Supporting Text S1: Characterizing the transmission potential of zoonotic infections from minor outbreaks

Adam J. Kucharski* and W. John Edmunds
Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London WC1E 7HT, United Kingdom

*Corresponding author. Tel: +44 (0)20 7927 2407. E-mail: adam.kucharski@lshtm.ac.uk.

S1 Model derivation

S1.1 Single-type branching process

Suppose we have a branching process with offspring generating function $g(s)$, where

$$g(s) = \sum_{i=0}^{\infty} p_i s^i \tag{S1}$$

and p_i is the probability that an infectious individual generates i secondary cases. If we define X to be the total outbreak size (i.e. the sum of infectious individuals across all generations), the outbreak size distribution is as follows [1],

$$\mathbb{P}(X = n \mid x_0 = a) = \frac{a}{n} p_{n-a}^{(n)} \tag{S2}$$

where $p_{n-a}^{(n)}$ are the coefficients in the n^{th} power of the offspring distribution,

$$[g(s)]^n = p_0^{(n)} + p_1^{(n)} s + p_2^{(n)} s^2 + ... \tag{S3}$$

and hence

$$p_{n-a}^{(n)} = \frac{1}{(n-a)!} \frac{d^{n-a}}{ds^{n-a}} [g(s)]^n \bigg|_{s=0} . \tag{S4}$$

S1.2 Offspring distribution in two group model

The generating function for the offspring distribution of individual i is

$$h_i(s_1, s_2) = \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} p_{s_1,s_2} s_1^{j_1} s_2^{j_2} \tag{S5}$$

where p_{s_1,s_2} is the probability that an infectious individual of type i generates s_1 secondary cases of type 1 and s_2 cases of type 2.

As branching processes are independent, and stochasticity in transmission is represented by a Poisson process, the offspring distribution for an individual of type i is given by

$$h_i(s_1, s_2) = \int_0^\infty \int_0^\infty \exp\left[-\sum_{j=1}^{2} u_{ij}(1 - s_j)\right] f_{\nu_{i1}}(u_{i1}) f_{\nu_{i2}}(u_{i2})\, du_{i1}\, du_{i2} \tag{S6}$$

where $f(u_{ij})$ is the distribution of the reproduction number to i from j. Hence we can separate the p.g.f. as follows,

$$h_i(s_1, s_2) = \int_0^\infty \exp\left[-u_{i1}(1 - s_1)\right] f_{\nu_{i1}}(u_{i1})\, du_{i1} \int_0^\infty \exp\left[-u_{i2}(1 - s_2)\right] f_{\nu_{i2}}(u_{i2})\, du_{i2} \tag{S7}$$

$$= g_{1i}(s_1) g_{2i}(s_2). \tag{S8}$$

We can implement different assumptions about the offspring distribution by choosing different forms for $f(u_{ij})$. For example, if $f_{\nu_{ij}}(u_{ij}) = \delta(u_{ij} - R_{ij})$, where $\delta(x - R_{ij})$ is the Dirac delta function centred at R_{ij}, the offspring distribution is Poisson. If $f(u_{ij})$ is exponentially distributed with mean R_{ij}, the offspring distribution will be geometric (this is equivalent to assuming exponentially distributed recovery times, as in the standard SIR model). We can also impose a negative binomial offspring distribution by having $f(u_{ij})$ be gamma distributed with mean R_{ij} and dispersion parameter k [2]. Hence we have:

$$h_i(s_1, s_2) = \left(1 + \frac{R_{1i}}{k}(1 - s_1)\right)^{-k} \left(1 + \frac{R_{2i}}{k}(1 - s_2)\right)^{-k}. \tag{S9}$$

In this paper, we limit our attention to the geometric distribution (i.e. $k = 1$).

S1.3 Outbreak size distribution

Extending the calculation of Equation S4 for a multi-type model, we obtain the probability that n infectives of type i generate z_1 cases of type 1 and z_2 cases of type 2:

$$h_i^{(n)}(z_1, z_2) = \frac{1}{z_1!\; z_2!} \frac{\partial^{z_1} \partial^{z_2}}{\partial s_1^{z_1} \partial s_2^{z_2}} [h_i(s_1, s_2)]^n \bigg|_{s_1 = s_2 = 0}. \tag{S10}$$

To calculate the outbreak size distribution, we first consider the outbreak size distribution in two simple situations. First, let $R_{21} = R_{22} = 0$. The probability an outbreak starting with a_0 cases in group 1 will result in a total of n_1 cases in group 1 and n_2 cases in group two is [3],

$$\mathbb{P}(X_1 = n_1, X_2 = n_2 \mid x_0 = a_0) = \frac{a_0}{n_1} h_1^{(n_1)}(n_1 - a_0, n_2) \tag{S11}$$

Second, let $n_2 > 0$ and $R_{12} = 0$. The joint outbreak size distribution in this case is [3],

$$\mathbb{P}(X_1 = n_1, X_2 = n_2 \mid x_0 = a_0) = \frac{a_0}{n_1} \sum_{a_{21}=0}^{n_2} \frac{a_{21}}{n_2} h_1^{(n_1)}(n_1 - a_0, a_{21}) h_2^{(n_2)}(0, n_2 - a_{21}) \tag{S12}$$

Finally, we specify the joint distribution for a non-negative next generation matrix. In the two group model with single introduction from an external reservoir, with $n_1 > 0$, $n_2 > 0$ and $R_{ij} \geq 0$ for all i and j, we have [4]

$$\mathbb{P}(X_1 = n_1, X_2 = n_2 \mid x_0 = 1) = \frac{1}{n_1} \sum_{a_{12}=0}^{n_1-1} \sum_{a_{21}=0}^{n_2} \frac{a_{21}}{n_2} h_1^{(n_1)}(n_1 - 1 - a_{12}, a_{21}) h_2^{(n_2)}(a_{12}, n_2 - a_{21}) \, . \tag{S13}$$

Note that if $n_2 = 0$, we instead use Equation S11 with $a_0 = 1$.

S2 Incorporating pathogen introduction

S2.1 Probability of introduction into each age group

Here we calculate the probability the outbreak starts a particular age group, assuming that the infection is introduced randomly across the entire susceptible population. Let P_2 be the proportion of the population over age 20, and S be proportion of this group susceptible to infection. If σ_1 is the probability the outbreak starts in the under 20 group, and σ_2 is the probability it starts in the over 20s,

$$\sigma_1 = \frac{(1 - P_2)}{(1 - P_2) + SP_2} \quad \text{and} \quad \sigma_2 = \frac{SP_2}{(1 - P_2) + SP_2} \, . \tag{S1}$$

S2.2 Inference when introduction probabilities are equal to dominant eigenvector of next generation matrix

In a homogeneously mixing population with $R < 1$, the mean outbreak size, $\bar{\mu}$, is related to R as follows [5]:

$$\bar{\mu} = \sum_{k=0}^{\infty} R^k = \frac{1}{1 - R} \tag{S2}$$

In an age-structured population, is also possible to obtain unbiased estimates of R using the overall mean outbreak size as long as the probabilities of introduction in each age group are equal to dominant eigenvector of the next generation matrix, $\mathbf{R}$. To prove this, we first define the overall mean outbreak size, μ, in the age-structured population to be

$$\mu = \mathbf{1}^{\top}.\boldsymbol{\mu}.\mathbf{1} \tag{S3}$$

where $\mathbf{1} = \binom{1}{1}$ and $\boldsymbol{\mu}$ is as in Equation 3 in the main text. Next we define $\mathbf{v} = \binom{v_1}{v_2}$ to be the dominant eigenvector of the matrix $\mathbf{R}$. If the probabilities that the infection will be introduced to each age group are equal to the entries in this vector, by definition we have

$$\mathbf{A}.\mathbf{1} = \mathbf{v} \iff \mathbf{R}\mathbf{A}.\mathbf{1} = R\mathbf{A}.\mathbf{1}\ . \tag{S4}$$

As the entries of $\mathbf{A}$ sum to one, we also have $\mathbf{1}^{\top}\mathbf{A}.\mathbf{1} = 1$.

Using the above relationships, we can show that outbreak size in the heterogeneous model is equal to the outbreak size in the homogeneous model:

$$\begin{aligned} \mu = \mathbf{1}^{\top}(\mathbf{I}-\mathbf{R})^{-1}\mathbf{A}.\mathbf{1} = \mathbf{1}^{\top}\sum_{k=0}^{\infty}\mathbf{R}^{k}\mathbf{A}.\mathbf{1} = \mathbf{1}^{\top}\sum_{k=0}^{\infty}R^{k}\mathbf{A}.\mathbf{1} \\ = \mathbf{1}^{\top}\mathbf{A}.\mathbf{1}\sum_{k=0}^{\infty}R^{k} = \mathbf{1}^{\top}\mathbf{A}.\mathbf{1}\bar{\mu} = \bar{\mu}\ . \end{aligned} \tag{S5}$$

S3 Accounting for censoring during real-time analysis

Outbreak size analysis can also provide information about supercritical infections, with $R > 1$ [6]. However, some introductions will lead to large epidemics when $R > 1$. Making inferences using data on finite outbreak sizes is therefore equivalent to conditioning on outbreaks having gone extinct, which means the distribution of observed outbreak sizes is improper [7]. It is therefore not possible to distinguish $R > 1$ and $R < 1$. One way to remove the condition on extinction is to include observations that are censored [7]. An outbreak of size X that is still in progress at the time of observation could eventually have any size in the set $\{X, X+1, ...\}\bigcup\{\infty\}$. As this includes the possibility that the infection will go on to cause a major epidemic, R no longer has an improper distribution.

Let $c = 1$ if an outbreak k is censored. Recall that $r^{1}_{n,m}$ is the probability that an infection that starts in group 1 results in n cases in group 1 and m cases in group 2. If $c = 0$, then our likelihood function is unchanged. If $c = 1$, the likelihood function is:

$$L(n, m \mid c = 1) = 1 - \sum_{j_1=1}^{\tilde{n}}\sum_{j_2=0}^{\tilde{m}} r^{1}_{j_1,j_2} \tag{S1}$$

where $\tilde{n} = \max\{1, n-1\}$ and $\tilde{m} = \max\{0, m-1\}$. We can also write an equivalent expression for an outbreak that starts in group 2.

For the MERS-CoV data, we used the serial interval of infection to specify whether a particular outbreak was censored. It has been suggested that the MERS serial interval is 7.6 days (95% CI 2.5–23.1 days) [8]. We therefore assumed that any outbreak cluster containing a case with onset date less than 23 days from the last date in our list was censored.

S4 Construction of confidence intervals

To find the 95% confidence interval for a given parameter (such as R_0 or S) we constructed a likelihood profile. For each value the parameter could take, we found the maximum likelihood estimate across all possible values of the other parameter. The 95% confidence interval was equivalent to the region of parameter space that was within 1.92 log-likelihood points of the maximum-likelihood estimate for both parameters [9].

S5 Estimating pre-existing immunity

S5.1 Monkeypox

As the smallpox vaccine also provided cross-immunity against the monkeypox virus, the presence of smallpox vaccine scars has been used to measure pre-existing immunity to monkeypox [10]. In 1978–79, a survey of villages where monkeypox cases occurred found vaccine scars in 44% of children under age 4 and 77% of individuals in the 5–14 and over 15 age groups [10]. We used demographic data from the 1980 census of the Democratic Republic of Congo to merge these categories into two age groups: under and over 20 years old. Based on the age distribution of the population, we estimated that 66% of individuals under age 20 and 77% over age 20 in these villages would have had vaccine scars, which implies that susceptibility in the over 20 age group was reduced by a factor (1-0.77)/(1-0.66)=0.68.

A survey of the villages surrounding those in which cases were identified found scars in 60% of children under age 4, 91% of individuals in the 5–14 age group and 92% of individuals over age 15 [10]. Again using the population age distribution, we estimated that 81% of individuals under age 20 and 92% over age 20 in villages would have had vaccine scars, suggesting a reduction factor of (1-0.92)/(1-0.81)=0.42. Hence it is plausible that S was between 0.4–0.7 in the 1970–79 monkeypox outbreaks.

S5.2 Influenza A(H5N1)

It has been suggested that populations might have pre-existing to influenza A(H5N1) as a result of cross-immunity from previous infection with H1N1, which shares the neuraminidase protein, N1 [11]. The proportion of individuals aged a with pre-existing immunity can therefore be expressed as $1 - e^{-a\Lambda}$, where Λ is the average annual probability of gaining cross-immunity to H5N1 as a result of H1N1 infection. An analysis of influenza A(H5N1) cases in Egypt and Southeast Asia using this model of cross-immunity estimated that Λ= 0.036 (95% CI 0.029-0.043) [11]. Combining this model with demographic data for Indonesia, we estimated that a proportion 0.22 (0.16–0.28) of the under 20 age group would have had immunity to H5N1, and 0.60 (0.54–0.65) of the over 20 group. The relative susceptibility of the two groups was therefore S=0.36 (0.30–0.44).

References

[1] Dwass M (1969) The total progeny in a branching process and a related random walk. Journal of Applied Probability 6: 682–686.

[2] Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM (2005) Superspreading and the effect of individual variation on disease emergence. Nature 438: 355-9.

[3] Bertoin J (2009) The structure of the allelic partition of the total population for galton-watson processes with neutral mutations. The Annals of Probability : 1502–1523.

[4] Chaumont L, Liu R (2014) Coding multitype forests: application to the law of the total population of branching forests. Transactions of the American Mathematical Society (in press).

[5] De Serres G, Gay N, Farrington C (2000) Epidemiology of transmissible diseases after elimination. American Journal of Epidemiology 151: 1039–1048.

[6] Blumberg S, Lloyd-Smith JO (2013) Inference of R_0 and transmission heterogeneity from the size distribution of stuttering chains. PLoS Comput Biol 9: e1002993.

[7] Farrington C, Kanaan M, Gay N (2003) Branching process models for surveillance of infectious diseases controlled by mass vaccination. Biostatistics 4: 279.

[8] Assiri A, McGeer A, Perl TM, Price CS, Al Rabeeah AA, et al. (2013) Hospital outbreak of middle east respiratory syndrome coronavirus. N Engl J Med 369: 407-16.

[9] Burnham K, Anderson D (2002) Model selection and multimodel inference: a practical information-theoretic approach. Springer-Verlag, New York, second edition.

[10] Breman J, Kalisa-Ruti M, Zanotto E, Gromyko A, Arita I (1980) Human monkeypox, 1970-79. Bulletin of the World Health Organization 58: 165.

[11] Kucharski AJ, Edmunds WJ (2014) Cross-immunity and age patterns of influenza A(H5N1) infection. Epidemiology and Infection .