Yes, it matters who is spreading monkeypox: Supplemental Information

Jonathan Smith, PhD, MPH

July 25, 2022

Contents

1 Background

We assume the number of secondary cases produced from each infectious case is distributed according to a negative binomial distribution with mean R_0 (the basic reproductive number) and dispersion parameter k. The parameter k quantifies the degree of individual heterogeneity by measuring overdispersion in the distribution (e.g., higher than expected variation). For a given R_0, smaller values of k ($k << 1$) suggest increased heterogeneity in the number of secondary cases between individual index cases. As k increases, transmission becomes more uniform and centered around the mean, R_0; as k approaches infinity the negative binomial distribution converges to the Poisson distribution, wherein all differences in the number of secondary cases is attributed to demographic stochasticity.

Code to implement the included methods and recreate figure is available at the following GitHub repository: https://github.com/jpsmithuga/Comment-Supplementary

2 Proportion of secondary cases attributed to the proportion of infectious cases

Lloyd-Smith *et al.* (in in supplementary information 2.2.5) propose a distribution for describing transmission based from the distribution of the individual reproduction number ν:

$$F_{\text{trans}}(x) = \frac{1}{R_0}\int_0^x u\, f_\nu(u)\, du$$

This gives the cumulative distribution function (CDF) in terms of the individual reproduction number density f_ν, as the proportion of all transmission due to infectious individuals with reproduction number $\nu < x$.
In the specific case when ν is gamma distributed with shape parameter $k > 0$ and rate parameter k/R_0, where R_0 is the mean (expected) individual reproduction number, we have

$$f_\nu(u) = \frac{k^k}{R_0^k\,\Gamma(k)} u^{k-1} e^{-ku/R_0}$$

It follows that

$$\begin{aligned} f_{\text{trans}}(x) = F'_{\text{trans}}(x) &= \frac{1}{R_0} x f_\nu(x) \\ &= \frac{k^k}{R_0^{k+1}\,\Gamma(k)} x^k e^{-kx/R_0} \\ &= \frac{k^{k+1}}{R_0^{k+1}\,\Gamma(k+1)} x^k e^{-kx/R_0} \\ &\sim \text{Gamma}(k+1, k/R_0) \end{aligned}$$

The rate parameter remains the same and the shape parameter increases by 1. To calculate t_p, the expected proportion of transmission due to the most infectious $100p\%$ of cases, we first find the $(1-p)$th centile of the individual reproduction number, $x_p = F_\nu^{-1}(1-p)$, then calculate $t_p = 1 - F_{\text{trans}}(x_p)$. As both random variables are gamma distributed, this is implemented in R via the qgamma (F^{-1}) and pgamma (F) functions.
The rate parameter remains the same and the shape parameter increases by 1. To calculate t_p, the expected proportion of transmission due to the most infectious $100p\%$ of cases, we first find the $(1-p)$th centile of the individual reproduction number, $x_p = F_\nu^{-1}(1-p)$, then calculate $t_p = 1 - F_{\text{trans}}(x_p)$. As both random variables are gamma distributed, this is implemented in R via the qgamma (F^{-1}) and pgamma (F) functions (see here).

3 The relationship between p and t_p

Parameterizing by x_p, and noting that $p = 1 - F_\nu(x_p)$, the first derivative is

$$\frac{dt_p}{dp} = \frac{dt_p/dx_p}{dp/dx_p} = \frac{-F'_{\text{trans}}(x_p)}{-F'_\nu(x_p)} = \frac{R_0^{-1} x_p f_\nu(x_p)}{f_\nu(x_p)} = \frac{x_p}{R_0} > 0$$

which confirms that t_p is an increasing function of p. When $x_p = 0$, $(p, t_p) = (1, 1)$ at which the point the derivative is zero (horizontal tangent). Similarly, when $x_p = +\infty$, $(p, t_p) = (0, 0)$ at which the point the derivative is infinite (vertical tangent).
The second derivative is negative implying concavity of the graph of t_p as a function of p.

$$\frac{d^2 t_p}{dp^2} = \frac{d}{dx_p}\left(\frac{dt_p}{dp}\right) \times \frac{dx_p}{dp} = \frac{1}{R_0} \times \frac{1}{dp/dx_p} = -\frac{1}{R_0 f_\nu(x_p)} < 0$$

These properties can be seen in Figure 1A.

4 Final Outbreak Size

Let Y represent the sum of all cases in a transmission chain (including the index case), resulting in a final outbreak size (Dwass, 1969). Using branching process theory, the distribution of the total number of secondary infections in a single transmission chain (including the index case at generation) can be defined by the probability generating function (pgf) yielding the implicit relationship (Yan 2008):

$$G_Y(s) = sG_Z((G_Y(s))$$

We can treat this as $G_Y(s) = \sum_{y=1}^{\infty} q_y s^y$, where $q_y = P(Y = y)$, allowing us to extract the probability (Becker 1974; Blumberg 2014):

$$P(Y = y) = \frac{1}{y!} \frac{d_y G_Y(s)}{ds^y}\bigg|_{s=0}$$

Where $Y = 1, 2, 3 \ldots$
Plugging in the negative binomial distribution yields (see Smith 2022 ([supplementary information, section 4]()) and Blumberg 2012)

$$P(Y = y) = \binom{n}{y} \frac{\Gamma(ky + y - n)}{\Gamma(ky)(y - n)!} \frac{(\frac{R_0}{k})^{y-n}}{(1 + \frac{R_0}{k})^{ky+y-n}}$$

Where n represents the number of index cases initiating the outbreak and $Y = 1, 2, 3 \ldots$.
The this is implemented in R software by exponentiating the log-likelihood (for computational ease; see here)

5 Acknowledgements

Special thanks to Dr. Andrew Hill for providing mathematical oversight into this analysis.

6 References

1. Lloyd-Smith, JO, et al. Superspreading and the effect of individual variation on disease emergence, Nature 438, 355–359 (2005)
2. Yan, P. Distribution Theory, Stochastic Processes and Infectious Disease Modelling, pages 229. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
3. Becker, N. On parametric estimation for mortal branching processes. Biometrika, 61(3):393, 1974.
4. Dwass, M. The total progeny in a branching process and a related random walk. Journal of Applied Probability, 6(3):682, 1969.
5. Blumberg S and Lloyd-Smith JO. Inference of r0 and transmission heterogeneity from the size distribution of stuttering chains. PLoS Comput Biol., 5(9):393, 2013.
6. Smith, JP, *et al*. A cluster-based method to quantify individual heterogeneity in tuberculosis transmission. Epidemiology 33, 217–227. 2022.