THE LANCET

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Breban R, Riou J, Fontanet A. Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk. *Lancet* 2013; published online July 5. http://dx.doi.org/10.1016/S0140-6736(13)61492-0.

Inter-human transmissibility of MERS-CoV: estimation of pandemic risk

WebAppendix

Romulus Breban[1], Julien Riou[1], Arnaud Fontanet[1,2]

[1] *Institut Pasteur, Emerging Diseases Epidemiology Unit, Paris, France*
[2] *Conservatoire National des Arts et Métiers, Paris, France*

Contents

1 The choice of distribution for the number of secondary cases

Recent applications of homogeneous branching process theory to the epidemiology of emerging diseases have used a negative binomial distribution to model the number of secondary cases of an infected individual [1]. This modeling choice has several important benefits such as:

1. The two parameters of the distribution allow for an independent description of the intensity and heterogeneity of transmission through the distribution average, represented by the basic reproduction number R_0, and variance determined by R_0 and the dispersion parameter k. This provides an elegant modeling framework for transmission with superspreading events;

2. The negative binomial distribution reduces to the Poisson and geometric distributions for various particular values of the dispersion parameter k.

However, applying the likelihood model based on the negative binomial distribution [1] to MERS-CoV tree size data, we found that the dispersion parameter k was *not* resolved by the data and could be very large (even as large as 1000). Thus, applying Ocam's razor, we considered k to be essentially infinite, which parsimoniously reduced the negative binomial distribution to the Poisson distribution. Furthermore, we note that, in analogy to the case of pre-pandemic SARS, superspreading events may not be common in the transmission of MERS-CoV. Hence, we believe that our choice of a Poisson distribution for the number of secondary cases is appropriate for the MERS-CoV dataset that we used.

It should further be noted that R_0 results obtained from tree sizes are quite robust to the choice of the distribution of secondary cases. The average tree size can be written in terms of R_0 as

$$1 + R_0 + R_0^2 + R_0^3 + \ldots, \tag{1}$$

independently of the choice of the distribution. In turn, this result has major consequences for estimating R_0 from tree sizes. Reference [1] shows that a geometric, Poisson or negative binomial distribution of secondary cases yields the same (i.e., independent of the dispersion parameter k) maximum likelihood estimation of R_0 based on tree sizes.

2 Bayesian methodology for estimating R_0

We analysed MERS-CoV transmission using the theory of homogeneous branching processes. The key element of this theory is the distribution of cases caused by an infected individual. Its average is the basic reproduction number of the pathogen R_0 and plays a fundamental role for the transmission dynamics. If $R_0 < 1$, then all transmission trees terminate, otherwise transmission trees may be infinite and the disease becomes an epidemic. Assuming that the distribution of secondary cases is Poisson (see Sec. 1), we inferred the R_0 of MERS-CoV using Bayesian analysis. We used a flat, non-informative prior from 0 to an arbitrary large constant M

$$\pi(R_0) = 1/M, \tag{2}$$

where later, in the formula of the posterior distribution, we take the limit $M \to \infty$. The likelihood was constructed as follows. We first calculated the probability that a branching tree has size n, given that the distribution of secondary cases is Poisson with average R_0 [1]

$$p(n, R_0) = \frac{(R_0 n)^{n-1} \exp(-R_0 n)}{n!}. \tag{3}$$

Then, the likelihood of observing a set of trees $\mathbb{T}$ can be written as [1]

$$\mathcal{L}(\mathbb{T}|R_0) = \prod_{n=1}^{\infty} [p(n, R_0)]^{s_n} \tag{4}$$

where s_n is the number of observations of trees of size n in the dataset $\mathbb{T}$. Finally, we obtained the posterior distribution for R_0, combining the prior distribution $\pi(R_0)$ and the likelihood $\mathcal{L}(\mathbb{T}|R_0)$ according to Bayes rule

$$\hat{\pi}(R_0|\mathbb{T}) = \frac{\mathcal{L}(\mathbb{T}|R_0)\pi(R_0)}{\int_0^{\infty} \mathcal{L}(\mathbb{T}|R_0)\pi(R_0)dR_0} \xrightarrow{M\to\infty} \frac{\mathcal{L}(\mathbb{T}|R_0)}{\int_0^{\infty} \mathcal{L}(\mathbb{T}|R_0)dR_0}. \tag{5}$$

It is important to note that, according to the above equation, the maximum likelihood and most probable Bayesian estimates of R_0 are identical.

The posterior distribution $\hat{\pi}(R_0|\mathbb{T})$ allows for the calculation of the expected R_0 value and its corresponding confidence interval, according to the dataset $\mathbb{T}$. Of note, analysis of the MERS-CoV dataset yields that the maximum likelihood estimate of R_0 and expectation of the posterior R_0-distribution in the Bayesian analysis are very close (within 0.02), owing to the fact that the posterior distribution $\hat{\pi}(R_0|\mathbb{T})$ is unimodal and highly symmetric.

We further used Bayesian analysis to understand the impact of future MERS-CoV outbreaks and compute the probability that R_0 exceeds 1 as a function of the tree size which would be observed next. In particular, we performed the Bayesian analysis adding a hypothetical tree to the dataset $\mathbb{T}$ and then using the resulting posterior distribution to compute

$$\int_1^{\infty} \hat{\pi}(R_0|\mathbb{T}')dR_0, \tag{6}$$

for each hypothetical tree size, where $\mathbb{T}'$ denotes the extended dataset. This methodology can be adapted straightforwardly to the case where the hypothetical data represents a number of secondary cases by simply rewriting the likelihood as

$$\mathcal{L}'(\mathbb{T}'|R_0) = \mathcal{L}(\mathbb{T}|R_0)f \tag{7}$$

where f is the probability mass function of the Poisson distribution evaluated at the hypothetical number of secondary cases.

3 Sensitivity analyses

We ran two sets of sensitivity analysis. First, we analyzed the variation of R_0 with changing the size of the Jordanian cluster between 2 (the number of confirmed cases) and 13 (the total number of confirmed and probable cases). The change in the expected R_0 in both scenarios was less than 0.07, not significant when compared to the amplitudes of the corresponding 95% confidence intervals.

Second, we ran a sensitivity analysis based on the outstanding event of six secondary cases caused by patient C of the Al Hasa cluster in the dialysis ward [2]. Hence, removing this event, we re-evaluated R_0 under both scenarios. The main rational for these computations is that the outstanding transmission event might be unlikely to repeat. We found that the expected R_0 changed by very little (by less than 0.1 for scenario 1 and 0.01 for scenario 2), confirming the robustness of our results.

4 Estimation of the MERS-CoV introduction rates

The rate of MERS-CoV introductions into the human population was calculated from the estimated number of index patients over the period of data collection (see Table 1 in the main text). Corresponding confidence intervals were assigned assuming that introduction events are *rare events* and follow the Poisson statistics.

References

[1] Seth Blumberg and James O Lloyd-Smith. Inference of R0 and Transmission Heterogeneity from the Size Distribution of Stuttering Chains. *PLoS Computational Biology*, 9(5):e1002993, May 2013.

[2] Abdullah Assiri, Allison McGeer, Trish M Perl, Connie S Price, Abdullah A Al Rabeeah, Derek A T Cummings, Zaki N Alabdullatif, Maher Assad, Abdulmohsen Almulhim, Hatem Makhdoom, Hossam Madani, Rafat Alhakeem, Jaffar A Al-Tawfiq, Matthew Cotten, Simon J Watson, Paul Kellam, Alimuddin I Zumla, and Ziad A Memish. Hospital Outbreak of Middle East Respiratory Syndrome Coronavirus. *The New England Journal of Medicine*, June 2013.