

# Control of equine influenza: scenario testing using a realistic metapopulation model of spread

M. Baguelin<sup>1,2,\*</sup>, J. R. Newton<sup>1</sup>, N. Demiris<sup>3</sup>, J. Daly<sup>4</sup>, J. A. Mumford<sup>2</sup>  
and J. L. N. Wood<sup>2</sup>

<sup>1</sup>*Animal Health Trust, Lanwades Park, Kentford, Newmarket CB8 7UU, UK*

<sup>2</sup>*Department of Veterinary Medicine, Cambridge Infectious Disease Consortium, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK*

<sup>3</sup>*MRC Biostatistics Unit, Institute of Public Health, Cambridge CB2 0SR, UK*

<sup>4</sup>*Viral Brain Infections Group, 8th Floor Duncan Building, Daulby Street, Liverpool L69 3GA, UK*

We present a metapopulation model of the spread of equine influenza among thoroughbred horses parametrized with data from a 2003 outbreak in Newmarket, UK. The number of horses initially susceptible is derived from a threshold theorem and a published statistical model. Two simulated likelihood-based methods are used to find the within- and between-yard transmissions using both exponential and empirical latent and infectious periods. We demonstrate that the 2003 outbreak was largely locally driven and use the parametrized model to address important questions of control. The chance of a large epidemic is shown to be largely dependent on the size of the index yard. The impact of poor responders to vaccination is estimated under different scenarios. A small proportion of poor responders strongly influences the efficiency of vaccine policies, which increases risk further when the vaccine and infecting strains differ following antigenic drift. Finally, the use of vaccinating in the face of an outbreak is evaluated at a global and individual management group level. The benefits for an individual horse trainer are found to be substantial, although this is influenced by the behaviour of other trainers.

**Keywords:** equine influenza; metapopulation; stochastic modelling; influenza control; simulated likelihood

## 1. INTRODUCTION

Equine influenza, as all mammalian influenza viruses, is transmitted by close contact, including via aerosolization (Turkington & Ashby 1998). Transmission at the population level is therefore determined by contact patterns within and organization of the host population, and these vary markedly between species (e.g. pig, horse, humans). Important drivers of transmission in horse populations include demographic structure and movements of horses; these are associated with exercise and competition requirements as well as sale and purchase of horses.

To date, models for equine influenza have studied different, but generally small-scale aspects of equine influenza transmission including the spread of influenza in a non-vaccinated population (Glass *et al.* 2002), the

extension to realistic seasonal demographic variations (Park *et al.* 2003) and the consequences of strain heterology (i.e. the antigenic difference between the infecting strains and the strain in the vaccine) on epidemics (Park *et al.* 2004). However, these models have been largely restricted to the study of single premises, even though the dynamics of between-yard transmission are obviously extremely important in the understanding of larger scale influenza outbreaks.

An important underlying reason for increased risk of outbreaks, as in humans, is the phenomenon of antigenic drift of the influenza virus (Potter 2002; Daly *et al.* 2004; Park *et al.* 2004). Indeed, compulsory equine vaccination policies, although offering benefit, fail to prevent outbreaks, due to the emergence of antigenically distant strains of virus from those contained in vaccines (for a description of that issue for equine influenza see Daly *et al.* (1996), for a way of quantifying and visualizing these ‘antigenic’ distances in humans see Smith *et al.* (2004)).

The antigenic drift of viral strains drives the risk of epidemics and thus has to be constantly monitored

\*Author and address for correspondence: Department of Veterinary Medicine, Cambridge Infectious Disease Consortium, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK (mb556@cam.ac.uk).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2009.0030> or via <http://rsif.royalsocietypublishing.org>.

(Mumford 1999), but at any particular time point, with a population reasonably well protected by vaccination, other factors are also important in determining population-level susceptibility. In particular, the structure of the population and the contact and mixing patterns are of primary importance. The work presented here includes development of metapopulation models to represent the transmission of equine influenza among a population of thoroughbred racehorses. Such models are useful in the development of risk assessments and definition of optimal vaccine policies, and are important because it is not possible to conduct experiments at a population level.

The aim of this work was to design a metapopulation model that incorporates complex population structures. We then developed methods to parametrize this model using epidemiological data from an outbreak that occurred during the spring of 2003 in Newmarket, UK, where a large number of racehorses were affected (Newton *et al.* 2006). The model was then used to address a series of applied questions related to the control of outbreaks.

## 2. MATERIAL AND METHODS

### 2.1. Data

The latent and infectious periods for influenza in individual horses were derived from published results of experiments carried out previously at the Animal Health Trust in Newmarket (Park *et al.* 2004). The remaining data, including horse demography and location, were from an outbreak of equine influenza that occurred in spring 2003 in Newmarket, UK, and for which a good description was available (Newton *et al.* 2006). Demographic data came from various specific sources, including trainer-targeted questionnaires and Bell (2003). Data also included the results of laboratory tests performed on samples collected during the 2003 epidemic, including serological data on pre-outbreak antibody levels for horses in some of the yards, estimates of the total number of infected horses at the end of the epidemic in some yards and vaccine histories. A statistical model describing predictors for individual horses becoming infected was also used (Barquero *et al.* 2007). Data are summarized in appendix A (tables 1 and 2) and a yard level movie of the recorded epidemic is provided in the electronic supplementary material 2.

### 2.2. Definition of the model

As for previous models of the spread of equine influenza (Glass *et al.* 2002; Park *et al.* 2003, 2004), the approach adopted here was based on the classical susceptible–exposed–infectious–recovered (SEIR; Allman & Rhodes 2004) model. Owing to the relatively small sizes (in terms of number of horses) of the yards that the horses were kept in (median=30, highest=190; Bell 2003), it was necessary to incorporate stochasticity into the development of an epidemic as, when the virus enters a yard, there is a non-zero probability that the virus dies out without transmitting (Renshaw 1991).

Such stochasticity also underlies the fact that, in real outbreaks, only a proportion of the yards get affected during any outbreak.

*2.2.1. Metapopulation structure of the model.* Two important assumptions of the SEIR model are that the population is ‘well mixed’ (equal probabilities of making contacts for all horses) and homogeneous (same response to infection). Scaling up the model to a multiple yard training centre with a geographically fragmented population, different qualities of vaccination and shared training facilities invalidates both assumptions.

To tackle this issue, the model groups animals with the same behaviour in the epidemic under the same index, representing animals, for example, in the same training yard or animals with the same vaccine history, etc. Then, the  $S$ ,  $E$ ,  $I$  and  $R$  sections of the population break down in vectors  $S = (S_i)_{0 < i \leq n}$ ,  $E = (E_i)_{0 < i \leq n}$ ,  $I = (I_i)_{0 < i \leq n}$  and  $R = (R_i)_{0 < i \leq n}$ , where, within each value of the index  $i$ , the populations are homogeneous and well mixed. The total number of horses indexed by  $i$  is represented by  $N_i = S_i + E_i + I_i + R_i$ .

In the simplest case, the index matches the different yards and  $n$  is equal to the total number of these yards. However, in more complex cases, such as heterogeneous response to vaccination,  $n$  can be bigger (see §2.4.2).

*2.2.2. Protection afforded by antibody.* Protection against influenza infection through inactivated vaccines in use in 2003 was largely antibody mediated, as evidenced experimentally (Mumford *et al.* 1983) and epidemiologically (Newton *et al.* 2000a). To increase the level of antibodies in horses and thus improve population-level protection against the virus, vaccination has been compulsory among racehorses in Great Britain since 1981. There is considerable variation in individual response to vaccination and, particularly when less than four doses have been given, a rapid exponential decline in antibody levels after vaccination (Newton *et al.* 2000a). Other factors such as vaccine type can also have an impact (Barquero *et al.* 2007). The choice of product and schedule varies between trainers, and so, at the start of any epidemic, not all yards are protected equally.

Extending the studies of Park *et al.* (2003), in order to represent the variability between yards, we defined a level of susceptibility  $\alpha_i$  for each yard. Then, the level of initially susceptible horses is given by  $S_i(0) = \alpha_i N_i$  and the initially removed (protected) horses by  $R_i(0) = (1 - \alpha_i) N_i$ .

*2.2.3. Contact matrix and transmission.* The metapopulation structure implies that we have to consider the way the animals ‘mix’, i.e. the structure of infectious contacts between animals from different yards. For this, we defined the matrix  $T = (T_{ij})_{0 < i, j \leq n}$ , which summarizes how the animals mix with each other.  $T_{ij}$  is the transmission rate from yard  $i$  to yard  $j$ , and thus is a measure of the intensity of contacts between the two yards. The different forces of infection are given by  $\Delta = (\delta_i)_{0 < i \leq n} = TI$ . Thus, a horse passes from  $S_i$  to  $E_i$  with a rate  $\delta_i = \sum_{j=1}^n T_{ij} I_j$ .

**2.2.4. Latent and infectious periods.** The latent and infectious periods (respectively, the time spent by a horse in classes  $E$  and  $I$ ) are often modelled with exponential distributions with transition rate  $a_i$  (from class  $E_i$  to class  $I_i$ ) and  $g_i$  (from  $I_i$  to  $R_i$ ); such an approach has convenient mathematical properties. Park *et al.* (2004) demonstrated that this was inappropriate given experimental data especially as the infectious period appeared to be bimodal (see figure 1 in the electronic supplementary material 1). Here, both the exponential distributions and the empirical distributions (as adopted in Park *et al.* 2004) were used. The distributions corresponding to vaccination with heterologous viruses, rather than those homologous with epidemic strains, were used (Park *et al.* 2004).

### 2.3. Fitting the model with available data

The aim of parametrizing the model is to enable reproduction of epidemics similar to the outbreak seen in Newmarket in 2003 and then to explore various options for control.

To be able to use the data, assumptions regarding the contact structure had to be made.

The yard-level susceptibility levels were derived by scaling up the risk predicted by the horse-level statistical model from Barquero *et al.* (2007) using the relationship between final size, local transmission and susceptibility levels from the threshold theorem (Andersson & Britton (2000), see equation (2.1) for details). The last two parts of this section describe the methods used to numerically estimate the local and global transmissions parameters from data.

**2.3.1. Choice of the contact structure.** Previously observed patterns of transmission of influenza virus in Newmarket suggested that the geographical location of the yards probably had an impact on the way that the infection spreads (J. L. N. Wood, J. R. Newton & J. A. Mumford 1990–2005, unpublished data). However, estimation of the spatial effects on transmission would need some knowledge of the pathway of viral transmission or at least some clear patterns identified at the end of the epidemics. Analysis of the final picture of and the data during the epidemic in 2003 did not reveal any clear pattern (see movie in the electronic supplementary material 2).

In the absence of any clear way of incorporating geographical data, we considered whether the epidemic could be explained by two levels of mixing averaging both local (within yard) and global (between yard) transmissions. It is thus assumed that the ‘density’ for all yards is the same, i.e. the contact intensity is the same for each combination of two yards. As there is no information to model it differently, and as the range of contact between yards is similar (with some variations for yards sharing training for example), it is a reasonable approximation. This model is very similar to the one used for human households (Ball *et al.* 1997), although we consider that the transmission within the yard follows a true mass action law, as the transmission

between horses should be dependent on the density of infectious horses present in the yard and not on the size of the yard (for a discussion on the use of a pseudo or a true mass action, see de Jong *et al.* 1995).

The contact matrix  $T$  has thus the following form:

$$T_{ij} = \begin{cases} \frac{\lambda_L}{N_i} + \frac{\lambda_G}{N}, & \text{if } i = j, \\ \frac{\lambda_G}{N}, & \text{otherwise,} \end{cases}$$

with  $\lambda_L$  the local transmission;  $\lambda_G$  the global transmission;  $N = \sum_{i=1}^n N_i$ ,  $N_i$  the number of horses in yard  $i$ ; and  $n$  the number of yards on the site.

We assumed that the epidemic started with an individual horse that was infected in the yard where the disease was first detected during the 2003 outbreak.

**2.3.2. Estimation of yard-level susceptibility.** Previous models estimating susceptibility levels used results from challenge experiments to assess the probability of a horse seroconverting (viewed as a significant increase in antibody) as a function of its antibody level prior to infection (Park *et al.* 2004). However, for the 2003 outbreak in Newmarket that we consider here, antibody levels were not sufficient to explain differences in risk (Barquero *et al.* 2007), and six variables (sex, age of first vaccination, date and type of last vaccination, first time a N/1/93 strain was administered and level of N/1/93 antibodies) were necessary to predict the probability that a given horse would get infected if its yard was infected.

From this statistical model, it is thus possible to derive the mean expected numbers of infected horses in a yard if the epidemic kicks off. The susceptibility levels can then be estimated by scaling up the means produced by the statistical model using a version of the threshold theorem approximation (Andersson & Britton 2000) for a vaccinated population. It gives a relationship between the transmission rate, the mean infectious period, the susceptibility level and the final number of infectious horses (which can be predicted by the model) inside a single yard for a sufficiently large population (unpublished simulations show that the approximation is reasonably precise even for small-size populations such as the ones considered here). If we call  $\gamma_i$  the final proportion of infected animals predicted by the model inside the yard  $i$  (as an average of the individual probabilities to get infected), we then have  $\gamma_i \approx \alpha_i \tau$  with  $\tau$  being the non-trivial solution of  $1 - e^{-\alpha_i \tau \lambda_L / g_i} = \tau$ , with  $\alpha_i$  the proportion of fully susceptible horses;  $\lambda_L$  the local transmission rate; and  $g_i$  the local recovery rate. This can be rearranged to express  $\alpha_i$  as a function of  $\lambda_L$  as

$$\alpha_i(\lambda_L) \approx \frac{\gamma_i}{1 - e^{-\gamma_i \lambda_L / g_i}}. \quad (2.1)$$

The protection induced by vaccination affects the dynamic aspects of the disease. For example, for a single yard with a transmission rate ( $S \rightarrow E$ ) following

a true mass action law  $\lambda SI/N$  and a level of susceptibility of  $\alpha$ , the progression of the disease is equivalent to a model with a fully susceptible initial population  $S(0)$  and a transmission rate  $\lambda' = \alpha\lambda$ . The removal of some of the susceptible animals through vaccination is thus effectively equivalent to a reduction in the transmission parameter.

*2.3.3. Estimation of local and global transmissions (first method).* For a grid of values of the local and global transmissions, it is possible to simulate epidemics and count the occurrences of:

- (i) the number of times the overall size of the final simulated epidemic is in the range of the actual one (between 24 and 25% of the total population), and
- (ii) the number of times this epidemic involves the same number of yards as in the real epidemic (infected horses were detected in 21 yards, see [Newton \*et al.\* 2006](#)).

It is then possible to estimate  $\lambda_L$  and  $\lambda_G$  by considering that the frequency of occurrences respecting both conditions is a simulated likelihood of the given pair of parameters, in a similar way as for approximate Bayesian computation ([Beaumont \*et al.\* 2002](#); [Marjoram \*et al.\* 2003](#)). From this, it is possible to estimate the local and global transmissions as the means of these simulated samples.

As the data necessary to evaluate the risk inside yards were available only for a small proportion of the site, we needed to approximate the risk in the site to be the average one in the available samples (400 horses in 10 yards), we had set  $\gamma = 0.7287$  for all yards, the corresponding levels of susceptibility in each yard being derived for each level of  $\lambda_L$  following equation (2.1).

This method provides a first approximation and allows us to quantify how much of the transmission is local and global. If the transmission appears to be largely local, it would thus be possible to estimate also the local and global transmissions by using the available data (yard and overall proportions of infected horses) for 10 individual yards in the site (see appendix A, table 2).

*2.3.4. Estimation of local and global transmissions (second method).* We use a similar simulated likelihood approach here, but we also assume that epidemics are mainly driven locally (i.e.  $\lambda_G \ll \lambda_L$ ). It is then possible to consider the 10-yard final sizes from the epidemic and predicted sizes from the statistical model to estimate the likelihoods of the different values of the local transmission for each of these yards—and then go on to estimate  $\lambda_G$  based on these  $\lambda_L$  estimates. A more precise value of local transmission can then be worked out by considering the likelihood of the event where all the 10 yards simultaneously end with the same number of infected horses as in the actual outbreak data.

It is theoretically possible to derive, knowing the level of protection of the yard, the probability that an

outbreak of a certain size occurred as a function of the contact rate ( $\lambda$ ) and of the infectious period distribution ( $\mathcal{I}$ ). This can be done through the recursive resolution of a set of triangular equations (see [Andersson & Britton 2000](#)). Unfortunately, the evaluation of these probabilities numerically breaks down because of the appearance of terms smaller than the precision of standard programming. One solution was to increase the precision of the calculations ([Demiris & O'Neill 2006](#)), but limits were met very quickly when increasing the size of studied yards. We tackled this issue by estimating these probabilities through repeated simulations.

If, for a given infectious period distribution  $\mathcal{I}$ , we call  $P_{N,Z,\alpha}(\lambda_L)$  the probability that an epidemic of size  $Z$  occurs in a yard of size  $N$  with a level of susceptibility  $\alpha$  as a function of the transmission rate  $\lambda_L$ , then the probability that a set of yards with known sizes  $N_i$  and known levels of susceptibility  $\alpha_i$  endure exactly sub-epidemics of different sizes  $Z_i$  is

$$P(\lambda_L) = \prod_{i=1}^{10} P_{N_i, Z_i, \alpha_i}(\lambda_L).$$

$\lambda_L^{\text{est}}$  is then estimated as

$$\lambda_L^{\text{est}} = E(\lambda_L) = \frac{\int_0^\infty \lambda_L P(\lambda_L) d\lambda_L}{\int_0^\infty P(\lambda_L) d\lambda_L}.$$

Once the level of local transmission ( $\lambda_L$ ) has been estimated, assuming the structure of contacts presented in §2.3.1, for any global transmission parameter ( $\lambda_G$ ), there corresponds a probability  $Q(\lambda_G)$  that the model yields the actual final proportion of infected horses (estimated to be between 24 and 25%). The global transmission parameter,  $\lambda_G$ , is then given by

$$\lambda_G^{\text{est}} = \frac{\int_0^\infty \lambda_G Q(\lambda_G) d\lambda_G}{\int_0^\infty Q(\lambda_G) d\lambda_G}.$$

## 2.4. Exploration of different control measures and epidemic seeding pattern

*2.4.1. Quantifying the impact of the size of the seeding yard.* The previous simulations in this paper were all seeded from the yard believed to be where the virus was introduced in 2003.

In a series of numerical experiments, we wanted to quantify the impact of the size of the yard where the virus entered. For this, we plotted distributions of series of epidemics for different values of  $N_i$  from the smallest yard ( $N_{\min} = 7$ ) to the biggest one ( $N_{\max} = 190$ ), using both exponential and empirical distributions for latent and infectious periods. We also plotted the probability that an epidemic is larger than 10 per cent of the total population, as well as the mean size of such large epidemics, as a function of the number of horses in the index yard.

*2.4.2. The impact of poor responders.* In recent studies, it has been shown that some 7.5 per cent (22 out of 292) of racehorses present near to zero antibody levels



(Daly *et al.* 2004). This could be the result of horses responding weakly to vaccination, poor compliance or inadequate attention to optimal vaccine administration (e.g. bad storage of the vaccines). The model was used to test the consequences of incorporating a proportion of animals that we assumed to be unvaccinated among the vaccinated ones; scenarios that differed in how this unvaccinated population was distributed between yards were then explored.

Two different distributions of the poor responders were tested. In one, the ‘unvaccinated’ horses are randomly distributed between the yards, as would be expected if a proportion of all horses failed to respond adequately to vaccination. This was achieved through assigning such animals a different index to others in the same yards, with the mixing pattern inside the yard remaining the same, but with different latent and infectious periods and initial susceptibility in the poorly vaccinated group. In the other scenario, the unvaccinated animals are found in only a few yards, as would be expected if these animals were due to poor trainer compliance. To do so, we sequentially and randomly chose yards until the sum of their horses exceeded the number of observed poor responders. The last yard was then divided into two patches in order to have the appropriate overall proportion of poor responders in the population. These two scenarios are two extreme situations representing the case where poor compliance is totally random or entirely due to human intervention. The real situation probably lies somewhere between. Both of these two scenarios were also compared with the ideal ‘perfectly vaccinated’ situation for different level of vaccine protection.

*2.4.3. Vaccination in the face of an outbreak.* Vaccination in the face of an epidemic is an important measure in the veterinary armoury for the control of disease outbreaks. The result of such a measure as applied to large outbreaks of equine influenza among previously vaccinated racehorses is unclear. On the one hand, well-targeted vaccination could accelerate epidemic extinction by removing susceptible animals; on the other hand, vaccination when an epidemic has already started can be a costly operation implemented too late to prevent the development of the epidemic due to delayed responses to killed virus vaccines. In addition, failure to comply with good biosecurity during the conduct of mass vaccination programmes may exacerbate infectious transmission.

To test the role possibly played by revaccinating horses when an epidemic has already started, we had to address the fact that the level of susceptibility changes with the time of the year due to the natural decline of antibodies after seasonal vaccination (Park *et al.* 2004). Thus, all horses were considered potentially susceptible at the start of the epidemic, but with an antibody level  $l(t)$  attached to each one; this changes in response to vaccination. When a horse in yard  $i$  has been exposed to the virus, it can at a rate  $a_i E_i p(l(t))$  become infectious and at a rate  $a_i E_i (1 - p(l(t)))$  be removed as being protected by its antibody level, with  $p(l(t))$  being the

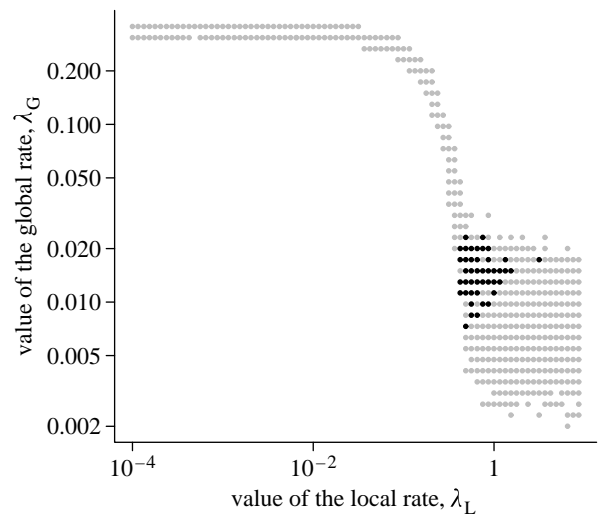


Figure 1. Parameter values where overall final epidemic size matches that observed in 2003 (final sizes between 24 and 25% of the overall population; grey points) and where the final size and number of yards affected in 2003 were both matched (additional exact same number of yards affected, 21—black points). These simulations used empirical distributions for the latent and infectious periods. For each of the 6400 points in the grid, 5000 simulations were run.

probability of seroconverting as a function of mean antibody level, varying with time (probability of seroconversion taken from the heterologous case in Park *et al.* 2004).

To determine the profile of the mean antibody level, we used the demographic model from Park *et al.* (2003) that integrates the compulsory vaccine schedule with demographic variation due to the sales and purchases of young horses. The profile of antibody levels for a ‘typical yard’ during the course of an entire year is shown in figure 2 of the electronic supplementary material 1 (measured with respect to one of the vaccine strains).

Once an epidemic (associated with a particular season) has started, it continues unabated until ‘detected’. Detection occurs in the model when a certain variable number of infectious horses appear in a 3-day interval (as in Savill *et al.* 2006). After detection, a control policy is chosen, each yard being able ‘to choose’ to vaccinate or not. The epidemic continues with two types of populations, one in which antibody continues to decline and the other in which antibodies are increasing in response to vaccination. The boosting is represented by a 6-day delay during which the antibody level of the boosted horse remains stable and then a linear increase in the antibody level to reach the peak value after 11 days (Powell 1988).

We first quantified the impact of vaccination when 80 per cent of trainers boosted following detection. Then, we studied the impact of varying of the detection threshold on the final size of the outbreak. Finally, we analysed the impact of the behaviour of other trainers on the benefit for individual trainers vaccinating.

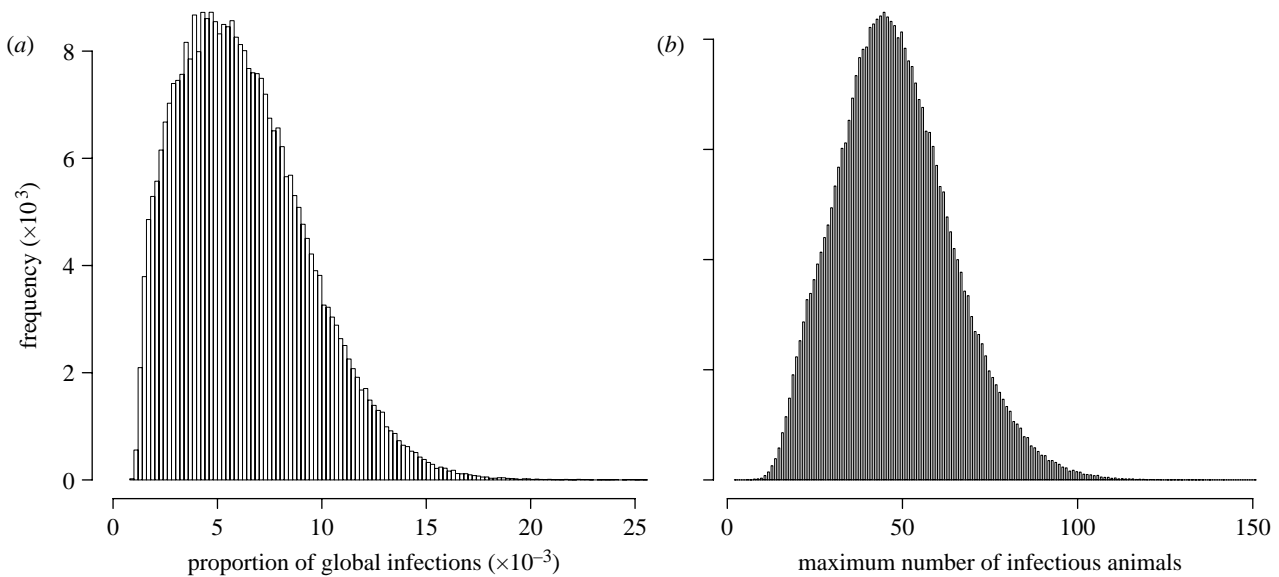


Figure 2. (a) Distribution of the ratios of the globally derived infections inside a yard after a virus has entered a yard. (b) Distribution of the maximum number of infectious animals during an epidemic. The total number of animals in the site is 2030. Only epidemics affecting more than one yard were considered here.

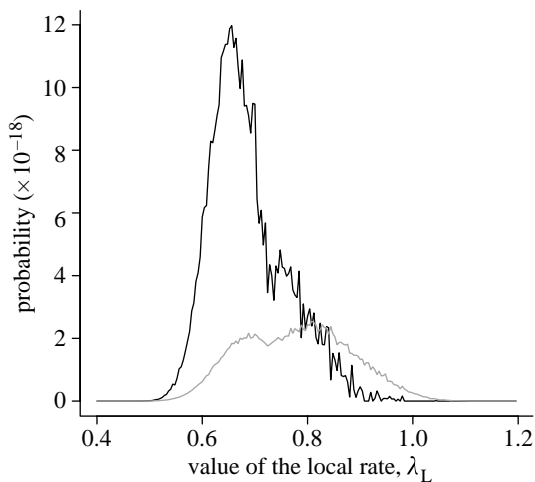


Figure 3. Probability of observing sub-epidemics, with the same sizes as 2003, as a function of  $\lambda_L$  for both exponential (grey curve) and empirical (black curve) infectious period distributions (for each value of  $\lambda_L$ ,  $2.5 \times 10^6$  simulations were run for the exponential distribution and  $2.5 \times 10^6$  for the empirical distribution).

### 3. RESULTS

#### 3.1. Estimation of model parameters

The data from the 2003 outbreak (Newton et al. 2006) were used to estimate the levels of susceptibility in each yard and the local and global transmission rates.

##### 3.1.1. Local and global transmissions (first method).

The first constraint on the final size of the epidemic gives a number of possible couples  $(\lambda_L, \lambda_G)$  (see grey points in figure 3 of the electronic supplementary material 1; figure 1). For the lowest values of  $\lambda_L$ , the epidemic is driven globally by the yard to yard transmission, thus the set of parameters is hardly

influenced by the variation of  $\lambda_L$ . As  $\lambda_L$  increases, global transmission has to go down to keep the balance. At high values of  $\lambda_L$ , the corresponding potential coupled values of  $\lambda_G$  are more variable than when  $\lambda_L$  is low.

Adding the constraint that the final number of yards affected should match exactly that observed in 2003 allows discrimination between these scenarios, as shown by the black points in figure 3 of the electronic supplementary material 1 and figure 1. The mean of  $\lambda_L$  and  $\lambda_G$  from this simulated likelihood provides an estimate of the pair  $(\lambda_L = 1.03; \lambda_G = 1.5 \times 10^{-2})$  for the exponential distributions and  $(\lambda_L = 0.7; \lambda_G = 1.5 \times 10^{-2})$  for the empirical ones.

Simulations show (figure 2) that when a yard becomes infected, less than 2 per cent (0.63% on average) of cases will come from reintroduction (from another yard), while the rest is due to local transmission. Unsurprisingly, reintroduction peaks when the number of infectious horses in the overall population peaks (median of 47 animals) and is also more common in larger yards. Epidemics are thus largely locally driven.

##### 3.1.2. Local and global transmissions (second method).

Given estimated levels of susceptibility, the probabilities of the observed final sizes in the 10 yards for different values of  $\lambda_L$  were derived and are shown in figure 3. The distribution of  $P(\lambda_L)$  for the exponential and empirical distributions are both shown. The estimated values of the intra-yard transmission  $\lambda_L$  were 0.78 for the exponential distribution and 0.69 for the empirical distribution.

The probability of matching the 2003 data, given different rates of  $\lambda_G$ , is shown in figure 3 of the electronic supplementary material 1, using  $\lambda_L$  derived from both infectious period distributions. The estimates of the global transmission rates are  $\lambda_G = 1.7 \times 10^{-2}$  for the exponential distribution and  $1.6 \times 10^{-2}$  for the empirical distribution.

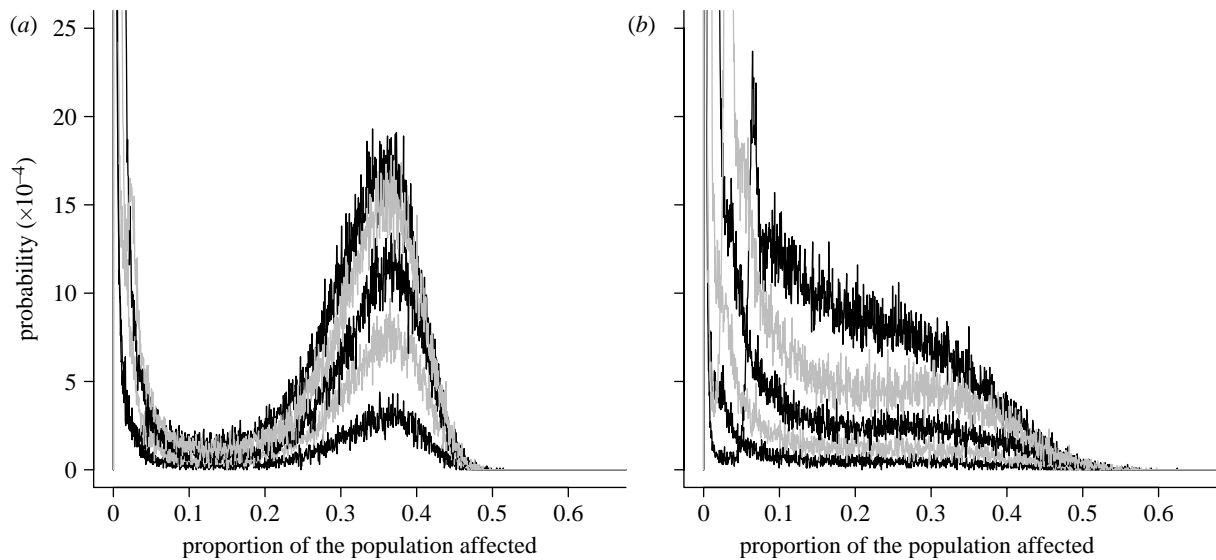


Figure 4. Change in the probability distribution of final epidemic sizes, by size of index yard, for the (a) exponential and (b) experimental distributions. Sizes of initial seeding yards shown are 7 (lower black curves), 23 (lower grey curves), 47 (middle black curves), 97 (upper grey curves) and 190 (upper black curves) horses.

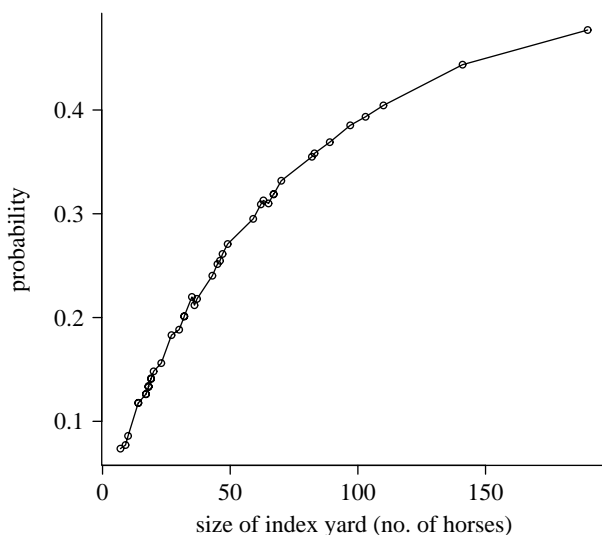


Figure 5. Probability of an epidemic involving more than 10% of the population, by size of the index yard initially.



Figure 6. Mean large (more than 10% of the population) epidemic size, by size of index yard.

### 3.2. Practical issues

For the rest of the simulations, we have used the values of  $\lambda_L$  and  $\lambda_G$  yielded by the second method (see §3.1.2; with the corresponding distributions), as there was greater uncertainty about the number of finally affected yards due to the difficulty of detecting yards in which some horses were infected but the virus did not take off.

#### 3.2.1. Impact of index yard size and control measures.

The distribution of the final sizes of the simulated global epidemic for both the exponential and experimental distributions is shown in figure 4. The bimodal aspect of the distribution is more evident for the exponential distribution, with a higher risk of large epidemics.

The probability of a large epidemic (defined as an epidemic involving more than 10% of the population) increases with the size of the initial index yard (figure 5). By contrast, the mean sizes of these epidemics remain approximately similar (mainly driven by the value of  $\lambda_L$  and  $\lambda_G$ ), and, after reaching a peak around the median yard size, the mean size even slightly decreases when the size of the seeding yard increases (figure 6). Indeed, epidemics starting from bigger yards are more 'explosive', thus lasting for a shorter duration of time, and leaving consequently more susceptible horses remaining.

**3.2.2. Poor responders.** A number of different scenarios were compared when considering the impact of poor responders on the risk of a large epidemic. We chose 10 per cent of the overall population affected to be the

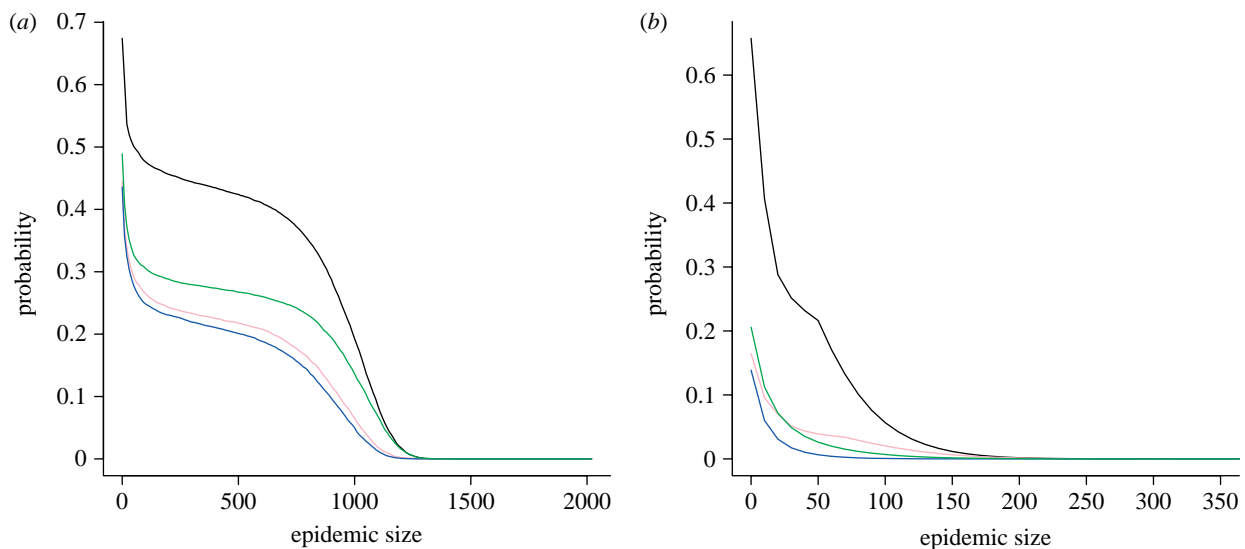


Figure 7. The impact of poor responders on the probability of different epidemic sizes. The blue curves represent the normal compulsory vaccination policy assuming all horses respond, and the pink, green and black curves represent the impact of different distributions of 7.53% of the population poorly responding. The pink represents random distribution of poor responders, the green has them concentrated in a few yards and the black has them concentrated in a few yards including the index yard. (a) Epidemic situation: 86% susceptible and (b) heterologous vaccination: 50% susceptible.

threshold at which epidemics are considered as being large. This value is rather arbitrary but corresponds approximately to the least frequent final epidemic size between bimodal peaks (figure 4). A ‘fully’ vaccinated population provided a baseline against which randomly spread poor responders, poor responders concentrated in a few yards and concentrated poor responders acting as the index yard were compared. Two series of simulations were run, one corresponding to the 2003 epidemic with a population 86 per cent susceptible and the other with a 50 per cent susceptible population, corresponding to a situation where vaccination is heterogeneous (Park *et al.* 2004).

There is a substantial increase in risk of an epidemic when part of the population is not protected and the impact depends on the level of susceptibility in the vaccinated population. For example, when 86 per cent of the population were susceptible, the probability that the epidemics involve more than 500 horses is 20 per cent for the fully vaccinated population, 22 per cent for the concentrated poor responders, 27 per cent for the randomly spread poor responders and 44 per cent if the index yard had concentrated poor responders. When the vaccine is more efficient (50% susceptibility), the probability of large epidemics is reduced but risks of having medium (between 50 and 150 horses) outbreaks remain and again are very sensitive to the presence of poor responders. In contrast to the first scenario, a concentration of poor responders is more detrimental than a random distribution, mainly due to large local epidemics.

Given the substantial impact of poor responders on the risks of epidemics and the sensitivity to the initial proportion susceptible (figure 7), we explored the impact of population susceptibility level further. Susceptibility level increases with time since seasonal vaccination and also with viral antigenic drift. The effects

of population-level susceptibility on the probabilities of final epidemic sizes more than 10 per cent of the population were compared between the same different poor responder scenarios as above (see figure 5 in the electronic supplementary material 1). To reach a 5 per cent chance of an epidemic size more than 10 per cent, susceptibility needs only to be 53 per cent where the index yard and others have a concentration of poor responders, it needs to be 66 per cent for both the random distribution and where poor responders are concentrated and 71 per cent if all horses respond to vaccination.

*3.2.3. Vaccination in the face of an epidemic.* The likely impact of vaccinating 80 per cent of the yards following detection of influenza and the benefits for an individual trainer alone revaccinating were explored. Finally, the consequences of changing the diagnosis or detection threshold were considered.

*Vaccination of 80 per cent of the yards.* The impacts of revaccinating 80 per cent of yards following detection of influenza on the distribution of epidemic final sizes for two different times of a year are shown in figure 8. The diagnostic threshold of 10 horses infected in 3 days was chosen for these simulations. The population is relatively protected in March, but in October is rendered relatively susceptible by the seasonal ingress of young horses. In March, 23 per cent of introductions take off and in October this increases to 44 per cent. In both cases, the impact of vaccination in the face of the outbreak considerably reduces the risk of a large epidemic. The efficiency of the measure is smaller in mid-October than in March, where very large epidemics are quite unlikely. Vaccination in October reduces median epidemic size from 1643 to 379 horses (mean sizes: 1570 to 411).



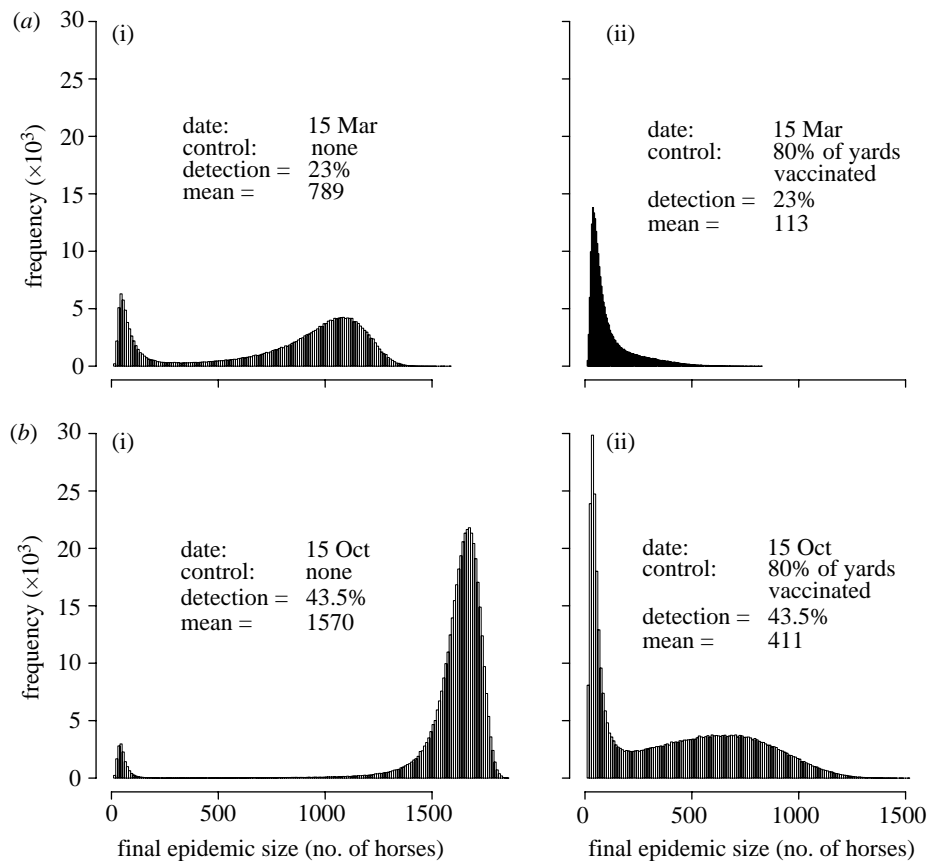


Figure 8. Impact of vaccinating 80% of the population following detection in (a) March and (b) October. (a(i),b(i)) The distribution of outbreak sizes with no further controls implemented and (a(ii),b(ii)) the same with 80% vaccination following detection. Each plot represents the output from  $10^6$  runs and the proportion of runs resulting in detection is also given.

*Benefits of vaccination for an individual trainer.* We considered the impact in the largest yard (190 horses) of boosting following diagnosis to maximize the detection of any effects of individual trainer behaviour. The individual benefits for a trainer boosting (figure 9) are highly dependent on the behaviour of other trainers. The chances of an outbreak in the yard are substantially reduced when other yards revaccinate, but the size of the within-yard outbreak remains large when the trainer does not boost. While the probability of an outbreak in the yard is not much reduced when the trainer alone boosts, the likely size of that outbreak is almost halved. The optimal situation is when there is boosting in the yard in question as well as in 80 per cent of others.

*Change in the detection threshold.* Finally, we explore the importance of surveillance and early detection by exploring the relationship between the mean size of the overall epidemic and the threshold of detection (figure 10). The efficiency of the vaccination measure is very sensitive to the threshold, with the mean final sizes of epidemics increasing almost linearly with the threshold being set at up to 15 horses. Comparing thresholds of 5, 10 and 15 horses, the mean epidemic sizes are 215, 411 and 557 horses, respectively. The slope of the curve flattens slightly above 15 horses. These results demonstrate the use of surveillance (and response to diagnosis) in reducing the severity of the outbreak.

#### 4. DISCUSSION

This paper describes a realistic model that captures the essential characteristic of the spread of an epidemic of equine influenza in a multi-yard scenario as in centres such as Newmarket where young racehorses are often concentrated. Being based on and parametrized from real data, it captures the demographic details of horse populations, the spread of the virus and its control and thus provides a reasonably realistic tool to study equine influenza control at the population level. The results of simulations using this model provide several important pointers for control, at both the training yard and the vaccine development levels. The results also strongly support the use and importance of sensitive disease surveillance when considering control of epidemic diseases.

Although practiced for several decades based on empirical evidence, this work quantifies the likely benefits of trainers revaccinating their horses following local diagnosis of influenza. While this measure is very effective at reducing the impact on the trainer's own horses, it is greatly enhanced when a significant number of other trainers revaccinate. Although this may sometimes be outside of the standard dosing schedule of the vaccine, this study provides clear evidence of the benefits of trainers acting together in this way. Although the extent of the benefits can vary between seasons, there are always benefits of reboosting at the

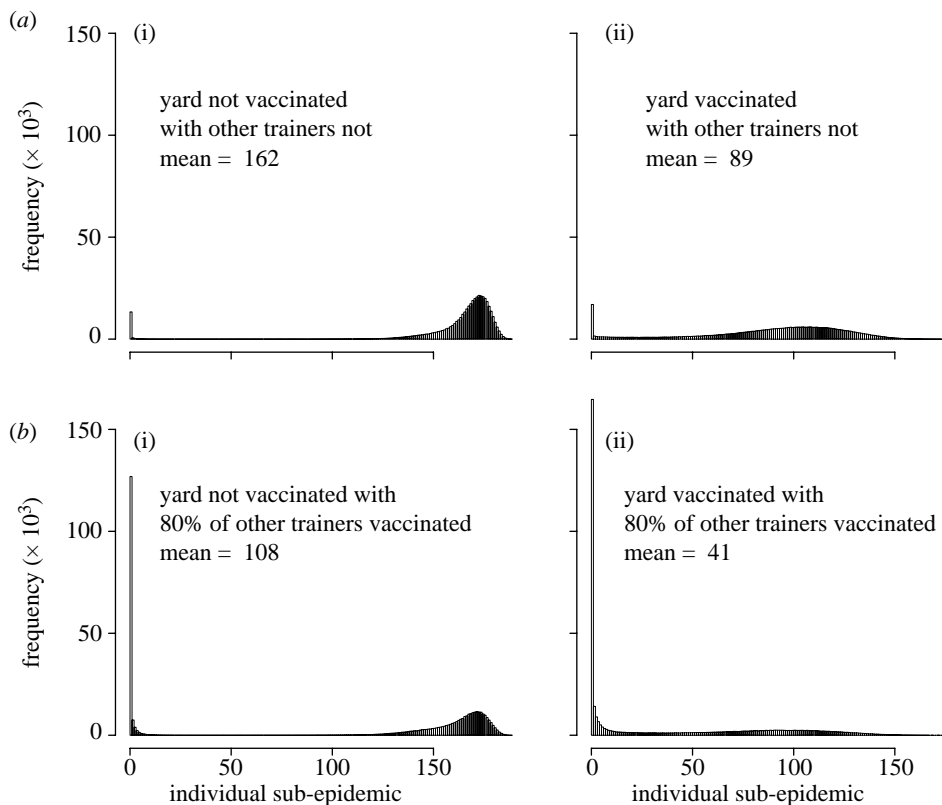


Figure 9. Individual benefit for a trainer to vaccinate his/her horses when an epidemic is detected depending on whether 80% of other trainers (*a*(i,ii)) do not vaccinate or (*b*(i,ii)) do vaccinate. Each plot is the result of  $10^6$  runs, the yard chosen houses 190 horses (the biggest yard of the site).

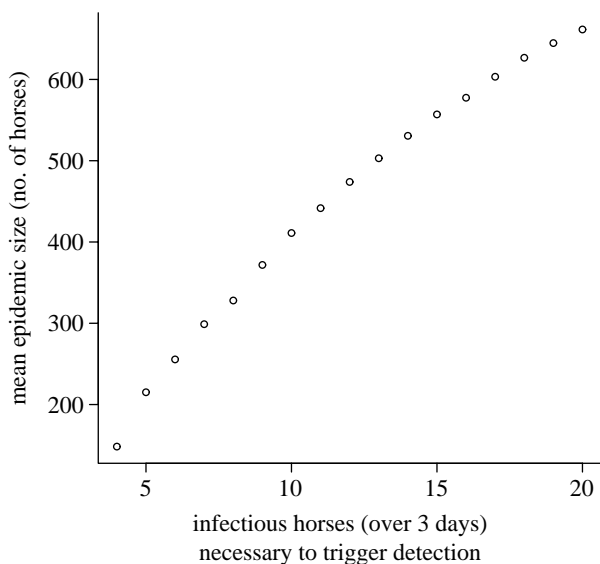


Figure 10. The impact of detection threshold on mean final epidemic size.

population level. Unsurprisingly, the strength of this effect was sensitive to the level of influenza surveillance. Respiratory disease is common in young racehorses (Burrell *et al.* 1996; Wood *et al.* 2005) and trainers will often ignore the first few cases of disease in their yard before investigating it. This study demonstrates how important it can be to differentiate low-level endemic

disease from an incursion of epidemic disease, as is also the case when considering control of avian influenza (Savill *et al.* 2006).

The results suggest that the proportion of horses responding poorly to vaccination do have an important impact on the risks for the population. Thus, all measures should be taken to try to reduce such poor responses in individual animals. Newer, better adjuvanted vaccines seem to reduce the proportion of such animals (J. A. Mumford 1990–2005, unpublished data), but further efforts to diminish this subset of the population further are justified. It has been known for some time that individual trainers can diminish the likely impact of epidemics by their general management (Newton *et al.* 2000*b*). It is now also clear that vaccine programmes associated with a reduced proportion of poorly responding animals will be of substantial benefit in the control of disease. Furthermore, any trainer not fully complying with compulsory vaccination programmes could have a substantial impact on the risks of disease in the rest of the population and the results justify continued efforts to ensure full compliance.

While the model developed to describe transmission of equine influenza is very general, the parametrization was only possible with a very simple structure of population (two levels of mixing). The question therefore arises concerning the usefulness of such a sophisticated model, given that the data were not optimal in providing information about the structure of contacts and this is the case for all the sets of data we reviewed in the

literature. Model improvement should not only depend on strictly fitting existing data, but it has also a role in exploring new pathways, even if they remain theoretical. We believe that this approach, by revealing which data might be suitable for better understanding the epidemic, will inform data collection from future outbreaks. The versatility of the model also allows a broad range of scenarios to be tested in an exploratory way and is consequently a potentially useful tool for policy making. Some of these scenarios will be the subject of future study.

We used different sources of data for modelling the effect of vaccination and this means that they do not always involve the same strains. Our conceptual approach to the effects of vaccination within groups is the same as taken by Park *et al.* (2003, 2004), and is based on detailed knowledge of responses of horses to vaccines and the close correlation between circulating antibody levels and the probability of becoming infected during heterologous challenge (Park *et al.* 2004); this has been validated by field observations (Newton *et al.* 2000a). Ideally, models would be based on assessments of vaccination that do not make assumptions about transmission between individuals, but would directly measure overall transmission rates in the context of specific infectious periods, as was done in experiments in chickens (van der Goot *et al.* 2005). However, the aim here was to estimate epidemiological transmission rates during an epidemic in the context of specific and rather predictable antibody levels, so that extrapolation to other situations might be possible. Furthermore, stabled horses are typically managed in a very different manner to experimental animals, management markedly affecting transmission, and realistic experimental studies are difficult to conceive in large groups of horses.

Increasingly, infectious disease models are moving away from assuming exponential distributions for infection and latent periods (Keeling & Grenfell 1998; Park *et al.* 2004). It is noticeable that the area under the probability curve for the experimental distribution of infection is larger (2.3 times bigger) than that for the exponential distribution, which suggests also that the experimental scenario seems more likely.

One issue concerns the fact that one estimation of local transmission rates depended on the assumption that sub-epidemics were considered to be independent. We went on to estimate the coupling of these different sub-epidemics when considering the global transmission  $\lambda_G$ . There is an apparent contradiction in this reasoning, as we assume something to be zero before estimating it as non-zero. In fact, the small difference in values between the estimated values for the two rates and the fact that only a limited portion of the whole site is affected at instant  $t$  shows that, in a multi-yard outbreak or epidemic, most of the infectious contacts occur between horses in the same yards, rather than between horses in different yards. The assumption of independence is therefore believed to be a reasonable approximation. An alternative approach would consider the random graph framework of Demiris & O'Neill (2005). That method could be slightly more accurate but requires assessment of algorithmic

convergence, and repeated epidemic simulations would again be necessary for the different scenarios we consider. Additionally, our method can be viewed as an approximate Bayesian computation approach with small  $\varepsilon$  (see Beaumont *et al.* (2002) and Marjoram *et al.* (2003) for details).

The coefficients  $\alpha_i$  will depend strongly, among other factors, on the match between the antibodies produced by vaccination and the infecting viral strain. Owing to the phenomenon of antigenic drift occurring in influenza viruses over time, protection is substantially reduced for heterologous (antigenically different) infecting and vaccinal viral strains (Park *et al.* 2004).

From a more theoretical point of view, the model presented in this paper gives a new insight into the way we see epidemics and their control. Classic SEIR models have stressed the importance of the basic reproductive ratio  $R_0$  on predicting the final issues of epidemics. This number can also be used for assessing vaccination policies. Our model also confirms that the existence of this threshold number is also connected to the structure of the population. Owing to the simplicity of having such a number (one quantity to assess the risk or not of an epidemic), some more complex models appear to have been designed to make similar threshold numbers evident. In our metapopulation model, different possible local transmission rates result in different local basic or effective reproductive ratios. Furthermore, other prospective simulations, with less restrictive assumptions about the coefficients describing more complex contact patterns, can lead to a lot more scenarios than the binary, small/big epidemics.

This work also paves the way to a population model that would take into consideration the evolutionary aspects of influenza viruses. This should allow us to address questions relating to the process that gives a certain shape to the phylogenetic trees of influenza viruses depending on the populations targeted, but also assess the impact of compulsory vaccination on the observed change of the nature of the disease.

This work was conducted through the generous financial support of the Horserace Betting Levy Board (M.B., J.R.N. and J.D.) and also through the DEFRA and HEFCE support for the Cambridge Infectious Diseases Consortium (J.L.N.W. and J.A.M.). We would also like to thank Bryan Grenfell for his advice at an early stage of this work and Joshua Ross, Roberto Saenz, Olivier Restif and Julia Gog whose criticisms significantly improved this manuscript. We thank three anonymous referees for their extremely useful criticisms.

## APPENDIX A. DATA

Table 1. Data provided by Richard Newton. (For more details about methods used, see Newton *et al.* (2006).)

number of yards affected	21
overall population affected	24.44%

Table 2. Final sizes of the sub-epidemics inside 10 yards during the 2003 outbreak in Newmarket and predicted risks given by the statistical model from Barquero et al. (2007). (The corresponding proportions of initial susceptible horses are given for one particular value of the parameters ( $\lambda_L = 0.78$  and  $g_i = 0.3$ ). Susceptibility levels are relatively similar for values of  $\lambda_L$  in the same range and can be calculated from the predicted risk using equation (2.1).)

yard	1	2	3	4	5	6	7	8	9	10
final size	0.3	0.94	0.78	0.56	0.18	0.76	0.85	0.73	0.93	0.5
predicted risk	0.6	0.92	0.71	0.63	0.26	0.68	0.9	0.88	0.88	0.69
susceptibility ( $\lambda_L = 0.78$ )	0.76	1	0.84	0.78	0.53	0.82	1	0.98	0.98	0.83

## REFERENCES

- Allman, E. S. & Rhodes, J. A. 2004 *Mathematical models in biology: an introduction*. Cambridge, UK: Cambridge University Press.
- Andersson, H. & Britton, T. 2000 *Stochastic epidemic models and their statistical analysis*. Lecture Notes in Statistics, vol. 151. New York, NY: Springer.
- Ball, F., Mollison, D. & Scalia-Tomba, G. 1997 Epidemics with two levels of mixing. *Ann. Appl. Probab.* **7**, 46–89. (doi:10.1214/aoap/1034625252)
- Barquero, N., Daly, J. M. & Newton, J. R. 2007 Risk factors for influenza infection in vaccinated racehorses: lessons from an outbreak in Newmarket, UK in 2003. *Vaccine* **25**, 7520–7529. (doi:10.1016/j.vaccine.2007.08.038)
- Beaumont, M., Zhang, W. & Balding, D. 2002 Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025–2035.
- Bell, L. 2003 *Horses in training 2003*. London, UK: Raceform Ltd.
- Burrell, M. H., Wood, J. N., Whitwell, K. E., Chanter, N., Mackintosh, M. E. & Mumford, J. A. 1996 Respiratory disease in thoroughbred horses in training: the relationships between disease and viruses, bacteria and environment. *Vet. Rec.* **139**, 308–313.
- Daly, J. M., Lai, A. C., Binns, M. M., Chambers, T. M., Barrandeguy, M. & Mumford, J. A. 1996 Antigenic and genetic evolution of equine H3N8 influenza viruses. *J. Gen. Virol.* **77**, 661–671. (doi:10.1099/0022-1317-77-4-661)
- Daly, J. M., Yates, P. J., Newton, J. R., Park, A., Henley, W., Wood, J. L. N., Davis-Poynter, N. & Mumford, J. A. 2004 Evidence supporting the inclusion of strains from each of the two co-circulating lineages of H3N8 equine influenza virus in vaccines. *Vaccine* **22**, 4101–4109. (doi:10.1016/j.vaccine.2004.02.048)
- de Jong, M. C. M., Diekmann, O. & Heesterbeek, H. 1995 How does transmission of infection depend on population size? In *Epidemic models: their structure and relation to data*, pp. 84–94. Cambridge, UK: Cambridge University Press.
- Demiris, N. & O'Neill, P. D. 2005 Bayesian inference for stochastic multitype epidemics in structured populations via random graphs. *J. R. Stat. Soc. B* **67**, 731–745. (doi:10.1111/j.1467-9868.2005.00524.x)
- Demiris, N. & O'Neill, P. D. 2006 Computation of final outcome probabilities for the generalised stochastic epidemic. *Stat. Comput.* **16**, 309–317. (doi:10.1007/s11222-006-8320-4)
- Glass, K., Wood, J. L. N., Mumford, J. A., Jesset, D. & Grenfell, B. T. 2002 Modelling equine influenza 1: a stochastic model of within-yard epidemics. *Epidemiol. Infect.* **128**, 491–502. (doi:10.1017/S0950268802006829)
- Keeling, M. J. & Grenfell, B. T. 1998 Effect of variability in infection period on the persistence and spatial spread of infectious diseases. *Math. Biosci.* **147**, 207–226. (doi:10.1016/S0025-5564(97)00101-6)
- Marjoram, P., Molitor, J., Plagnol, V. & Tavaré, S. 2003 Markov chain Monte Carlo without likelihoods. *Proc. Natl Acad. Sci. USA* **100**, 15 324–15 328. (doi:10.1073/pnas.0306899100)
- Mumford, J. A. 1999 The equine influenza surveillance program. *Adv. Vet. Med.* **41**, 379–387. (doi:10.1016/S0065-3519(99)80028-7)
- Mumford, J. A., Wood, J. M., Scott, A. M., Folkers, C. & Schild, G. C. 1983 Studies with inactivated equine influenza vaccine. 2. Protection against experimental infection with influenza A/equine/Newmarket/79 (H3N8). *J. Hyg.* **90**, 385–395.



- Newton, J. R., Townsend, H. G. G., Wood, J. L. N., Sinclair, R., Hannant, D. & Mumford, J. A. 2000a Immunity to equine influenza: relationship of vaccine-induced antibody in young thoroughbred racehorses to protection against field infection with influenza A/equine-2 viruses (H3N8). *Equine Vet. J.* **32**, 65–74. (doi:10.2746/042516400777612116)
- Newton, J. R., Lakhani, K. H., Wood, J. L. N. & Baker, D. J. 2000b Risk factors for equine influenza serum antibody titres in young thoroughbred racehorses given an inactivated vaccine. *Prev. Vet. Med.* **46**, 129–141. (doi:10.1016/S0167-5877(00)00144-6)
- Newton, J. R., Daly, J., Spencer, L. & Mumford, J. A. 2006 Description of the outbreak of equine influenza (H3N8) in the United Kingdom in 2003, during which recently vaccinated horses in Newmarket developed respiratory disease. *Vet. Rec.* **158**, 185–192.
- Park, A. W., Wood, J. L. N., Newton, J. R., Daly, J., Mumford, J. A. & Grenfell, B. T. 2003 Optimising vaccination strategies in equine influenza. *Vaccine* **21**, 2862–2870. (doi:10.1016/S0264-410X(03)00156-7)
- Park, A. W., Wood, J. L. N., Daly, J., Newton, J. R., Glass, K., Henley, W., Mumford, J. A. & Grenfell, B. T. 2004 The effects of strain heterology on the epidemiology of equine influenza in a vaccinated population. *Proc. R. Soc. Lond. B* **271**, 1547–1555. (doi:10.1098/rspb.2004.2766)
- Potter, C. W. (ed) 2002 *Influenza*. Perspectives in Medical Virology 7, pp. 49–85. Amsterdam, The Netherlands: Elsevier.
- Powell, D. G. (ed.) 1988 Equine infectious diseases V. In *Proc. 5th Int. Conf.* Lexington, KY: University Press of Kentucky.
- Renshaw, E. 1991 *Modelling biological populations in space and time*. Cambridge, UK: Cambridge University Press.
- Savill, N. J., St Rose, S. G., Keeling, M. J. & Woolhouse, M. E. J. 2006 Silent spread of H5N1 in vaccinated poultry. *Nature* **442**, 757. (doi:10.1038/442757a)
- Smith, D. J., Lapedes, A. S., de Jong, J. C., Bestebroer, T. M., Rimmelzwaan, G. F., Ostertaus, A. D. & Fouchier, R. A. 2004 Mapping the antigenic and genetic evolution of influenza virus. *Science* **305**, 371–376. (doi:10.1126/science.1097211)
- Turkington, C. & Ashby, B. 1998 *Encyclopaedia of infectious diseases*. New York, NY: Facts on File.
- van der Goot, J. A., Koch, G., de Jong, M. C. M. & van Boven, M. 2005 Quantification of the effect of vaccination on transmission of avian influenza (H7N7) in chickens. *Proc. Natl Acad. Sci. USA* **102**, 18 141–18 146. (doi:10.1073/pnas.0505098102)
- Wood, J. L., Newton, J. R., Chanter, N. & Mumford, J. A. 2005 Inflammatory airway disease, nasal discharge and respiratory infections in young British racehorses. *Equine Vet. J.* **37**, 236–242. (doi:10.2746/0425164054530579)