

# The role of superspreading in Middle East respiratory syndrome coronavirus (MERS-CoV) transmission

A J Kucharski<sup>1</sup>, C L Althaus (christian.althaus@alumni.ethz.ch)<sup>2</sup>

1. Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom
2. Institute of Social and Preventive Medicine (ISPM), University of Bern, Berne, Switzerland

## Citation style for this article:

Kucharski AJ, Althaus CL. The role of superspreading in Middle East respiratory syndrome coronavirus (MERS-CoV) transmission. *Euro Surveill.* 2015;20(25):pii=21167. Available online: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=21167>

Article submitted on 15 June 2015 / published on 25 June 2015

As at 15 June 2015, a large transmission cluster of Middle East respiratory syndrome coronavirus (MERS-CoV) was ongoing in South Korea. To examine the potential for such events, we estimated the level of heterogeneity in MERS-CoV transmission by analysing data on cluster size distributions. We found substantial potential for superspreading; even though it is likely that  $R_0 < 1$  overall, our analysis indicates that cluster sizes of over 150 cases are not unexpected for MERS-CoV infection.

## MERS-CoV transmission

There have been 1,288 cases of Middle East respiratory syndrome (MERS) reported worldwide as at 15 June 2015 [1]. Many of these have been index cases, likely to have been infected from an animal reservoir, but there have also been several clusters of human-to-human transmission. An imported MERS case with a travel history to the Arabian Peninsula resulted in a new cluster in South Korea, with 150 cases reported as at 15 June 2015 [2]. This raises two important questions about the transmission dynamics of MERS coronavirus (MERS-CoV). First, how much heterogeneity is there in MERS-CoV transmission in the absence of animal-human infection? Second, given such heterogeneity, what are the chances of observing an outbreak as large as the one in South Korea?

The dynamics of an outbreak depend on both  $R_0$  – the average number of secondary cases generated by a typical infectious individual – and individual heterogeneity in transmission. Such heterogeneity can be estimated by describing the distribution of secondary cases as a negative binomial distribution with dispersion parameter  $k$ , where  $k < 1$  suggests that transmission is overdispersed, and hence outbreaks can include superspreading events [3,4]. However, there is currently no measure of transmission heterogeneity for MERS-CoV. Using reported outbreak data, we examined the extent of individual variation in MERS-CoV

transmission, and estimated the probability of observing clusters as large as the one in South Korea.

## Analysing cluster data

We analysed data on MERS cluster sizes for cases reported up to 31 August 2013 [5]. For comparison, we also considered data from two other reports, up to 21 June 2013 [6] and 8 August 2013 [7]. Cases with known epidemiological links were classified as a cluster. Single index cases were considered as independent clusters of size one. Although more cases have since been reported [1], it is not entirely clear how many clusters there have been. We therefore chose to focus on published cluster data (Table), which also made it possible to compare our results with previous analyses.

To estimate  $R_0$  and  $k$  from the distribution of cluster sizes, we used a likelihood-based inference method based on branching processes with the offspring distribution following a negative binomial distribution with mean  $R_0$  and dispersion parameter  $k$ . This distribution is widely used to describe overdispersed count data in biology and epidemiology [4], and has the useful property that Poisson ( $k = \infty$ ) and geometric offspring distributions ( $k = 1$ ) are special cases of it. The probability that an index case generates a cluster of size  $j$  is [8,9]:

$$r_j = \frac{\Gamma(kj + j - 1)}{\Gamma(kj)\Gamma(j + 1)} \frac{\left(\frac{R_0}{k}\right)^{j-1}}{\left(1 + \frac{R_0}{k}\right)^{kj+j-1}}$$

Therefore the likelihood of observing  $n_j$  clusters of size  $j$  is:

$$L = \prod_{j=1}^{\infty} r_j^{n_j}$$

For given values of  $R_0$  and  $k$ , the probability that an index case generates a transmission cluster of size  $j$  or greater is:

$$p_j = 1 - \sum_{i=1}^{j-1} r_i.$$

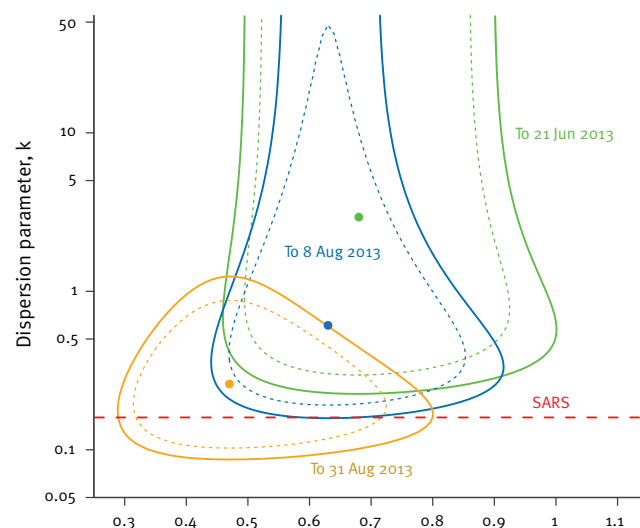
Assuming  $N$  introductions of infections into the human population, the probability that at least one cluster of size  $j$  or greater occurs is  $1 - (1 - p_j)^N$ . All analyses were done in the R software environment for statistical computing [10].

### Findings\*

Using available cluster data, we jointly estimated  $R_0$  and the dispersion parameter  $k$  for MERS-CoV (Figure 1). Analysis of severe acute respiratory syndrome (SARS) coronavirus transmission during the early stages of the outbreak in Singapore suggested  $k = 0.16$  (90% confidence interval (CI): 0.11–0.64) [3] (the study cited 90% CI owing to the paucity of available data). Our estimate for MERS-CoV is similar, with  $k = 0.26$  (90% CI: 0.11–0.87, 95% CI: 0.09–1.24). As it is not always clear from case reports which cases are epidemiologically linked, we also estimated  $k$  using data from two other studies of clusters [6,7]. These data included fewer clusters

**FIGURE 1**

Joint estimates of basic reproduction number,  $R_0$ , and dispersion parameter,  $k$ , for MERS-CoV\*



MERS-CoV: Middle East respiratory syndrome coronavirus; SARS: severe acute respiratory syndrome.

Estimates are based on reported cluster size distributions until 31 August 2013 (orange), 8 August 2013 (blue) and 21 June 2013 (green) reported by Poletto et al. [5], Cauchemez et al. [7] and Breban et al. [6], respectively. Points indicate maximum likelihood estimates and lines show 90% (dashed) and 95% (solid) confidence intervals. The red dashed line indicates the dispersion parameter  $k = 0.16$  that was reported for SARS coronavirus [3].

**TABLE**

Data sources used for MERS clusters<sup>a</sup> of a given size (including index case), based on laboratory-confirmed MERS case reports worldwide<sup>b</sup>\*

Cluster size	Number of MERS clusters of a given size		
	Breban et al. [6] <sup>c</sup>	Cauchemez et al. [7]	Poletto et al. [5] <sup>c</sup>
1	11	27	42
2	2	2	7
3	3	4	2
4	1	3	–
5	2	2	2
7	–	1	–
10	–	–	1
13	–	1	–
22	–	–	1
24	1	–	–
26	–	1	–

MERS: Middle East respiratory syndrome.

Dashes indicate that there were no such reports.

- <sup>a</sup> Cases with known epidemiological links were classified as a cluster. Single index cases were considered as independent clusters of size one.
- <sup>b</sup> We analysed data on MERS cluster sizes for cases reported up to 31 August 2013 [5]. For comparison, we also considered data from two other reports, up to 21 June 2013 [6] and 8 August 2013 [7].
- <sup>c</sup> These studies listed more than one set of possible clusters, depending on how cases were interpreted. We therefore considered data from the most pessimistic scenario in each study, which included the probable cases in the Jordan outbreak in April 2012.

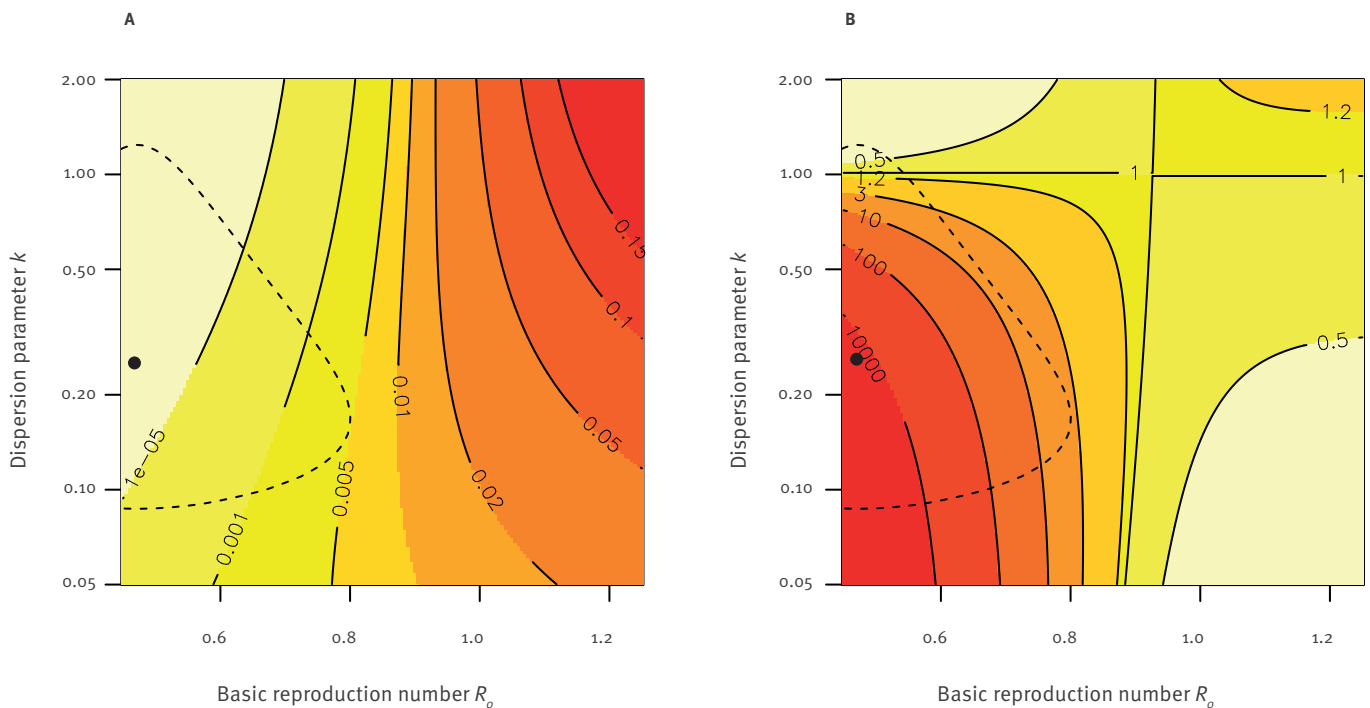
and were less conclusive regarding the amount of over-dispersion, with  $k = 0.61$  (95% CI: 0.16–∞) [7] and  $k = 2.94$  (95% CI: 0.23–∞) [6].

Our estimate for  $R_0$  was 0.47 (95% CI: 0.29–0.80). The maximum likelihood estimate (0.47), which is independent of  $k$  [8], agrees with previous work [5-7]. However unlike earlier studies, which assumed the distribution of secondary cases to be either geometric (i.e.  $k = 1$ ) [5,7] or Poisson ( $k = ∞$ ) [6], our upper 95% CI is larger. This is because allowing for potential over-dispersion increases the uncertainty surrounding the estimate of  $R_0$  (Figure 1).

There is an intricate relationship between the basic reproduction number,  $R_0$ , the dispersion parameter,  $k$ , and the probability of observing a large transmission cluster (Figure 2A). For a given value of  $k$ , increasing  $R_0$  also increases the probability of observing large clusters. If  $R_0$  is low, a higher variation in the number of secondary cases (i.e. smaller  $k$ ) increases the probability of observing large transmission clusters owing to the potential for superspreading. The effect of  $k$  is reversed for values of  $R_0$  near one, where a smaller  $k$  reduces the probability of observing large clusters. This is because a higher variation in the number of secondary cases increases the probability that an infected index case does not generate further cases [3]. Interestingly, the

**FIGURE 2**

Relationship between the basic reproduction number,  $R_0$ , the dispersion parameter,  $k$ , and the probability that a transmission cluster reaches at least 150 cases\*



MERS-CoV: Middle East respiratory syndrome coronavirus.

Panel A: Probability that a single index case generates a transmission cluster of 150 cases or greater. Panel B: Relative risk of seeing a cluster of at least 150 cases, compared with the scenario where  $k = 1$  (geometric distribution of secondary cases). The points indicate maximum likelihood estimates of  $R_0$  and  $k$  for MERS-CoV, and the dashed lines show the 95% confidence intervals.

area where the effect of overdispersion for a given value of  $R_0$  switches from increasing to decreasing the probability to observe large cluster sizes lies near the maximum likelihood estimate for MERS-CoV (Figure 2B).

Finally, we calculated the expected probability of observing a MERS-CoV transmission cluster of a given size or greater, by integrating across the full parameter distribution in Figure 1. Using the estimated distribution of  $k$  substantially increases the probability that index cases generate large clusters (Figure 3A), compared with the situation in which the number of secondary cases are assumed to be geometrically distributed ( $k = 1$ ). The probability that a single index case infected with MERS-CoV results in a cluster of 150 cases or more – as observed in South Korea – is 0.04%. Assuming different numbers of MERS-CoV introductions into human populations, the probabilities that at least one such outbreak occurs are 2.5% (100 introductions), 5.6% (500 introductions), 7.4% (1,000 introductions) and 9.3% (2,000 introductions).

## Discussion

Our results suggest that MERS-CoV transmission is highly overdispersed, and hence there is substantial potential for superspreading events. This finding is corroborated by a similar analysis of MERS-CoV outbreak

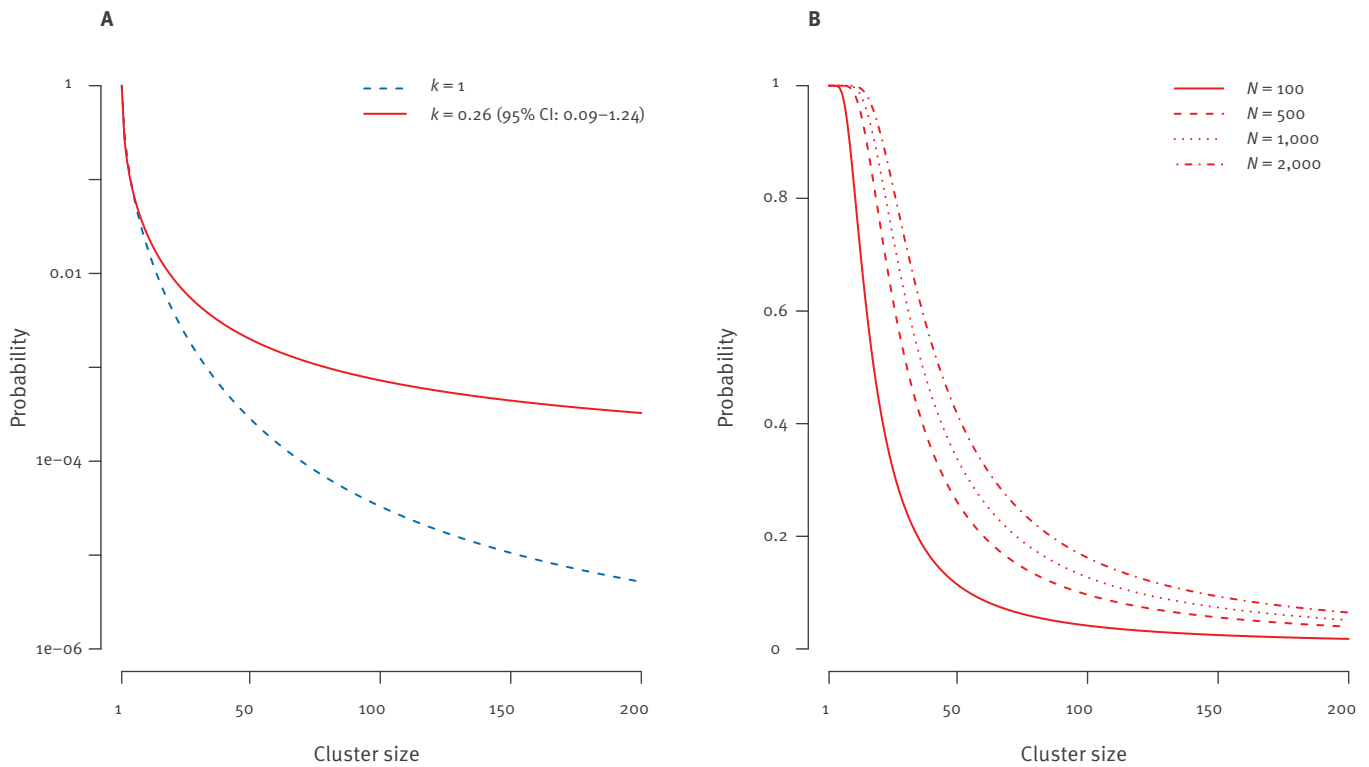
size distributions [11]. Given that hundreds of MERS-CoV index cases have been reported to date, our analysis indicates that occasional cluster sizes of over 150 cases – such as the one in South Korea – should not be unexpected. We also found a non-linear relationship between the basic reproduction number,  $R_0$ , dispersion parameter,  $k$ , and outbreak size: when  $R_0 < 0.9$ , the probability of obtaining a large cluster increases as the process becomes more overdispersed; as  $R_0$  approaches one, the effect is reversed and a higher level of overdispersion reduces the chances of a large cluster for a given value of  $R_0$ .

There are some limitations to our study. Case data may be subject to bias or under-reporting. However, such factors will generally drive up estimates of overdispersion [4] and hence are unlikely to alter our overall conclusions. It can also be difficult to conclusively identify outbreak clusters from case data. We therefore considered three different data sources, and found evidence of overdispersion in the two largest and most recent data sets.

Other infections, including SARS [3] and Ebola virus disease [12], also exhibit overdispersed transmission patterns. However, it can be difficult to establish precisely which factors drive superspreading events. For MERS-CoV, the observed overdispersion may result

### FIGURE 3

Probability that a MERS-CoV transmission cluster of a given size or greater occurs\*



MERS-CoV: Middle East respiratory syndrome coronavirus;  $N$ : introductions of infections into the human population.

Panel A: Probability that a single index case generates a transmission cluster of a given size or greater. Panel B: Probability that at least one transmission cluster of a given size or greater occurs assuming different numbers of MERS-CoV introductions into human populations. We assumed  $R_0 = 0.47$  (0.29–0.80). Note the logarithmic and linear scale of the horizontal axis in panels A and B.

from a combination of factors, including individual viral shedding and contact rates, hospital procedures and location, as well as population structure and density [13]. Even if such factors cannot be disentangled, measuring the overall extent of overdispersion – as we have done here – can help with the interpretation of surveillance data, and enable more realistic analysis of disease transmission and control [14].

#### \* Authors' correction

A typo in the code that was used for the analysis resulted in erroneous estimates for the dispersion parameter  $k$  and the confidence intervals surrounding the basic reproduction number  $R_0$ . All numbers in the text and the figures have been updated using the corrected estimates of  $k$ . Furthermore, figures have been updated to include parameter estimates derived from the largest data set of cluster sizes as reported by Poletto et al. [5]. The study now refers to a more recent analysis of MERS-CoV outbreak size distributions that showed very similar results [11]. These changes were made on 10 August 2015, at the request of the authors.

The authors have made the code available on GitHub (<https://github.com/calthaus/MERS>) to ensure reproducibility of the analysis.

#### Acknowledgements

AK is supported by the Medical Research Council (MR/K021524/1) and the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health. CLA received funding through an Ambizione grant from the Swiss National Science Foundation (grant 136737).

#### Conflict of interest

None declared.

#### Authors' contributions

AJK and CLA designed the study, carried out the analysis, and wrote the manuscript.

#### References

1. European Centre for Disease Prevention and Control (ECDC). Middle East respiratory syndrome coronavirus (MERS-CoV). 17th update, 11 June 2015. Rapid risk assessment. Stockholm: ECDC; 2015. Available from: <http://ecdc.europa.eu/en/publications/Publications/middle-east-respiratory-syndrome-coronavirus-rapid-risk-assessment-11-june-2015.pdf>
2. World Health Organization (WHO). Middle East respiratory syndrome coronavirus (MERS-CoV), 15 June 2015. Geneva: WHO. [Accessed 15 Jun 2015]. Available from: [http://www.who.int/csr/disease/coronavirus\\_infections/en/](http://www.who.int/csr/disease/coronavirus_infections/en/)

3. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature*. 2005;438(7066):355-9. <http://dx.doi.org/10.1038/nature04153> PMID:16292310
4. Lloyd-Smith JO. Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. *PLoS ONE*. 2007;2(2):e180. <http://dx.doi.org/10.1371/journal.pone.0000180> PMID:17299582
5. Poletto C, Pelat C, Levy-Bruhl D, Yazdanpanah Y, Boelle PY, Colizza V. Assessment of the Middle East respiratory syndrome coronavirus (MERS-CoV) epidemic in the Middle East and risk of international spread using a novel maximum likelihood analysis approach. *Euro Surveill*. 2014;19(23):20824. <http://dx.doi.org/10.2807/1560-7917.ES2014.19.23.20824> PMID:24957746
6. Breban R, Riou J, Fontanet A. Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk. *Lancet*. 2013;382(9893):694-9. [http://dx.doi.org/10.1016/S0140-6736\(13\)61492-0](http://dx.doi.org/10.1016/S0140-6736(13)61492-0) PMID:23831141
7. Cauchemez S, Fraser C, Van Kerkhove MD, Donnelly CA, Riley S, Rambaut A, et al. Middle East respiratory syndrome coronavirus: quantification of the extent of the epidemic, surveillance biases, and transmissibility. *Lancet Infect Dis*. 2014;14(1):50-6. [http://dx.doi.org/10.1016/S1473-3099\(13\)70304-9](http://dx.doi.org/10.1016/S1473-3099(13)70304-9) PMID:24239323
8. Blumberg S, Lloyd-Smith JO. Inference of  $R(0)$  and transmission heterogeneity from the size distribution of stuttering chains. *PLOS Comput Biol*. 2013;9(5):e1002993. <http://dx.doi.org/10.1371/journal.pcbi.1002993> PMID:23658504
9. Nishiura H, Yan P, Sleeman CK, Mode CJ. Estimating the transmission potential of supercritical processes based on the final size distribution of minor outbreaks. *J Theor Biol*. 2012;294:48-55. <http://dx.doi.org/10.1016/j.jtbi.2011.10.039> PMID:22079419
10. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2015
11. Nishiura H, Miyamatsu Y, Chowell G, Saitoh M. Assessing the risk of observing multiple generations of Middle East respiratory syndrome (MERS) cases given an imported case. *Euro Surveill*. 2015;20(27). pii: 21181. PMID: 26212063.
12. Althaus CL. Ebola superspreading. *Lancet Infect Dis*. 2015;15(5):507-8. [http://dx.doi.org/10.1016/S1473-3099\(15\)70135-0](http://dx.doi.org/10.1016/S1473-3099(15)70135-0) PMID:25932579
13. Drosten C, Muth D, Corman VM, Hussain R, Al Masri M, HajOmar W, et al. An observational, laboratory-based study of outbreaks of Middle East respiratory syndrome coronavirus in Jeddah and Riyadh, kingdom of Saudi Arabia, 2014. *Clin Infect Dis*. 2015;60(3):369-77. <http://dx.doi.org/10.1093/cid/ciu812> PMID:25323704
14. Blumberg S, Worden L, Enanoria W, Ackley S, Deiner M, Liu F, et al. Assessing measles transmission in the United States following a large outbreak in California. *PLOS Curr*. 2015;7.pii: ecurrents.outbreaks.b497624d7043b1aecfbfd3dfda3e344a. <http://dx.doi.org/10.1371/currents.outbreaks.b497624d7043b1aecfbfd3dfda3e344a>