

Title: A computational analysis on Covid-19 transmission raises immuno-epidemiology concerns.

Authors: Anthony M. Kyriakopoulos^{1,*}, Shi Zhao^{2,3}.

5 **Affiliations:**

¹Nasco AD Biotechnology Laboratory, 11 Sachtouri Str., Piraeus 18536, Greece

²JC School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong
999077, China

³CUHK Shenzhen Research Institute, Shenzhen 518060, China

10 *Correspondence to: antkyriak@gmail.com.

Abstract: For Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-COV-2) the investigation of the heterogeneity of individual infectiousness becomes important due to the cross reactive immunity of general population. Using a sample of infected population with SARS-COV-2 in close geographical proximity to the initial Severe Advanced Respiratory Syndrome-1 (SARS-1) outbreak, we explored the association between infector's age and dispersion (or heterogeneity) of individual infectiousness (k) in order to investigate the relatedness with the age of an individual's capability to disperse SARS-COV-2. Interestingly, we find a negative association between k and increase of infector's age. Significantly this becomes more evident for the age group of 20-60 years comparing with the infectors with younger age. This raises important immuno-epidemiology concerns for effectiveness of public health measures to contain the disease.

One Sentence Summary: Dispersion of Coronavirus Disease-19 in China differed with age.

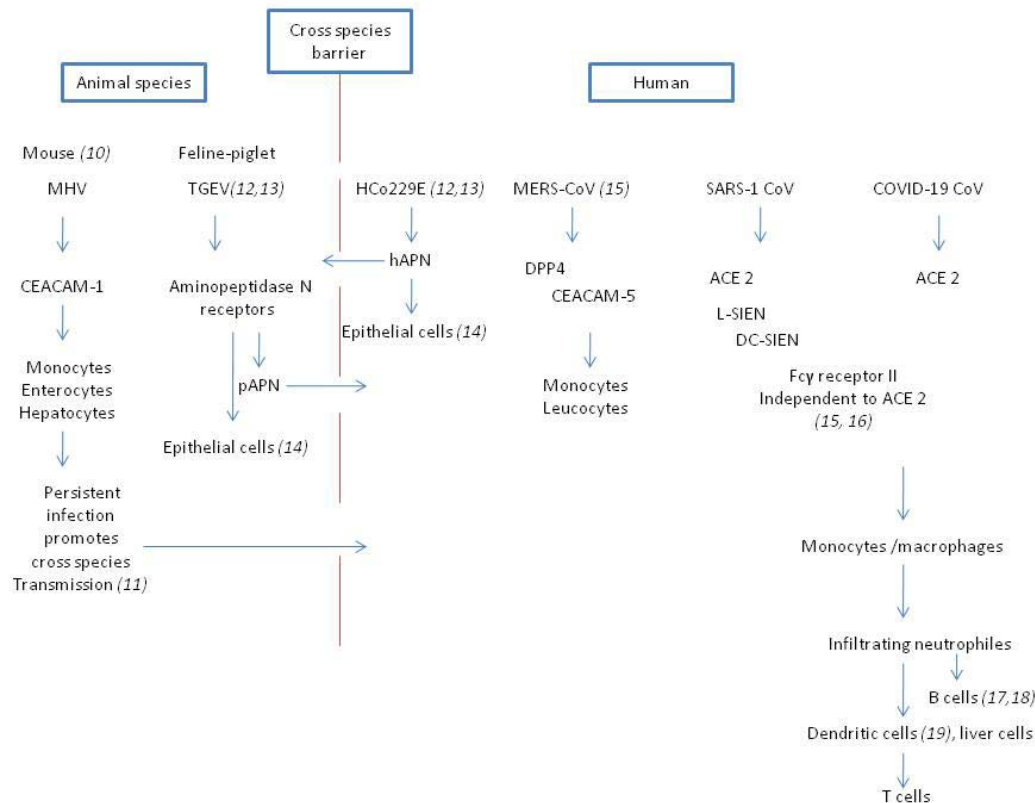
Main Text: The resistance of the "host-donor" defines the efficient transmission of the infectious agents (I). This resistance can be described as the immunity of individuals raised to disable transmission. The previous Severe Acute Respiratory Syndrome-1 (SARS-1) patients, even after 17 years post infection poses a sustained-well-developed, specific T cell memory response against SARS-1 Coronavirus (SARS-1-COV). This specific immunity is intensified according to the increased severity of previous SARS-1 clinical condition (2). These SARS-1 patients whilst all poses a long term cross reactive immunity against SARS-1-COV, also all (to a tested sample of 23/23), poses reacting T cells to the N peptides of SARS-COV-2 (3). Further, the unexposed individuals with no previous history of SARS-1 infection, and negative for Nucleocapsid protein (N) antibodies and neutralizing antibodies of SARS-COV-2, to a 51.35 % (19/37), have also a reactive T cell immunity against SARS-COV-2 N, and non structural proteins (nsp) proteins of SARS-COV-2 (50 % and 50 % respectively) (3). Also a widely

distributed T cell cross reactivity (35 % – 60 %) amongst unexposed individuals extends against to the SARS-COV-2 spike (S) protein (4,5).

The specific memory immunity of unexposed individuals is attributed either to immune cross reactions with common flu coronavirus infections (4,5) or to the involvement of animal species cross transmitting coronaviruses to human (3). Since there is a 17 year interval from SARS-1 epidemic and the developmental stages of immunity can be severely influenced (6,7) by the cross barrier transmission of coronaviruses (including common cold coronaviruses) between animal species and human individuals “Fig. 1”, we have selected specifically to investigate the variation of dispersion parameter k with age increase, bearing in mind the concept of higher $k \rightarrow$ less heterogeneity of age group \rightarrow more difficult to control the epidemic (8). We considered that investigation on this route could identify distinct variation in individual infectiveness with the new virus and thus identify variations on the efficiency of individuals to spread SARS-COV-2 between specific age groups.

Fig. 1. Cross species barrier transmission between human and animals by coronaviruses.

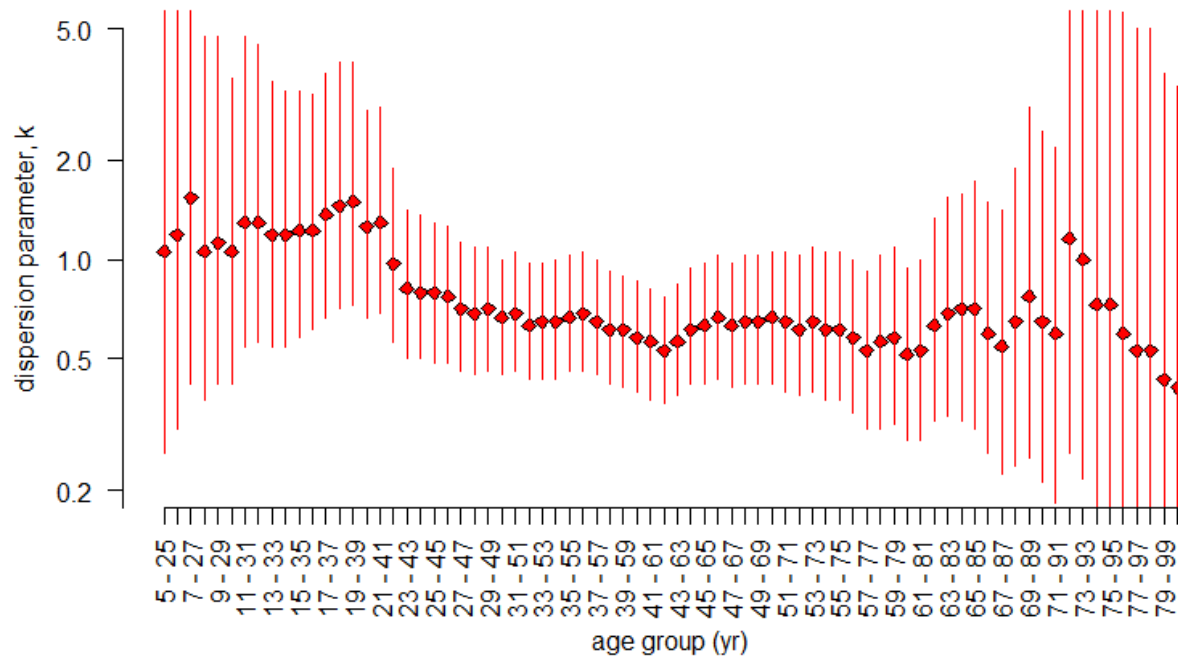
The cross species barrier infection is achieved by coronaviruses using suitable receptors that enable this transmission. Molecules like the carcinoembryonic antigen-related adhesion molecules (CEACAM) comprise a family of antigens which are highly preserved between animal and human species and their participation can lead to a wide immune cellular dispersal of coronaviruses throughout the human organism (9).



pAPN: Piglet aminopeptidase N receptor; hAPN: Human aminopeptidase N receptor; L-SIEN: Liver/Lymph node specific intracellular adhesion molecule 3-grabbing integrin; DC-SIEN : Dendritic cell specific intracellular adhesion molecule 3-grabbing non integrin

We have therefore employed the metric named dispersion of individual infectiousness, denoted by k , which was first proposed for SARS-1 (20) since it was used to quantify the role of heterogeneity of individuals in transmitting Coronavirus Disease -19 (Covid-19) as well as the difficulty in controlling the epidemics with Non Pharmaceutical Interventions (NPIs) at population scale (21,22). We performed a statistical calculation of k , and explored to define possible relatedness of infector's age as a determinant of k . We estimated the k ranges from 0.4 to 1.5 for different age bins, which is in line with previous studies (8, 21,22). We observe an evident downward trend in k as the infector's age increases, with p-value < 0.001 for linear trends testing using Student's t test. Moreover, we detect a structural break at age bin 20-40 years, in which k drops 47% with p-value < 0.001 comparing with the infector with younger age "Fig. 2".

Fig. 2. The dispersion parameter (k). The estimated dispersion parameter (k) in different age sub-groups. The dots are the estimates, and the vertical bars are the 95% CI.



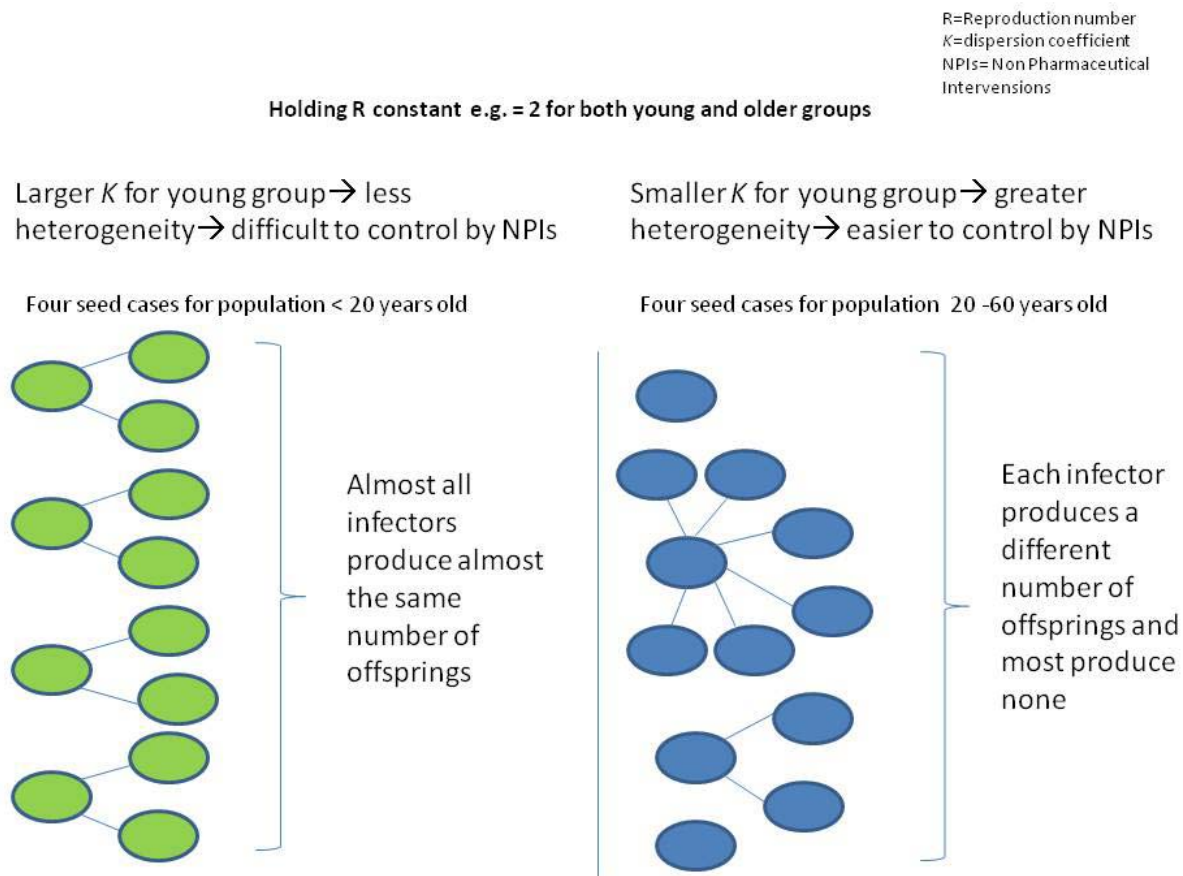
The dataset [see (23) for full list of transmission pairs] we used to investigate variation in heterogeneity of individual infectiveness with age initially contained 1407 transmission pairs (22), out of which we have identified 807 infectors. Out of these infectors, we selected 777 with age information.

In the investigation of individual infectiveness we kept the reproduction number R constant in order to measure the variation of individual dispersiveness, k , with age. The k value decreases with age increase, with the important difference being between the two age groups 0-20, and 20-60 years of age (Fig.2). This is important, as their difference lies in the heterogeneity of each sub-population group (8,20). For youth age as k is larger, the heterogeneity of subgroup

5

population is smaller and for the older age group as k , decreases the heterogeneity between individual increases (20, 22, 24). Hypothesizing that R equals to 2, which is a realistic scenario for SARS-COV-2 (25), for four seed cases of youth and old age respectively, the offspring cases will be eight for each age group. This means that almost all youth age infectors will produce two offspring cases, whereas for old age infectors, almost half will not be able to produce any offspring case, one will be able to produce six offspring cases and one two offspring cases “Fig.3”.

Figure 3. The difference of k between age groups concerns the ability of infectors to transmit the disease.



10

15

20

Reflecting the change of individual infectiveness with age and thus their heterogeneity, NPIs, are more applicable to the sub-population older than 20 years of age (20), whereas NPIs are not expected to provide an adequate solution (26), for disease spread containment from younger age infectors. Given the situation that a high proportion of youth age remains asymptomatic but highly infectious (6, 27), this makes the contact tracing effort in this group even more difficult but urgent. Specific screening strategy for the youth population to identify as more possible positivity will make restriction contact measures more efficient as almost all young infected individuals are likely to transmit the disease. The notable differences in the heterogeneity of individuals across the age groups of 0-20, 20-60 and over 60 years old may reflect de-similarities between developments of immune surveillance mechanisms due to environmental cross transmission reactions (10-19). These are encountered in previous SARS-1 patients and

unexposed individuals (3-5). Due to Covid-19 overspread and consistence, the targeted pharmaceutical interventions may be appropriate to lower seed cases in population groups with small heterogeneity. As our results show, by focusing on youth age population to prevent from spreading SARS-COV-2 to rest of general population, this will help to contain pandemic in a similar way to SARS-1 (26).

References and Notes:

1. J.F. Stephenson, K.A. Young, J. Fox, J. Jokela, J. Cable, S.E. Perkins, Host heterogeneity affects both parasite transmission to and fitness on subsequent hosts. *Philos. Trans. R. Soc. Lond. Biol. Sci.* **372**(1719):20160093 (2017). doi:10.1098/rstb.2016.0093.
2. O.W Ng, A. Chia, A.T. Tan, R.S. Jadi, H.N. Leong, HN, A. Bertoletti, Y.J. Tan. Memory T cell responses targeting the SARS coronavirus persist up to 11 years post-infection. *Vaccine.* **12**;34(17):2008-14 (2016). doi: 10.1016/j.vaccine.2016.02.063.
3. N. Le Bert, et al., SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature.* **584**(7821):457-462 (2020). doi: 10.1038/s41586-020-2550-z.
4. A. Grifoni, et al., Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell.* **25**;181(7):1489-1501.e15 (2020). doi: 10.1016/j.cell.2020.05.015.
5. L. Pia, SARS-CoV-2-reactive T cells in patients and healthy donors. *Nat. Rev. Immunol.* **20**(6):353 (2020). doi: 10.1038/s41577-020-0333-2.
6. P. Zimmermann, N. Curtis, Coronavirus Infections in Children Including COVID-19: An Overview of the Epidemiology, Clinical Features, Diagnosis, Treatment and Prevention Options in Children. *Pediatr. Infect. Dis. J.* **39**(5):355-368 (2020). doi:10.1097/INF.0000000000002660
7. M. Mesel-Lemoine, et al., A human coronavirus responsible for the common cold massively kills dendritic cells but not monocytes. *J. Virol.* **86**(14):7577-7587 (2012). doi:10.1128/JVI.00269-12
8. D. He, et al. Low dispersion in the infectiousness of COVID-19 cases implies difficulty in control. *BMC Public Health.* **20**(1):1558 (2020). doi: 10.1186/s12889-020-09624-2.
9. B. Rizeq, Z. Zakaria, A. Ouhtit. Towards understanding the mechanisms of actions of carcinoembryonic antigen-related cell adhesion molecule 6 in cancer progression. *Cancer Sci.* **109**(1):33-42 (2018). doi:10.1111/cas.13437
10. F. Taguchi, S. Matsuyama, Soluble Receptor Potentiates Receptor-Independent Infection by Murine Coronavirus. *J. Virol.* **76**(3):950-8 (2002).
11. R.S. Baric, E. Sullivan, L. Hensley, B. Yount, W. Chen, Persistent infection promotes cross-species transmissibility of mouse hepatitis virus. *J. Virol.* **73**(1): 638-649 (1999).
12. D.E. Wentworth, K.V. Holmes. Molecular determinants of species specificity in the coronavirus receptor aminopeptidase N (CD13): influence of N-linked glycosylation. *J. Virol.* **75**(20):9741-9752 (2001). doi:10.1128/JVI.75.20.9741-9752.2001
13. G.H. Hansen, et al., The coronavirus transmissible gastroenteritis virus causes infection after receptor-mediated endocytosis and acid-dependent fusion with an intracellular compartment. *J. Virol.* **72**(1):527-534 (1998).
14. J.W. Rossen, C.P. Bekker, W.F. Voorhout, G.J. Strous, A. van der Ende, P.J. Rottier, Entry and release of transmissible gastroenteritis coronavirus are restricted to apical surfaces of polarized epithelial cells. *J. Virol.* **68**(12):7966-7973 (1994).
15. C.M. Chan, et al., Carcinoembryonic Antigen-Related Cell Adhesion Molecule 5 Is an Important Surface Attachment Factor That Facilitates Entry of Middle East Respiratory Syndrome Coronavirus. *J. Virol.* **90**(20): 9114-9127 (2016). doi:10.1128/JVI.01133-16

16. M. Jaume, et al., Anti-severe acute respiratory syndrome coronavirus spike antibodies trigger infection of human immune cells via a pH- and cysteine protease-independent Fc γ R pathway. *J Virol.* **85**(20):10582-10597 (2011). doi:10.1128/JVI.00671-11
17. R. Kammerer, W. Zimmermann, Coevolution of activating and inhibitory receptors within mammalian carcinoembryonic antigen families. *BMC Biol.* **8**:12 (2010). doi:10.1186/1741-7007-8-12
18. J.W. Booth, et al., Phosphatidylinositol 3-kinases in carcinoembryonic antigen-related cellular adhesion molecule-mediated internalization of *Neisseria gonorrhoeae*. *J. Biol. Chem.* **278**(16):14037-14045 (2003). doi:10.1074/jbc.M211879200
19. M. Spiegel, K. Schneider, F. Weber, M. Weidmann, F.T. Hufert. Interaction of severe acute respiratory syndrome-associated coronavirus with dendritic cells. *J. Gen. Virol.* **87**(Pt 7):1953-1960 (2006). doi:10.1099/vir.0.81624-0
20. J.O. Lloyd-Smith, S.J. Schreiber, P.E. Kopp, W.M. Getz. Superspreading and the effect of individual variation on disease emergence. *Nature.* **17**;438(7066):355-9 (2005). doi: 10.1038/nature04153.
21. D.C. Adam, et al., Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat. Med.* [published ahead of print <https://doi.org/10.1038/s41591-020-1092-0>] (2020)
22. Kwok KO, Chan HHH, Huang Y, Hui DSC, Tambyah PA, Wei WI, Chau PYK, Wong SYS, J.W.T. Tang, Inferring super-spreading from transmission clusters of COVID-19 in Hong Kong, Japan, and Singapore. *J. Hosp. Infect.* **105**(4):682-685 (2020). doi: 10.1016/j.jhin.2020.05.027. 32446721.
23. X.K. Xu, et al., Reconstruction of Transmission Pairs for novel Coronavirus Disease 2019 (COVID-19) in mainland China: Estimation of Super-spreading Events, Serial Interval, and Hazard of Infection. *Clin. Infect. Dis.* [published ahead of print 2020 Jun 18:ciaa790. doi: 10.1093/cid/ciaa790] (2020).
24. M.E. Woolhouse, et al., Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl. Acad. Sci. U S A.* **94**:338-42 (1997). <https://doi.org/10.1073/pnas.94.1.338>
25. A.M. Kyriakopoulos, A. Papaefthymiou, N. Georgilas, M. Doulberis, J. Kountouras. The Potential Role of Super Spread Events in SARS-COV-2 Pandemic; a Narrative Review. *Arch. Acad. Emerg. Med.* **8**(1):e74 (2020).
26. S. Riley S, et al., Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science.* **300**(5627): 1961-6 (2003). doi:10.1126/science.1086478
27. D.S. Oliveira, N.I. Medeiros, J.A.S. Gomes, Immune response in COVID-19: What do we currently know?. *Microb Pathog.* **148**:104484 (2020).
28. A.J. Kucharski, C.L. Althaus, The role of superspreading in Middle East respiratory syndrome coronavirus (MERS-CoV) transmission. *Euro Surveill.* **20**(25):14-8 (2015). doi: 10.2807/1560-7917.es2015.20.25.21167
29. O. Diekmann, J.A.P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis, and Interpretation*. S. Levin, Ed., Wiley series in mathematical and computational biology (John Wiley & sons, New York, 2000).
30. https://en.wikipedia.org/wiki/Structural_break

Acknowledgments: We thank Professor J. Kountouras and Professor Markus Nagl for their valuable communication on Covid-19 epidemiology. We thank our families for moral support in conducting this research. **Funding:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. **Author contributions:** AK inspired the hypothesis of propagation of Covid-19 through different immune surveillance of population groups and hypothesized for age differences. AK has written the manuscript and SZ conducted the proper mathematical analysis to investigate the AK hypothesis and performed the mathematical part of this manuscript. SZ has written the materials and methods. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** All data is available in the main text or the supplementary materials.

Supplementary Data:

Materials and Methods

Supplementary text

References (28-30)

Supplementary Data for

A computational analysis on Covid-19 transmission raises imuuno-epidemiology concerns.

5 Anthony M. Kyriakopoulos^{1,*}, Shi Zhao^{2,3}

Correspondence to: antkyriak@gmail.com

10 **This supplementary file includes:**

Materials and Methods
Supplementary Text
References

15

Materials

Data collection and extraction

5 We used the COVID-19 surveillance data previously published (23) and the dataset can be accessed freely via the public respiratory https://github.com/linwangidd/covid19_transmissionPairs_China/blob/master/transmission_pairs_covid_v2.csv.

10 The dataset contains 1407 transmission pairs that are identified and reconstructed according to the previous studies (8, 20), governmental new release, and official situation reports.

We identified 807 infectors, and who act as source of infection to transmit to the infectee. We extract the information, including age and gender, of each infector as well as the number of offspring infectees generated by each infector. After excluding the infector with missing information on age, we collected 777 infectors for further analysis.

15 Methods

1. Heterogeneity of individual infectiousness: a statistical modeling perspective

20 We consider the variation in the individual-level infectiousness as a quantifiable scale that affects the distribution of offspring infectee generated by an infector. Following the previous study (20), we introduce the number of offspring infectee generated by an infector, denoted by r , as a random variable from a Gamma distribution, denoted by $h()$, with mean $R (> 0)$ and dispersion parameter $k (> 0)$. Thus, we have $r \sim h(R, k)$. Here, R is the reproduction number that is defined as the expected (or average) number of secondary cases caused by one typical infected individual.

25 The dispersion parameter k governs the dispersiveness of the Gamma distribution. As demonstrated theoretically in the previous study (20), with R fixed, a larger k results in a lower effectiveness of non-pharmaceutical interventions in controlling the epidemics, which is also discussed in another study (8).

30 Poisson process with rate r , denoted by $f(r)$, is adopted to address the stochastic effects in transmission, and to govern the number of infectee caused by each infector, denoted by $Z (\geq 0)$. (29). Thus, we have $Z \sim f(r) = f(R, k)$. Straightforwardly, $f(R, k)$ is an Negative Binomial (NB) distribution with mean R and variance $R \cdot (1 + R/k)$. By the definition of NB distribution, the probability that one infector generates $(j-1)$ offspring infectees, i.e., cluster size of $j (\geq 1)$, which is denoted by $\Pr(Z = j-1) = L_j$, is given in Eqn (1).

$$\Pr(Z = j - 1) = L_j = \frac{\Gamma(kj + j - 1)}{\Gamma(kj) \cdot \Gamma(j + 1)} \cdot \frac{\left(\frac{R}{k}\right)^{j-1}}{\left(1 + \frac{R}{k}\right)^{kj + j - 1}}. \quad \text{Eqn (1)}$$

40 Here, $\Gamma(\cdot)$ denotes the Gamma function. Specially, the NB distribution $f(\cdot)$ reduces to a Geometric distribution when $k = 1$, and it reduces to a Poisson distribution when k approaches infinity. Importantly, a smaller value of k indicates larger heterogeneity in individual infectiousness.

By fitting distribution $f(\cdot)$ to the real-world observations, we may estimate value of dispersion parameter k , and explore the determinants of k .

2. Likelihood inference framework and subgrouping by infector's age

We consider observed samples of number of offsprings from N infectors. We denote the number of infectors who have j infectees associated by n_j (≥ 1). Note that all $j > 1$ in our dataset though j may be 1 theoretically or observed in other datasets, and thus we adjusted for this truncation in our likelihood framework. Straightforwardly, we have $\sum_{j>1} n_j = N$. Then, following the previous studies (21, 28), we constructed the likelihood function, denoted by L , as in Eqn (2).

$$L = \prod_{j>1} \left(\frac{L_j}{1-L_1} \right)^{n_j}. \quad \text{Eqn (2)}$$

We estimated the dispersion parameter k using the maximal likelihood estimation approach. To explore the association between the infector's age and k , we repeated the above fitting and estimation procedure after sub-setting the dataset into subgroups by the infector's age. We considered 76 age bins, and they include 5-25, 6-26, ..., 79-99, and 80+ years. We estimate the value of k for each age bin to examine the association between the infector's age and k .

Supplementary text

Estimating results

We estimated the k ranges from 0.4 to 1.5 for different age bins, which is line with previous studies (8, 21, 22). We observe an evident downward trend in k as the infector's age increases, with p -value < 0.001 for linear trends testing using Student's t test. We detect a structural break (30) at age bin 20-40 years, in which k drops 47% with p -value < 0.001 comparing with the infector with younger age seen in "Fig. 2".

References for the supplementary information

28. A.J. Kucharski, C.L. Althaus, The role of superspreading in Middle East respiratory syndrome coronavirus (MERS-CoV) transmission. *Euro Surveill.* **20**(25):14-8 (2015). doi: 10.2807/1560-7917.es2015.20.25.21167
29. O. Diekmann, J.A.P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis, and Interpretation*. S. Levin, Ed., Wiley series in mathematical and computational biology (John Wiley & sons, New York, 2000).
30. https://en.wikipedia.org/wiki/Structural_break