Environmental Research 142 (2015) 319–327

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Time series regression model for infectious disease and weather

Chisato Imai [a,*], Ben Armstrong [b], Zaid Chalabi [b], Punam Mangtani [c], Masahiro Hashizume [a]

[a] Department of Pediatric Infectious Diseases, Institute of Tropical Medicine, Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan
[b] Department of Social and Environmental Health, London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH UK
[c] Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

ARTICLE INFO

Article history:
Received 21 January 2015
Received in revised form
26 June 2015
Accepted 28 June 2015

Keywords:
Time series
Method
Infectious disease
Weather
Climate

ABSTRACT

Time series regression has been developed and long used to evaluate the short-term associations of air pollution and weather with mortality or morbidity of non-infectious diseases. The application of the regression approaches from this tradition to infectious diseases, however, is less well explored and raises some new issues.

We discuss and present potential solutions for five issues often arising in such analyses: changes in immune population, strong autocorrelations, a wide range of plausible lag structures and association patterns, seasonality adjustments, and large overdispersion.

The potential approaches are illustrated with datasets of cholera cases and rainfall from Bangladesh and influenza and temperature in Tokyo. Though this article focuses on the application of the traditional time series regression to infectious diseases and weather factors, we also briefly introduce alternative approaches, including mathematical modeling, wavelet analysis, and autoregressive integrated moving average (ARIMA) models.

Modifications proposed to standard time series regression practice include using sums of past cases as proxies for the immune population, and using the logarithm of lagged disease counts to control autocorrelation due to true contagion, both of which are motivated from "susceptible-infectious-recovered" (SIR) models. The complexity of lag structures and association patterns can often be informed by biological mechanisms and explored by using distributed lag non-linear models. For overdispersed models, alternative distribution models such as quasi-Poisson and negative binomial should be considered. Time series regression can be used to investigate dependence of infectious diseases on weather, but may need modifying to allow for features specific to this context.

1. Introduction

Time series regression (TSR) is widely used among environmental epidemiologists to examine associations between environmental predictors and adverse health outcomes. The method has been developed to evaluate the associations of air pollution and weather with all-cause mortality or morbidity in places where this is overwhelmingly due to non-infectious diseases (e.g. cardiovascular diseases). More recently, TSR approaches from this tradition have been applied to communicable diseases (Hashizume et al., 2008; Jusot and Alto, 2011; Lin et al., 2013; Luque Fernandez et al., 2009; Mangtani et al., 2006). However, the use of TSR in this context is less well explored and raises some new issues (Imai and Hashizume, 2015).

This article aims to discuss and present solutions to the most important issues arising for studies using TSR models to investigate associations of weather with infectious diseases. Though few of the issues we discuss are unique to infectious diseases, they are posed in ways that require some adaptation of the approaches developed for non-infectious diseases, and our main aim is to describe such adaptations. We make reference to alternatives to TSR that have also been considered from mathematical modeling, signal processing, or econometric traditions in particular when aspects of them can be incorporated into a TSR approach, but those methods are not described in detail.

Where we propose solutions, we illustrate them using datasets of influenza in Tokyo and cholera in Bangladesh (see Supplemental material pages 2 and 7 for details of the data). These two infectious diseases demonstrate short term immunity (or diseases with frequent changes in antigenic strains or subtypes) and long term

* Corresponding author.
E-mail addresses: chisato.imai@gmail.com (C. Imai), ben.armstrong@lshtm.ac.uk (B. Armstrong), Zaid.Chalabi@lshtm.ac.uk (Z. Chalabi), punam.mangtani@lshtm.ac.uk (P. Mangtani), hashizum@nagasaki-u.ac.jp (M. Hashizume).

http://dx.doi.org/10.1016/j.envres.2015.06.040
0013-9351/

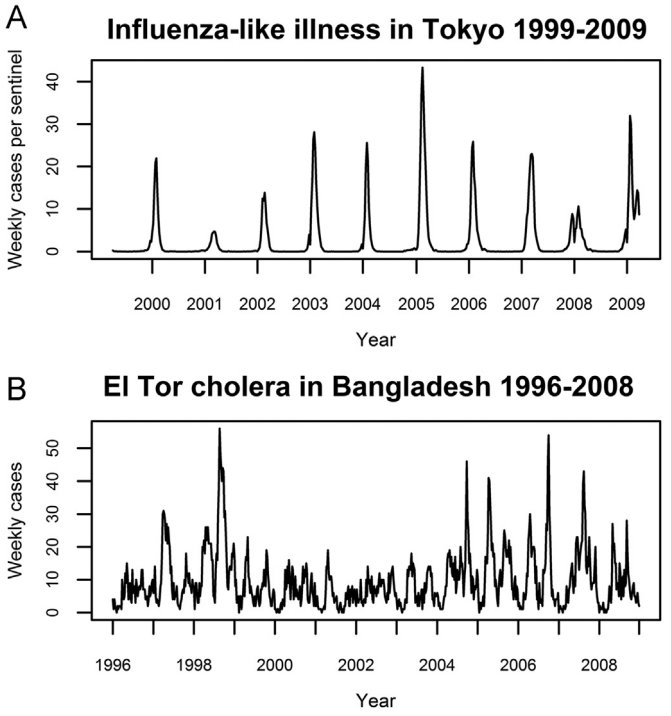

Fig. 1. (A) Weekly total influenza-like illness cases per sentinel medical facility in Tokyo, 1999–2009. (B) Weekly total El Tor cholera cases from laboratory confirmed infections from the hospital at ICDDR, B in Dhaka, Bangladesh, 1996–2008.

immunity respectively. These datasets are graphed in Fig. 1. Brief summary results are presented in the main text. More detailed results, R code, and the influenza data (the cholera data are not public) are available in Supplemental material.

This article begins with brief summaries of the TSR model typically used for non-infectious diseases in environmental epidemiology and a time series susceptible-infectious-recovery (SIR) model from the mathematical modeling tradition. These are then followed by five sections, each of which addresses an issue arising in the application of TSR to infectious diseases, and a discussion.

1.1. Overview of the time series regression model

The traditional TSR analysis seeks to identify how measured time-varying factors x_t (e.g. temperature) explain variation in an outcome series Y_t, usually daily counts of disease occurrence. The Poisson model is the most common TSR model, and can be presented as

$$\begin{aligned} Y_t &\sim Poisson(\mu_t) \\ \log(\mu_t) &= \beta_0 + \beta x_t + \sum_p \beta_p z_{p,t} + f(t) \end{aligned} \tag{1}$$

where $f(t)$ is a smooth function of time t designed to model and so avoid confounding by season and long term trend, x_t denotes an observed time varying variable of interest such as temperature, and $\{\beta_0, \beta, \beta_p\}$ are regression coefficients. Other measured risk factors are denoted as $z_{p,t}$. This model, in particular the choice of a suitable time function $f(t)$, has been reviewed in a recent tutorial paper (Bhaskaran et al., 2013). As in general in the TSR tradition, Bhaskaran focuses on acute effects to non-infectious disease outcomes.

1.2. Overview of the time series susceptible-infectious-recovery model

A feature of infectious diseases is that survivors of the disease are often immune to re-infection for some time. This causes potentially rapid changes in the population susceptible to infection. In particular, one possible explanation for the waning course of epidemics after the peak is that the susceptible population is exhausted, or at least, given herd immunity, susceptible contacts of infected cases become too sparse for infection propagation.

The SIR model is based on this and other known mechanisms for the dynamics of immunity and transmission among population. When combined with time series data this approach is called the TS-SIR (or sometimes TSIR) model. One variant of this model, simplified from Koelle (Koelle and Pascual 2004; Koelle et al., 2005), can be written in discrete time as

$$Y_t = \theta_{t-1} Y_{t-1}^{\alpha} \left(\frac{S_{t-1}}{N_{t-1}}\right)^{\gamma} \varepsilon_t \tag{2}$$

where, N_t is the total population size, S_t is the number of susceptible individuals, θ_t is pathogen transmissibility at time t and ε_t is multiplicative noise. α and γ are parameters associated with the type of mixing between individuals. S_t is not observed, but estimated from subtracting the sum of fractions of past incident cases where fractions immune κ_i are assumed to smoothly decline with the intervening time step $t - i$:

$$S_t = N_t - \sum_{i=0}^{m} Y_{t-i}\kappa_i \tag{3}$$

where m is the total duration of immunity (in time steps). At first sight this seems quite different from the traditional TSR framework, but taking logarithms and making a Taylor series approximation (details in Koelle and Pascual (2004)) reveal a strong similarity

$$\log(Y_t) \cong \log(\theta_{t-1}) + \alpha \log(Y_{t-1}) - \frac{\gamma}{N_t} \sum_{i=0}^{m} Y_{t-i}\kappa_i + \log(\varepsilon_t) \tag{4}$$

Koelle used this approach to estimate parameters θ_{t-1} for each time point (smoothed and separated from the seasonal component) and then considered associations with weather (and other explanatory factors) in a second stage, but there seems no reason why direct incorporation of explanatory variables in a single stage, as in the TSR would not be possible. However, given the large number of constrained parameters (the θ_{t-1} and κ_i), the complete Koelle model does not quite fall within the traditional TSR framework, therefore we decided not to pursue it here, though this approach may be an interesting subject for future research.

2. Topic 1: Immune population

The traditional TSR generally assumes that the population at risk of the outcome under study is more or less constant; however, as noted in the SIR overview above, immunity to infectious diseases causes variation in the susceptible population. Unless allowed for, such variations in the underlying population at risk may bias estimates of the associations with the weather.

If the size of the susceptible population were known at each time point, it could be allowed for in the model, but this information does not always exist. We review below some approaches to this problem. Choice of the approach is likely to depend on the specific infectious disease, given the large variations among infections in the duration of immunity.

2.1. Rely on smooth function of time to model changes in immunity

Much of the effect of changes in immune fractions of populations is often to induce seasonal and other long term variations of diseases (Grassly and Fraser, 2006; Pascual and Dobson, 2005).