Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 235 (2005) 326–337
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Scaling properties of childhood infectious diseases epidemics before and after mass vaccination in Canada

Helen Trottier*,[1], Pierre Philippe

Department of Social and Preventive Medicine, University of Montreal, Pavillon 1420 boul. Mont-Royal, Montréal, Canada H2V 4P3

Received 14 September 2004; received in revised form 20 December 2004; accepted 24 January 2005
Available online 7 March 2005

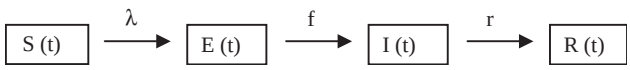
Abstract

The goal of this paper is to analyse the scaling properties of childhood infectious disease time-series data. We present a scaling analysis of the distribution of epidemic sizes of measles, rubella, pertussis, and mumps outbreaks in Canada. This application provides a new approach in assessing infectious disease dynamics in a large vaccinated population. An inverse power-law (IPL) distribution function has been fit to the time series of epidemic sizes, and the results assessed against an exponential benchmark model. We have found that the rubella epidemic size distribution and that of measles in highly vaccinated periods follow an IPL. The IPL suggests the presence of a scale-invariant network for these diseases as a result of the heterogeneity of the individual contact rates. By contrast, it was found that pertussis and mumps were characterized by a uniform network of transmission of the exponential type, which suggests homogeneity in the contact rate or, more likely, boiled down heterogeneity by large intermixing in the population. We conclude that the topology of the network of infectious contacts depends on the disease type and its infection rate. It also appears that the socio-demographic structure of the population may play a part (e.g. pattern of contacts according to age) in the structuring of the topology of the network. The findings suggest that there is relevant information hidden in the variation of the common contagious disease time-series data, and that this information can have a bearing on the strategy of vaccination programs.

Keywords: Childhood infectious diseases; Scaling properties; Inverse power law; Scale-invariant network; Exponential function

1. Introduction

Basically, there are three methods to extract the structure of infectious disease time-series data. Firstly, the deterministic compartmental models, such as SEIR, allow the identification of the main actors of the transmission of infections in populations (Anderson and May, 1991). This well-known method mirrors qualitatively the observed *trends* of the infectious disease time series from large populations insofar as the biological characteristics of the modeled disease are known (e.g. the duration of the period of infectivity, latent period, and immune status after infection). The SEIR model captures the mechanics of disease infection in the population and allows for successful predictions. Thus, the following SEIR model could be applied to measles:



where the rates (λ, f, r) of change (per unit time) in the number susceptible (S), exposed (E), infectious (I) and immune (R) are given by the following

*Corresponding author. Tel.: +1 514 398 1489; fax: +1 514 398 5002.
E-mail addresses: htrottier@sympatico.ca (H. Trottier), philippp@sympatico.ca (P. Philippe).
[1]Division of Cancer Epidemiology, Department of Oncology, Faculty of Medicine, McGill University, 546 Pine Avenue West, Room B1, Montreal, Canada, H2W 1S6.

0022-5193/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2005.01.013

equations:

$$dS/dt = -\lambda S(t),$$
$$dE/dt = \lambda S(t) - fE(t),$$
$$dI/dt = fE(t) - rI(t),$$
$$dR/dt = rI(t),$$

where $S(t)$ is the number susceptible at time t, $E(t)$ the number infected but not yet infectious at time t, $I(t)$ the number infectious at time t, $R(t)$ the number immune at time t, λ the rate (force) of infection per unit time, f the rate at which an infected individual becomes infectious per unit time, and r the rate at which an infectious individual recovers per unit time.

The SEIR model reflects a collective dynamic from the mean individual biological parameters of the disease. The SEIR is an *explanatory* model of the collective dynamic. The SEIR is useful in modeling moderately complex dynamics.

Secondly, there is the stochastic modeling technique of Box–Jenkins, also known as the autoregressive, integrated, moving average (ARIMA) method (Box et al., 1994). The method models the empirical time series by seeking the parameters that will reproduce the variations in the series as accurately as possible. The method actually models the dependent structure embedded in the time series. If we denote the values of a series at equally spaced time t, $t-1$, $t-2,\ldots$ by Z_t, Z_{t-1}, $Z_{t-2}\ldots$, the Box–Jenkins method distinguishes three possible processes that account for the dependent structure between successive observations:

(1) *Autoregressive process* (*AR*): $Z_t = a_t + \phi_1 Z_{t-1}$. Here, the current time-series observation (Z_t) is determined by a portion of the preceding observation (Z_{t-1}), and a current random shock (a_t) (order 1 process). The above model is a special case of the following general model (an autoregressive model of order p): $Z_t = a_t + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p}$.
(2) *Moving average process* (*MA*): $Z_t = a_t - \theta_1 a_{t-1}$. In this model, the current time-series observation, Z_t, is composed of a current random shock (a_t) and portion of the preceding random shock (a_{t-1}) (order 1 process). This simple model is a special case of the following general model (a moving average model of order q): $Z_t = a_t - \theta_1 a_{t-1} - \cdots \theta_q a_{t-q}$.
(3) *Integration* (*I*): $Z_t = a_t + Z_{t-1}$. If a series has to be *differenced* to stabilize the mean, then the model corresponding to the original series is called integrated. In this process, Z_t is the sum of the current random shock (a_t) and the preceding observation.

Empirical time series can be adjusted by any one of these processes. Also, more than one process can be used to fit the data such as an ARMA(p,q) model. Moreover, cyclical or periodic fluctuations that repeat themselves regularly in time should also be taken care of if they occur. It is also possible albeit rare to find a series that embeds more than one term of the same process. The aim of Box–Jenkins modeling is to transform a time series into a white noise (uncorrelated random residuals). It follows from the above that a model must remain as simple as possible in order to be consistent and should contain only short-term dependencies (autocorrelations) to be modeled adequately. If, on the contrary, long-term autocorrelations involving nonlinearities are embedded in the data, the Box–Jenkins model may fail to adjust the data properly unless one is ready to increase the order of the model; doing so, however, may jeopardize the stability of the estimation of the coefficients and invalidate the model. Therefore, the Box–Jenkins methodology, although well known, was rarely applied to time series of infectious diseases. The method allows one to *describe the collective pattern* of the dynamics while ignoring the biological specifics of the disease and the nonlinearity of the disease process. This result has been confirmed by our latest analyses (Trottier and Philippe, 2004a).

In order to gain more insight into the dynamic of population disease processes, a new model, the inverse power-law (IPL) distribution function, has been suggested for time series from small populations and large mass vaccinated populations (Rhodes and Anderson, 1996a,b; Rhodes et al., 1997,1998). The method is stochastic but allows one to recognize a deterministic self-organizational structure in the series (Barabasi and Albert, 1999). The frequency distribution of epidemic sizes over time can bring out the topological structure of the contact network that caused the epidemics. In other words, the structure of the interactive network of contacts, when fit to an IPL distribution function, can explain the type of propagation and the size of epidemics in the population over time (Barabasi and Bonabeau, 2003; Liljeros et al., 2001). This is usually called scaling analysis and consists in adjusting epidemic size distributions to an IPL of the type: $N(>s) = as^{-b}$, where s is the epidemic size, $N(>s)$ is the frequency of epidemics of atleast size s, and a and b are parameters to be estimated from the data.

The IPL is a hyperbolic function with no characteristic scale, i.e. it has no natural scale, and the function can fit extreme values of the distribution since $N(>s)$ decays according to b, the slope of the power law (Schroeder, 1991). The IPL is therefore characterized by no finite mean and an infinite variance (Bak and Chen, 1991; Barabasi and Albert, 1999; West and Shlesinger, 1990). To fix ideas, an exponential function has a characteristic scale with well-defined (finite) moments. The IPL is self-similar upon rescaling (scale invariance), i.e. it involves a hierarchy of multiple scales embedded into one another, a property that can explain why very small epidemics co-exist with very large ones in the