard SQL, (ii) it renders data in the DT environment using OSG, (iii) it allows users to select layers of data, associated with display properties such as the color and shape, and (iv) it has menus for launching algorithms and choosing parameters, but it does not run them itself. The key to coordinating model runs and data transfers lies with the methods used to access the database. Scripts, written in

Fig. 4. Analysis of West Nile Virus in the Decision Theater at ASU. The photo was taken during a working session at the Decision Theater with the Arizona Department of Health Services, Vector Borne and Zoonotic Disease Department staff. The interactive visualization environment shows spatial analysis of West Nile Virus cases during 2004 and 2005 including demographic and environmental data and spatial clustering algorithms.

R, connect to the database and execute an SQL command that imports a frame of current data into R. Additional R commands transform the data to match the required format for the spatial and statistical analysis. The results are uploaded back into the database into tables which can be queried and rendered by Minerva. Each time a layer is selected for visualization, Minerva starts an R script, waits for the database to indicate it has been updated, and renders the data. Minerva stores binary versions of the rendered layers so they can be turned on and off through the user interface without running the algorithms multiple times during a session.

The fully-developed system will allow the calculation of statistical metrics and mathematical forecasts, coupled with a real-time geographic information system. The power and flexibility of the Decision Theater visualization environment provides the computational and informational infrastructure to deliver a real-time epidemiology decision support system that can be used during public health events, for public health planning, and in developing a better understanding of the spatial spread of multiple diseases. In addition, it allows the visualization of multiple interchangeable data sets simultaneously. Thus, this system can be used for several critical features of a real-time public health information system, such as: 1) data acquisition, storage, and retrieval; 2) exploring the descriptive epidemiology

of the disease; 3) determining the major demographic and socio-economic factors that contribute to the spread of the disease, 4) the development of modeling and simulation exercises, and 5) assessing the effects of health interventions in human populations.

The value of using real-time epidemiological decision support environments such as the Decision Theater relies heavily on the access and use by public health officials and on approaches to research and modeling that drive these systems. Current approaches to disease transmission modeling rely too heavily on the analysis of historical data or on the ability to predict future outbreaks, but overlook the needs of public health officials to understand the current unfolding situations in ways that extract meaningful knowledge. The modeling approach in this paper takes a significant step towards the creation of quantitative tools that can deliver real time information to health officials who can best apply the knowledge gained. The visualization of that information in a real-time spatial context is necessary to base real-time decisions that include the current state of disease epidemics as they unfold. Table-top exercises, simulation, and analyses of current infections in the population in cooperation with public health officials also serve to educate health professionals on how to use the information system prior to outbreaks in meaningful ways that will be directly transferable to the monitoring and management of actual health events.

6   Conclusions and Outlook

An integrated quantitative approach to data assimilation, prediction and anomaly detection over real time public health surveillance data streams has been presented. The foundation of the system that we envisage, and its main difference to other current approaches, is the use of epidemiological models as the basis for statistical analyses and temporal prediction. Although these models require more assumptions than purely statistical approaches, they allow for the integration of dynamical variables essential for forecasting and for natural automatic uncertainty quantification and anomaly detection.

The approach developed here will build on the current syndromic surveillance efforts [1,2,3,4], by integrating them with epidemiological predictive modeling, which has a long and successful tradition in support of public health [5,6,7]. While much remains to be done, we hope that these kind of methodologies will enable a shift towards more quantitative active surveillance and primary prevention, resulting in more powerful strategies for monitoring infectious diseases. The integration of these efforts into sustainable solutions that strengthen public health worldwide remains the most important challenge and, simultaneously, the greatest new opportunity to international public health and policy organizations, requiring new levels of transparency, efficiency and cooperation among scientists, governments, the private sector and non-governmental organizations.

References

1. Lawson AB, Kleinman K (eds.). Spatial and Syndromic Surveillance for Public Health. Chichester: John Wiley & Sons, 2005.
2. Framework for evaluating public health surveillance systems for early detection of outbreaks; recommendations from the CDC working group. MMWR CDC Surveill. Summ. 2004; 53: 1-16.
3. Proceedings of the 2002 National Syndromic Surveillance Conference. New York, USA, September 23-24, 2002. J. Urban Health. 2003; 80.
4. Various. Abstracts from the 2005 Syndromic Surveillance Conference. Advances in Disease Surveillance. 2006; 1.
5. Anderson RM, May RM. Infectious Diseases of Humans. Oxford: Oxford university Press, 1991.
6. Brauer F, Castillo-Chavez C. Mathematical Models in Population Biology and Epidemiology. New York: Springer-Verlag, 2001.
7. Diekmann O, Heesterbeek JA. Mathematical Epidemiology of Infectious Diseases: model building, analysis and interpretation. Chichester: John Wiley & Sons, 2000.
8. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am. J. Epidemiol. 2004; 160: 509-516.
9. Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L et al. Transmission dynamics and control of severe acute respiratory syndrome. Science. 2003; 300: 1966-1970.
10. Riley S, Fraser C, Donnelly CA, et al. Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. Science 2003; 300:1961-1966.
11. Chowell G, Ammon CE, Hengartner NW, Hyman JM. Transmission Dynamics of the Great Influenza Pandemic of 1918 in Geneva, Switzerland: Assessing the Effects of Hypothetical Interventions. J. Theor. Biol. 2006; 241:193-204.
12. Ferguson NM, Donnelly CA, and Anderson RM. Transmission dynamics and epidemiology of dengue: insights from age-stratified sero-prevalence surveys. Phil. Trans. Roy. Soc. Lond. B, 354:757–768, 1999.
13. Koopman JS, Prevots DR, Vaca-Marin MA, Gomez-Dantes H, Zarate-Aquino ML, Longini IM Jr, and Sepulveda-Amor J. Determinants and predictors of dengue infection in Mexico. Am. J. Epidem. 1991; 133:1168–1178.
14. Farrington CP, Whitaker HJ. Estimation of effective reproduction numbers for infectious diseases using serological survey data. Biostatistics 2003; 4:621,632.
15. Hethcote HW. The Mathematics of Infectious Diseases. SIAM Rev. 2000; 42: 599-653.
16. Nishiura H, Schwehm M, Kakehashi M, Eichner M. Transmission potential of primary pneumonic plague: time inhomogeneous evaluation based on historical documents of the transmission network. J. Epidemiol. Community. Health. 2006;60: 640-645.

17. Committee on Modeling Community Containment for Pandemic Influenza. Modeling Community Containment for Pandemic Influenza: A letter report. (National Academies Press, 2006).
18. Castillo-Chavez C, Feng Z, Huang W. On the computation R0 and its role on global stability. In: Castillo-Chavez C, van den Driessche P, Kirschner D, Yakubu A-A (eds.) Mathematical Approaches for Emerging and Reemerging Infectious Diseases: An Introduction. Berlin: Springer-Verlag, 2002: 229-250.
19. Heffernan JM, Smith RJ, Wahl LM. Perspectives on the basic reproductive ratio. Journal of the Royal Society, Interface / the Royal Society. 2005; 2: 281-293.
20. van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. 2002; 180: 29-48.
21. Cauchemez S, Boelle P-Y, Donnelly CA, Ferguson NM, Thomas G, Leung GM et al. Real-time estimates in early detection of SARS. Emerg. Infect. Dis. 2006; 12: 110-113.
22. Cauchemez S, Boelle P-Y, Thomas G, Valleron A-J. Estimating in real time the efficacy of measures to control emerging communicable diseases. Am. J. Epidemiol. 2006; 164: 591-597.
23. Bettencourt LMA, Ribeiro RM. Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. 2007: submitted.
24. Chowell G, Nishiura H, Bettencourt LMA. Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. J. R. Soc. Interface. 2007; 4: 155-166.
25. Bettencourt LMA, Cintron-Arias A, Kaiser DI, Castillo-Chavez C. The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models Physica A. 2006; 364: 513-536.
26. Petitti DB. Meta Analysis, Decision Analysis and Cost-effectiveness Analysis: Methods for Quantitative Synthesis in Medicine. New York: Oxford University Press, 2000.
27. Website: `http://postgis.refractions.net/`
28. Website: `http://www.postgresql.org/`
29. Website: `http://www.r-project.org/`
30. `http://cran.r-project.org/src/contrib/Descriptions/DCluster.html`
31. `http://sourceforge.net/projects/cadkit`