USING CONTACT TRACING DATA TO MEASURE EPIDEMIC DYNAMICS

Chris Jewell & Gareth Roberts

[Centre for Research in Statisical Methodology, University of Warwick]
chris.jewell@warwick.ac.uk

Introduction

Numerical methods are used increasingly to advise on control policy for infectious disease epidemics. However, for these approaches to be truely quantitative, it is important to use observed data to make measurements on the dynamics of the outbreak. Once this is done, the results can be used to make predictions on how the epidemic might unfold, and hence provide information for decisions on control to be made.

Often, models are contructed to reflect assumptions about the population structure. Individual-level covariates are then related, via parameters, to infection rate. Inference on these model parameters then provides the necessary information to drive epidemic simulations for prediction purposes.

Here we show that the acquisition of epidemic contact tracing data serves to augment the information contained in these covariates in order to make more accurate epidemic predictions.

Contact Tracing Data

Contact Tracing Data (CTD) is typically collected from a newly detected case during an epidemic. It comprises of a list of known contacts with other individuals during a specified period of time prior to the detection - the **contact tracing window**. This has the aim of identifying further individuals who might have been infected (but are currently undetected), and also helps in reconstructing the path of the epidemic through the population.

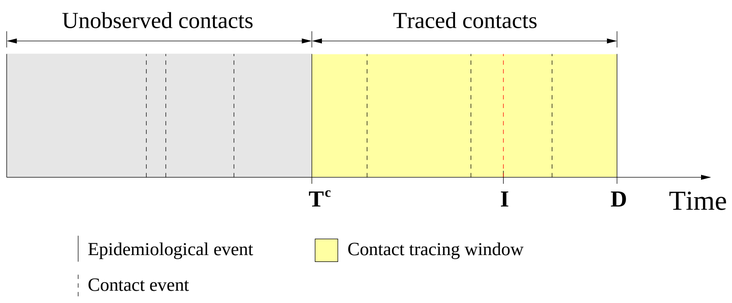

Figure 2: The timing of events leading up to an infected individual's (ie farm's) detection (**D**). Contacts occur at random intervals, and result in an infection with some probability p (provided they have originated from an infected individual). Contact events within the contact tracing window (interval $[T^c, N)$) are known (yellow area), whereas those which occur before T^c are only known to occur with a given rate (r).

The Model - Contact Tracing

To construct a model, we consider contacts in terms of arriving at individual j. If CTD were available for all time, we could assume that the number of contacts originating at infected individuals, and arriving at a suceptible j before it becomes infected, follows a negative binomial distribution:

$$\Pr(\mathcal{C}_j(I_j^-)|p) \propto p(1-p)^{\mathcal{C}_j(I_j^-)}$$

where $\mathcal{C}_j(I_j^-)$ is the set of *potentially* infectious contacts arriving at j before one results in an infection (all other contacts being statistically irrelevant).

The Model - Contact Rates

In the case where contact tracing is *not* available, information on the **rate** of contacts can be used in the classic stochastic epidemic setup. In our previous work, we have adopted a continuous time SINR model in which individuals progress according to:

Susceptible → Infected → Notified → Removed

for which the likelihood is:

$$f(\boldsymbol{N}, \boldsymbol{R}|\boldsymbol{I}, \boldsymbol{\theta}, \boldsymbol{X}) = \prod_{j=1, j\neq\kappa} (\tau_j(I_j^-)) \cdot \exp\left[\int_{I_\kappa}^{T_{obs}} \left(\sum_{j=1, j\neq\kappa}^{[\mathcal{S}_{T_{obs}}]} \tau_j(t)\right) dt\right]$$

$\tau_j(t)$ represents the rate of infections (infectious pressure) arriving at individual j at time t, typically a function of the covariates and parameters as shown in the next section; $\boldsymbol{N}, \boldsymbol{R}$ are the vectors of notification (ie detection) and removal times respectively; I_j^- is the time just before j is infected; T_{obs} is the analysis time; $\mathcal{S}_{T_{obs}}$ is the set of susceptibles at the analysis time.

Rates v CTD

At first glance, it appears from the two equations above that rate information and CTD are incompatible since they are different measures - rates can assume any non-negative value, whereas contacts either happen or not.
However, it can be shown that taking an expectation over the number of contacts occurring in the unknown periods gives the likelihood for the continuous time stochastic epidemic model[1]. By partitioning the likelihood into periods of unknown and observed contacts, inference can be made jointly using both rate and contact tracing information.

Example

These results are from work we have done to provide an inference and risk-prediction system for a potential outbreak of Highly Pathogenic Avian Influenza in the British Poultry Industry[2]. The Great Britain Poultry Register identifies the major production-type present on a farm j ($\boldsymbol{s}_j$), together with OSGB location data enabling the calculation of Euclidean distance between farms i and j (ρ_{ij}).

In addition, Network Data obtained by questionnaire identifies three matrices defining contact networks:

- Feedmills ($\boldsymbol{r}^{FM}$) - contact *rate* information for feed-lorry visits (median 1.1 lorry visits per farm per week).
- Slaughterhouses ($\boldsymbol{r}^{SH}$)) - contact *rate* information for abattoir-lorry visits (median 0.18 visits per farm per week).
- Company ($\boldsymbol{c}^{CP}$) - binary (0 or 1) information on business relationships between farms.

This model is therefore constructed to represent *infectious pressure* on susceptible farm j from infected farms i:

$$\tau_j(t) = \sum_{i\in\mathcal{I}(t)} \boldsymbol{\eta}\cdot\boldsymbol{s}_j \left[r_{ij}^{FM} p_1 + r_{ij}^{SH} p_2 + c_{ij}^{CP}\beta_1 + \beta_2 e^{-\psi(\rho_{ij})}\right]$$

Parameters to be estimated:

- $\boldsymbol{\eta}$ - a vector of production-type susceptibilities.
- p_1, p_2 - probabilities of an infection occurring *given* that a contact occurs between i and j for Feedmill and Slaughterhouse contacts respectively.
- β_1 - the infection rate between two farms connected by a company link.
- β_2, ψ - the spatial infection rate, and distance decay between two farms $\rho_i j$ apart.

Since $\boldsymbol{r}^{FM}$ and $\boldsymbol{r}^{SH}$ are rate matrices, we can incorporate contact tracing data and investigate how its addition improves the precision of our parameter inference on p_1 and p_2.

Results

For this study, a simulated epidemic was used since no sizeable outbreak of HPAI has yet occurred in Britain. Contact tracing data was simulated alongside, with a nominal contact tracing window beginning 21 days before case detection, and assuming perfect contact tracing data collection. Parameter estimation was performed using a Bayesian approach with Reversible Jump MCMC[3] both with and without contact tracing data.

Here we present our results for p_1 and p_2, having used Uniform$[0, 1]$ priors for both parameters. The true values (arbitrarily chosen for demonstration purposes) used to simulate the data are shown in relation to the posterior estimates. The graphs indicate that posterior variance is significantly reduced by the incorporation of contact tracing data into the analysis, thus enabling more efficient parameter estimation.

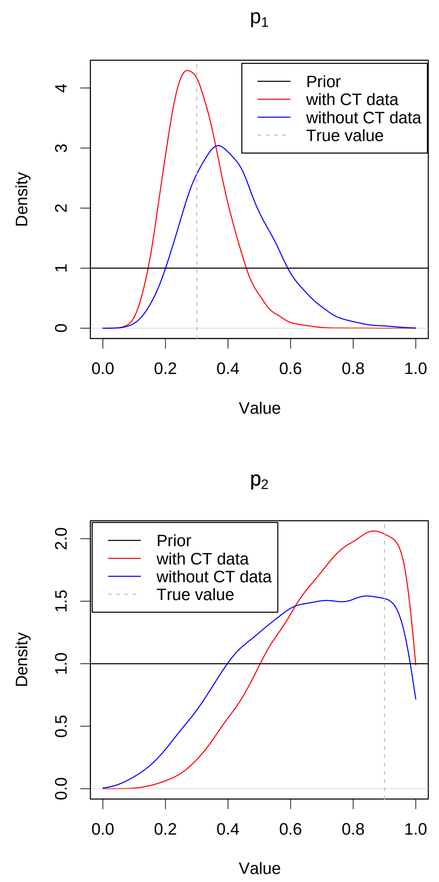

Figure 3: The decrease in posterior variance achieved by the addition of contact tracing data, supplementary to static network-frequency data

When is CTD useful?

Our early results suggest that using contact tracing data is only useful when contact frequencies are low. To see this, we consider how the variance of the parameter estimate vary in relation to the contact rate r.

Figure 4 shows how this variance varies in both the binomial contact tracing model (where contacts are known), and the Poisson Process model (where only contact rate is known). For all values of r, the binomial model gives a smaller variance. However, this difference is only appreciable for small values of r, in this case for r below about 3.

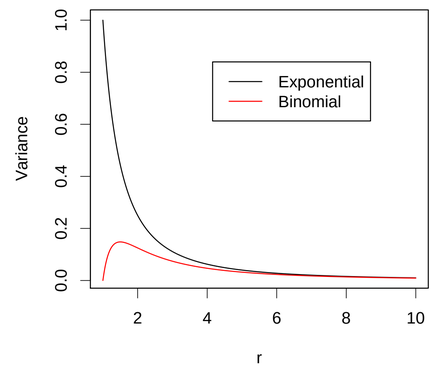

Figure 4: The variance of the binomial and exponential models as a function of the contact rate r, using a toy example of simple binomial and exponential models where $p = 0.5$.

Conclusions and Recommendations

Our results suggest that using CTD as an adjunct to static covariate data is a useful aid to providing accurate parameter estimates for epidemic prediction purposes. When contacts between individuals are relatively rare events, even small amounts of contact tracing information are valuable in providing large improvements in predictive uncertainty. Since "static" contact rate data is likely to change throughout an epidemic as control policy and farming behaviours change, the information that CTD gives will be vital in reducing the inherent inaccuracies in static data. Contact tracing data should therefore be considered an important part of modelling efforts during disease outbreaks.

References

[1] CP Jewell. *Real-time Inference and Risk-Prediction for Notifiable Diseases of Animals*. PhD thesis, Department of Mathematics and Statistics, Lancaster University, Lancaster University, Lancaster, LA1 4YF, UK., 2009.

[2] CP Jewell, T Kypraios, GO Roberts, and RM Christley. A novel approach to real-time risk-prediction for emerging infectious diseases: a case study in Avian Influenza H5N1. *Prev. Vet. Med.*, 2007. Submitted.

[3] CP Jewell, MJ Keeling, and GO Roberts. Predicting undetected infections during the 2007 foot-and-mouth disease outbreak. *JRS Interface*, 2008. [Epub ahead of print].