Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Theoretical Biology 241 (2006) 954–963

Journal of Theoretical Biology

www.elsevier.com/locate/yjtbi

A susceptible-infected model of early detection of respiratory infection outbreaks on a background of influenza

Mojdeh Mohtashemi[a,b,*], Peter Szolovits[b], James Dunyak[a], Kenneth D. Mandl[c,d]

[a] *MITRE Corporation, Bedford, MA, USA*
[b] *MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA*
[c] *Children's Hospital Informatics Program at the Harvard-MIT Division of Health Sciences and Technology, Boston, MA, USA*
[d] *Harvard Medical School, USA*

Received 5 October 2005; received in revised form 17 January 2006; accepted 26 January 2006
Available online 23 March 2006

Abstract

The threat of biological warfare and the emergence of new infectious agents spreading at a global scale have highlighted the need for major enhancements to the public health infrastructure. Early detection of epidemics of infectious diseases requires both real-time data and real-time interpretation of data. Despite moderate advancements in data acquisition, the state of the practice for real-time analysis of data remains inadequate. We present a nonlinear mathematical framework for modeling the transient dynamics of influenza, applied to historical data sets of patients with influenza-like illness. We estimate the vital time-varying epidemiological parameters of infections from historical data, representing *normal* epidemiological trends. We then introduce simulated outbreaks of different shapes and magnitudes into the historical data, and estimate the parameters representing the infection rates of anomalous deviations from normal trends. Finally, a dynamic threshold-based detection algorithm is devised to assess the timeliness and sensitivity of detecting the irregularities in the data, under a fixed low false-positive rate. We find that the detection algorithm can identify such designated abnormalities in the data with high sensitivity with specificity held at 97%, but more importantly, early during an outbreak. The proposed methodology can be applied to a broad range of influenza-like infectious diseases, whether naturally occurring or a result of bioterrorism, and thus can be an integral component of a real-time surveillance system.

Keywords: SI model; Transients; Early detection; Infectious disease; Outbreaks; Biosurveillance

1. Introduction

The global health, threatened by emerging infectious diseases, pandemic influenza, and biological warfare, is becoming increasingly dependent on the rapid acquisition, processing, integration and interpretation of massive amounts of data. In response to these pressing needs, new information infrastructures are needed to support active, real time surveillance. Critical for real time surveillance are two components: data collection and data analysis. Indeed today, there are several real time outbreak monitoring systems in place in major metropolitan cities (Mandl et al., 2004; Tsui et al., 2003; Lewis et al., 2002; Lober et al., 2002; Reis et al., 2003). Despite such progressive efforts, the state of the practice for detecting temporal abnormalities in surveillance data is not adequate.

For highly infectious agents, such as SARS or smallpox, a few infected individuals can propagate the infection at a rate that might initially elude the attention of public health authorities, thereby gaining time for the infected pool to grow silently but rapidly to the point after which public health measures will prove ineffective. Therefore, the challenge for any model of outbreak detection lies in the early recognition of such exponentially growing processes, when the exponential nature of the process is not easy to recognize.

Abbreviations: CH ED, Children's Hospital Emergency Department; ED, Emergency Department

[*] Corresponding author. MIT Computer Science, 77 Massachusetts Avenue, Cambridge, MA, USA.

E-mail addresses: mojdeh@mitre.org, mojdeh@mit.edu (M. Mohtashemi).

0022-5193/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2006.01.031

With respect to an infectious agent, a population consists of those who are susceptible, infected or immune to the disease. Depending on the clinical and epidemiological properties of the disease there may be other categories. Here, we model the short-term dynamic interaction between different subpopulations with respect to an infectious disease using a nonlinear system of difference equations. Such a system will comprise meaningful demographic and epidemiological parameters, representing transient properties of the underlying dynamics, which can be estimated from historic epidemiological data. The resulting body of information will be the basis for defining *normality* in epidemiological trends and therefore can be used to detect anomalous deviations from historically observed events early on. We make the key assumptions that such disease processes are highly infectious, highly contagious, and manifested with non-specific flu-like symptoms in patients early in the development.

The methodology presented here represents a fusion among the developing methods of syndromic surveillance (Mandl et al., 2004; Tsui et al., 2003; Lewis et al., 2002; Lober et al., 2002; Reis et al., 2003; Greenko et al., 2003; Goldenberg et al., 2002), the notion of transients in health and disease (Mohtashemi and Levins, 2001; Mohtashemi, 2001) and well-established concepts and approaches in mathematical epidemiology (Bailey, 1967; Edelstein-Keshet, 1988; Anderson and May, 1992).

2. Methods

The data set used for this study consists of the daily number of patients presenting to the emergency department (ED) of a large urban, academic pediatric or children hospital (CH) with respiratory syndromes during the period 6/1/1992–5/31/2003. Fig. 1 illustrates the historical time series during the 5-year period 6/1/1996–5/31/2000. ED chief complaints were used to select encounters for infectious respiratory illness that are highly reflective of population patterns of influenza (Bougeois et al., in press; Brownstein et al., 2005). Chief complaint codes were chosen during the triage process, from a pre-defined on-line list of 181 choices. A previously validated subset of the constrained chief complaint set was chosen a priori for inclusion in the respiratory syndromic grouping (Beitel et al., 2004). Institutional review board approval was obtained. It is important to note that the syndromic respiratory definition closely corresponds with influenza activity as shown in Brownstein et al. (2005), which further strengthens our key assumptions about the historical data representing flu-like illnesses.

2.1. Model

Consider the following first-order nonlinear system of difference equations:

$$\begin{aligned} S_{n+1} &= S_n - \beta S_n I_n, \\ I_{n+1} &= I_n + \beta S_n I_n - \delta I_n, \end{aligned} \tag{1}$$

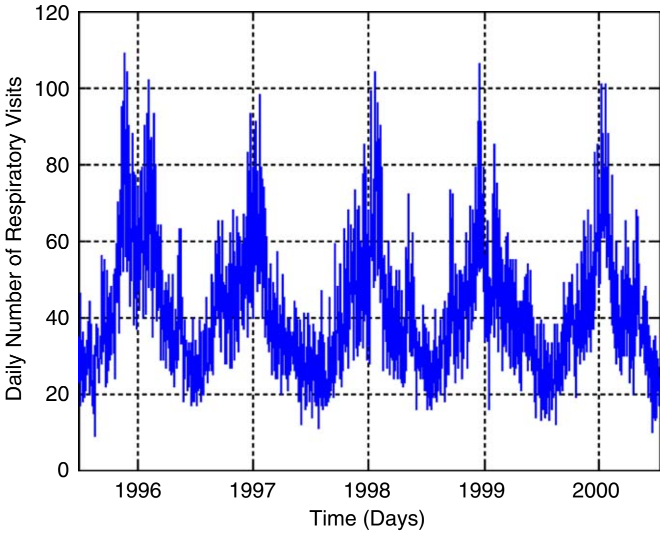

Fig. 1. Daily number of visits with respiratory infections to the CH ED during 6/1/1996–5/31/2000.

where S_n and I_n represent the respective number of susceptible and infected individuals at time n; β is the infection transmission rate; and δ is the average rate of recovery from infection so that $1/\delta$ is the mean duration of infectivity in days. Because we are interested in modeling transients in the population transmission dynamics of flu-like illnesses, implicit in our assumption is that there is not enough time for the temporary removal of the recovered population to be of significance to the dynamics. Delays introduced by temporary recovery and return to the susceptible class, although common in flu-like infections, are typically much longer than the infectivity period. Furthermore, there is no explicit representation for the latent population and the latency period in our model.

2.2. System transformation

Understanding population susceptibility is key to the control and prevention of epidemics of infectious disease. However, population susceptibility is difficult to account for. That is, the proportion or number of people susceptible at a point in time cannot be observed or measured systematically. On the other hand, the proportion or number of people infected at a point in time is an observable variable. Transforming an epidemiologically sensible system of two variables S_n and I_n, into an equation with one variable I_n, for which there are data, eliminates the unobservable variable while preserving the underlying dynamics (Mohtashemi, 2001). If we apply the elimination process to the above system of equations, then we end up with a second-order equation in the variable representing the infected population:

$$I_{n+2} = \frac{I_{n+1}^2}{I_n} - \beta I_{n+1}(I_{n+1} - (1-\delta)I_n). \tag{2}$$