Epidemics 17 (2016) 35–41

Contents lists available at ScienceDirect

Epidemics

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

Preserving privacy whilst maintaining robust epidemiological predictions

Marleen Werkman[a,b,*], Michael J. Tildesley[a,c], Ellen Brooks-Pollock[d], Matt J. Keeling[a]

[a] WIDER Centre, Mathematics Institute and School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK
[b] Current address: Department of Infectious Disease Epidemiology, School of Public Health, Faculty of Medicine, St Marys Campus, Imperial College London, London, UK
[c] Fogarty International Center, US National Institute of Health, Bethesda, MD 20892, USA
[d] School of Social and Community Medicine, University of Bristol, Oakfield Grove, Clifton BS8 2BN, UK

ARTICLE INFO

Article history:
Received 9 March 2016
Received in revised form 10 October 2016
Accepted 12 October 2016
Available online 13 October 2016

Keywords:
Metapopulation
Simulations
Stochastic model
Spatial aggregation

ABSTRACT

Mathematical models are invaluable tools for quantifying potential epidemics and devising optimal control strategies in case of an outbreak. State-of-the-art models increasingly require detailed individual farm-based and sensitive data, which may not be available due to either lack of capacity for data collection or privacy concerns. However, in many situations, aggregated data are available for use. In this study, we systematically investigate the accuracy of predictions made by mathematical models initialised with varying data aggregations, using the UK 2001 Foot-and-Mouth Disease Epidemic as a case study. We consider the scenario when the only data available are aggregated into spatial grid cells, and develop a metapopulation model where individual farms in a single subpopulation are assumed to behave uniformly and transmit randomly. We also adapt this standard metapopulation model to capture heterogeneity in farm size and composition, using farm census data. Our results show that homogeneous models based on aggregated data overestimate final epidemic size but can perform well for predicting spatial spread. Recognising heterogeneity in farm sizes improves predictions of the final epidemic size, identifying risk areas, determining the likelihood of epidemic take-off and identifying the optimal control strategy. In conclusion, in cases where individual farm-based data are not available, models can still generate meaningful predictions, although care must be taken in their interpretation and use.

1. Introduction

Mathematical models form an integral part of epidemic preparedness planning and real-time forecasting (Keeling et al., 2001; Ferguson et al., 2001, 2006; Germann et al., 2006; Brooks-Pollock et al., 2014). State-of-the-art individual farm-based models involve detailed data: these data are often not available or, if available, are often not in the public domain owing to privacy concerns. However, sharing data will hugely benefit developing, optimising and training disease simulation models (Webb et al., 2016). In some countries, only spatially aggregated data are available. However, the full heterogeneity of individual farms may not be captured with these data (Keeling et al., 2010), and in cases where limited data are available, simpler models may be a necessity (Buhnerkempe et al., 2014). Striking the optimal balance between detail and utility is an open and problem specific question.

Individual farm-based models have been utilised in the past to aid in the understanding of epidemiological processes of Foot-and-Mouth Disease (FMD) and testing potential control strategies such as (ring) vaccination, culling of livestock and quarantine of infected premises, most notably during and in the aftermath of the UK 2001 epidemic (Tildesley et al., 2009, 2006). These models typically rely on the availability of detailed spatial information regarding the size and location of all livestock farms (Keeling et al., 2001). Whilst these data are available for the UK, this is not the case for many countries around the world. For example in the USA, farm location data are aggregated at the county level to prevent privacy difficulties (Buhnerkempe et al., 2013) and in Australia precise farm locations are not known for all states (Garner and Beckett, 2005). In many other countries around the world precise farm locations are not known at all.

In situations where detailed demographic data are not available but aggregated data are available, it may be possible to adopt a metapopulation approach when developing a mathematical model. Metapopulation models are often used in ecology, theoretical biol-

* Corresponding author at: Department of Infectious Disease Epidemiology, School of Public Health, Faculty of Medicine, St Marys Campus, Imperial College London, London, UK.
E-mail address: m.werkman@imperial.ac.uk (M. Werkman).

http://dx.doi.org/10.1016/j.epidem.2016.10.004

ogy and epidemiology (Levin et al., 1997; Hanski, 1999, 1994; Bolker and Grenfell, 1995; Keeling and Gilligan, 2000; Gaff and Gross, 2007). In metapopulation models, epidemiological units, such as farms, are spatially aggregated into patches or subpopulations. Within a patch, farms are assumed to be well mixed (in the sense that transmission occurs randomly between all pairs of farms, in a density-dependent manner) and behave uniformly. Transmission within and between patches must capture the physical processes and can occur via different routes and over different spatial scales, such as local processes (i.e. aerosol spread, direct contact of animals, contaminated vehicles or farm equipment) or by long distance contact such as live animal movements (Keeling et al., 2001; Ferguson et al., 2001; Green et al., 2006; Gibbens et al., 2001).

Given that within a patch, farms are assumed to be well mixed and behave uniformly, a metapopulation model will not capture the impact of local spatial clustering of farms, heterogeneity of farm size nor species composition. Previous studies have shown that these characteristics may often play an important role in epidemic dynamics (Keeling et al., 2001; Tildesley et al., 2010; Rock et al., 2014). In this study, we therefore investigate whether, and under what circumstances, a metapopulation model is a good alternative to an individual farm-based (IFB) simulation model. Our goal is to determine whether a novel metapopulation model gives comparable predictions to the IFB model when considering key epidemiological quantities such as spatial spread, epidemic size and distribution of epidemic size. The results presented here will ultimately have implications for human and veterinary health settings where precise locations of farm are unknown.

2. Materials and methods

2.1. Data and model

Information on farm locations, sizes and species compositions was obtained from the 2010 agricultural census provided by the Department of Environment, Food and Rural Affairs. Early versions of the model used in this paper assumed one single set of parameters for the whole UK (Keeling et al., 2001). However, a more accurate fit to the 2001 UK outbreak can be achieved by fitting individual parameter sets to five distinct regions of the UK – Cumbria, Devon, the rest of England (excluding Cumbria and Devon), Scotland and Wales (Tildesley et al., 2008) (Table S1). This allows for the model to capture region-specific farming practices and control implementation. In particular, lower transmissibility values are found in Wales (Tildesley et al., 2008), possibly owing to the increase in road distances between farms (particularly in the hilly sheep farming regions). Previous work indicates that this regionalised model provides a more accurate fit to the 2001 outbreak than a model with a single set of parameters for the entire country (Tildesley et al., 2008). In this paper, we consider outbreaks in Cumbria, Devon and Aberdeenshire and therefore utilise model parameters for these three counties (where Aberdeenshire parameters are fixed to those of Scotland).

2.2. Within farm dynamics

Farms are classified as susceptible, exposed, infectious, reported or culled (SEIRC). The latent period (time for an exposed farm to become infectious) is set to five days, consistent with estimates from 2001 (Keeling et al., 2001). After this period, farms remain infectious for four days before being reported. There is then a two-day delay (during this period the farm remains infectious) from reporting to culling in line with previous work (Keeling et al., 2001; Tildesley et al., 2009). We adopt a Markovian approach for the transition between classes; these are modelled as a constant rate, leading to exponential distributed periods. In this paper, we assume that the virus spreads rapidly when introduced in a naïve farm, such that within-farm dynamics can be excluded from the model and all animals on a farm are assumed to belong to the same disease status (i.e. all animals on a farm are either susceptible or exposed etc.).

2.3. Between farm dynamics

2.3.1. Individual farm-based (IFB) model

In the 2001 IFB model, local spread, incorporating multiple routes of transmission (trucks, airborne transmission etc.) is modelled via the use of a distance dependent transmission kernel (Keeling et al., 2001; Buhnerkempe et al., 2014). The local transmission kernel exhibits power-law like behaviour, such that farms (j) that are in the closest proximity of an infected farm i experience the largest risk of transmission:

$$K\left(d_{ij}\right) = \frac{\zeta}{1+\left(\frac{d_{ij}}{\theta}\right)^{\sigma}}$$

Parameters σ, θ and ζ define the shape and scale of the kernel. These parameters are estimated from the 2001 local transmission kernel (Keeling et al., 2001; Buhnerkempe et al., 2013; Rorres et al., 2010) and are set to 3, 1 and 0.12 respectively. The variable d_{ij} defines the Euclidean distance between any two farms i and j. In this model, the risk of infection is determined by the number of cattle and sheep on infected and susceptible farms and the Euclidean distance between them. The susceptibility (α) and infectivity $\left(\beta\right)$ are species-specific and scale non-linearly with farm size using parameters p and q. The stochastic rate of transmission from farm i to j is therefore given by:

$$\text{rate}_{ij} = \left(\alpha_c N_{c,j}^{p_c} + \alpha_s N_{s,j}^{p_s}\right) \times \left(\beta_c N_{c,i}^{q_c} + \beta_s N_{s,i}^{q_s}\right) \times K\left(d_{ij}\right)$$

where N represents the number of animals on a farm for cattle (c) or sheep (s).

Only cattle and sheep farms are included in this study; other species such as pigs are susceptible as well, but did not appear to play an important role during the 2001 and 2007 FMD outbreaks (Gibbens et al., 2001; Ryan et al., 2008).

2.3.2. Homegenous metapopulation model

For the metapopulation model, the UK is divided into grids (two-dimensional squared cells) with an equal width and height. Farms are then allocated to the grid cells based on their Easting and Northing coordinates of the farmhouse taken from the cattle tracing system (Brooks-Pollock et al., 2014; Green et al., 2006). To investigate the effect of the resolution of the grid cells on model outcomes, we vary the scale of the grid cells from 200 m (in order for most grid cells to contain only a single farm) to 10 km (with increments of 200 m up to 1 km and 1 km increments from 1 km to 10 km grid cell sizes). The upper limit was chosen as this fits with many spatial scales used in control such as surveillance zones. As with most metapopulation models, we assume that all farms within a grid cell are well mixed.

In order to estimate the mean transmission rate within and between any two grid cells, we calculate the distance between two randomly located farms in each grid cell by integrating over all possible locations of farms of the two grid cells k and l:

$$\text{mean}\left(kernel_{kl}\right) = \frac{1}{\|A_k\|\|A_l\|}\int_k\int_l K\left(\|x^l - y^k\|\right)dxdy$$

where A_k is the area of grid cell k and A_l is the area of grid cell l, and x and y refer to the point locations in grid cells l and k.