medRxiv preprint doi: https://doi.org/10.1101/2020.06.20.20130476; this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Characterizing super-spreading events and age-specific infectivity of COVID-19 transmission in Georgia, USA

Max SY Lau[a], Bryan Grenfell[b,c], Kristin Nelson[d], and Ben Lopman[d]

[a]Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University
[b]Department of Ecology and Evolutionary Biology, Princeton University, Princeton
[c]Fogarty International Center, National Institute of Health, USA
[d]Department of Epidemiology, Rollins School of Public Health, Emory University

Abstract

As the current COVID-19 pandemic continues to impact countries around the globe, refining our understanding of its transmission dynamics and the effectiveness of interventions is imperative. In particular, it is essential to obtain a firmer grasp on the effect of social distancing, potential individual-level heterogeneities in transmission such as age-specific infectivity, and impact of super-spreading. To this end, it is important to exploit multiple data streams that are becoming abundantly available during the pandemic. In this paper, we formulate an individual-level spatio-temporal mechanistic framework to statistically integrate case data with geo-location data and aggregate mobility data, enabling a more granular understanding of the transmission dynamics of COVID-19. We analyze reported cases from surveillance data, between March and early May 2020, in five (urban and rural) counties in the State of Georgia USA. We estimate natural history parameters of COVID-19 and infer unobserved quantities including infection times and transmission paths using Bayesian data-augmentation techniques. First, our results show that the overall median reproductive number was 2.88 (with 95% C.I. [1.85, 4.9]) before the state-wide shelter-in-place order issued in early April, and the effective reproductive number was reduced to below 1 about two weeks by the order. Super-spreading appears to be widespread across space and time, and it may have a particularly important role in driving the outbreak in rural area and an increasing importance towards later stages of outbreaks in both urban and rural settings. Overall, about 2% of cases may have directly infected 20% of all infections. We estimate that the infected children and younger adults (<60 years old) may be 2.38 [1.30, 3.51] times more transmissible than infected elderly (>=60), and the former may be the main driver of super-spreading. Through the synthesis of multiple data streams using our transmission modelling framework, our results enforce and improve our understanding of the natural history and transmission dynamics of COVID-19. More importantly, we reveal the roles of age-specific infectivity and characterize systematic variations and associated risk factors of super-spreading. These have important implications for the planning of relaxing social distancing and, more generally, designing optimal control measures.

1 Introduction

The current COVID-19 pandemic continues to spread and impact countries across the globe. There is still much scope for mapping out the whole spectrum of the epidemiology and ecology of this novel virus. In particular, understanding of heterogeneities in transmission, which is essential for devising effective targeted control measures, is still limited. For instance, much is unknown about the variation of infectivity among different age groups [1, 2, 3]. Also, while super-spreading events have been documented, its impact and variation over space and time and associated risk factors have not yet been systematically characterized [1, 4, 5, 6].

For this reason, it is crucial to exploit the growing availability of multiple data streams during the pandemic, from which we may obtain a more comprehensive picture of the transmission dynamics of COVID-19. For example, shelter-in-place order could change the movement pattern (e.g. reduced

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.20.20130476; this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

distance of travel) of a population. Such a change needs to be taken into account as movement is a key factor that shapes transmission [7]. Failing to capture this change would also bias the estimates of key model parameters including intervention efficacy and transmissibility parameters that are correlated with movement [8, 9, 10]. De-identified mobility data from mobile phone users have been made available to state governments and research institutes through partnerships with private companies such as Facebook. Integration of such mobility data with surveillance data would allow us to account for the change in movement, and therefore more accurately infer the transmission dynamics [7, 8, 10, 11]. Geo-spatial location data and detailed spatial distribution of population are also important for capturing heterogeneous mixing in space [8, 9, 10]. A key step is to enable individual-level model inference that can properly synthesize these datastreams, which would go beyond most efforts so far that have focused on aggregated level dynamics [2, 11, 12].

In this paper, building on a previous framework we developed for modelling Ebola outbreaks in Western Africa [8, 9], we formulate an individual-level space-time stochastic model that describes the transmission of COVID-19 and captures the impact of state-wide social distancing measure in the state of Georgia, USA. Our model mechanistically integrates detailed individual-level surveillance data, geo-spatial location data and highly-resolved population density (grid-)data, and aggregate mobility data (see *Study Data*). We estimate model parameters and unobserved model quantities including infection times and transmission paths using Bayesian data-augmentation techniques in the framework of Markov Chain Monte Carlo (MCMC) (see *Materials and Methods*). Our individual-level modelling framework also allows us to compute population-level epidemiological parameter such as the basic reproductive number R_0 and quantify the degree of super-spreading over space and time.

2 Study Data

We analyzed a rich set of COVID-19 surveillance data collected by the Georgia Department of Public Health, between March 1, 2020, and May 3, 2020, in five counties which had the largest numbers of cases. These counties include four counties Cobb, Dekalb, Fulton and Gwinnett in urban area, and one rural Dougherty county. This dataset contains demographic information of 9,559 symptomatic cases which include age, sex and race, and symptom onset times. It also contains geo-location of the residence of cases. Highly-resolved population-density data over 100×100meter grids are obtained from *http://www.worldpop.org.uk.*, and are used to modulate the spatial spread of the virus (see *Materials and Methods*). Aggregate mobility data are used to characterize change of movement before and after the implementation of state-wide social distancing measure. Specifically, we used high-volume mobility data accessed through Facebook's Data for Good program [13]. These data represent Facebook users in Georgia who have location services enabled on their mobile device. These data provide information on the number of 'trips' (and trip distance) that occurred daily among users. A 'trip' is defined as a directional vector starting at the location where an individual spent most of their time during the previous 8-hour period and ending at the location where the same individual spent most of their time during the current 8-hour period.

3 Results

Natural History Parameters and Effectiveness of Social Distancing

We estimate that the median value of R_0 across five counties was overall between 2.88 with 95% C.I. [1.85, 4.9] before the intervention (i.e. shelter-in-place order). Dougherty county in the rural area had the largest prior-intervention R_0 (4.82 [4.31, 4.91]) and time-varying effective reproductive number R_{eff} at the earlier stage in March 2020 (Figure 1). Our results suggest the shelter-in-place order was effective, and in all the counties the R_{eff} declined below 1 in about two to three weeks since the intervention. It is also worth noting that R_{eff} appears to begin decline one to two weeks before shelter-in-place order in urban areas and earlier in Dougherty. We also estimate that the incubation period (i.e. waiting time from infection to symptoms onset) has a median value 6.34 days [5.17, 7.22]. These estimates are largely consistent with the literature [14, 15].

medRxiv preprint doi: https://doi.org/10.1101/2020.06.20.20130476; this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

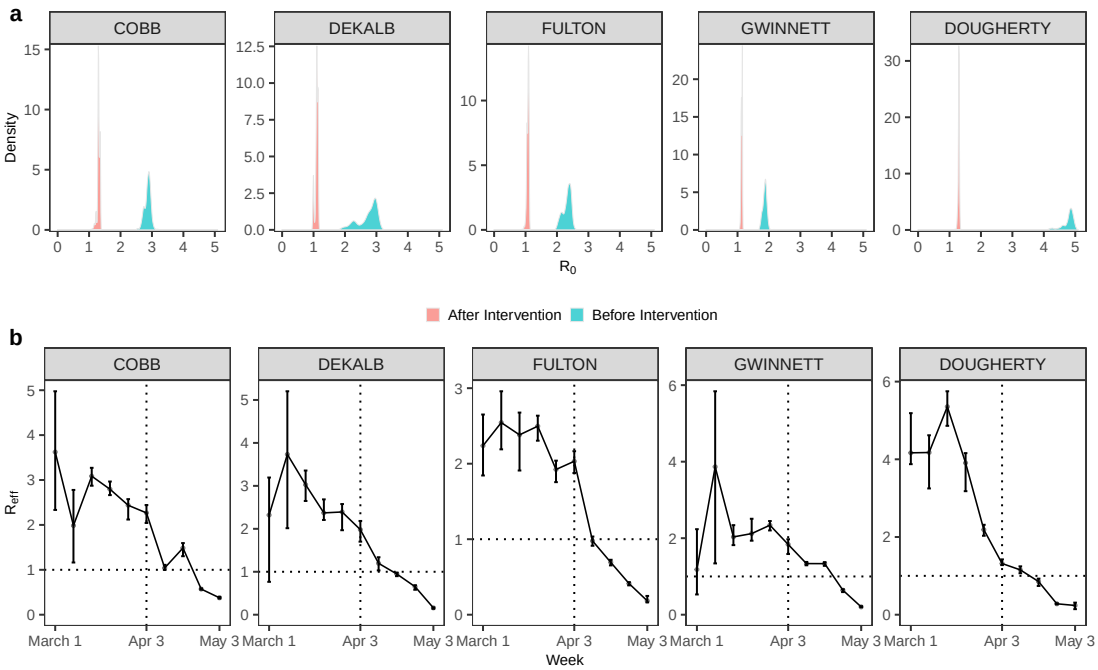

Figure 1: Posterior distribution of basic reproductive number R_0 and the effective reproductive number R_{eff}, before and after the implementation of state-wide shelter-in-place order on April 3, 2020. Errorbars represent 95% credible interval.

Systematic Characterization of Super-spreading

Super-spreading refers a phenomenon where certain individuals disproportionately infect a large number of secondary cases relative to an "average" infectious individual (whose infectivity may be well-represented by R_0). This phenomenon plays a key role in driving the spread of many pathogens including MERS and Ebola [8, 16]. A common measure of the degree of super-spreading is the dispersion parameter k, assuming that the distribution of the *offspring* (i.e. number of secondary cases generated) is negative binomial with variance $\sigma^2 = \mu(1 + \mu/k)$ where μ is the mean [16]. Generally speaking, a lower k corresponds to a higher degree of super-spreading; and k less than 1 implies substantial super-spreading. Our framework infers the transmission paths among all cases and therefore naturally generates the offspring distribution of each case (see *Materials and Methods*).

While super-spreading of COVID-19 has been observed [1, 4, 5, 6], systematic characterization of its impact and variation (e.g., over time and space) and associated risk factors is lacking. Our results (Figure 2a) suggest that super-spreading is a ubiquitous feature during different periods of the outbreak. Super-spreading may have a major impact for the rural area (Dougherty) during earlier periods – it has the lowest k in earlier periods among all counties (0.43 [0.39, 0.47] and 0.38 [0.34, 0.43]). Dougherty county has a disproportionately large outbreak compared to other more populated counties – having about only $\frac{1}{8}$ of the population of Cobb county (about 760,000), it has a comparable number of reported cases (1,628). Such an anomaly may be a consequence of the significant super-spreading and large (prior-intervention) R_0 in Dougherty (Figure 1a). This is also consistent with the evidence of super-spreading events due to a funeral in the area [17]. The increasing significance of super-spreading over time also highlights the importance of maintaining social distancing measures that may curtail close contacts (e.g. gatherings with densely packed crowds). Overall, top 2% of cases (that generate highest number of offsprings) are responsible directly for about 20% of the total infections. Our results also show that the younger infectees (children and younger adults who are <60) tends to the main driver of super-spreading (Figure 2b), which is also consistent with our estimates of higher infectivity in this age group (see *Age-specific Infectivity*).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.20.20130476; this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

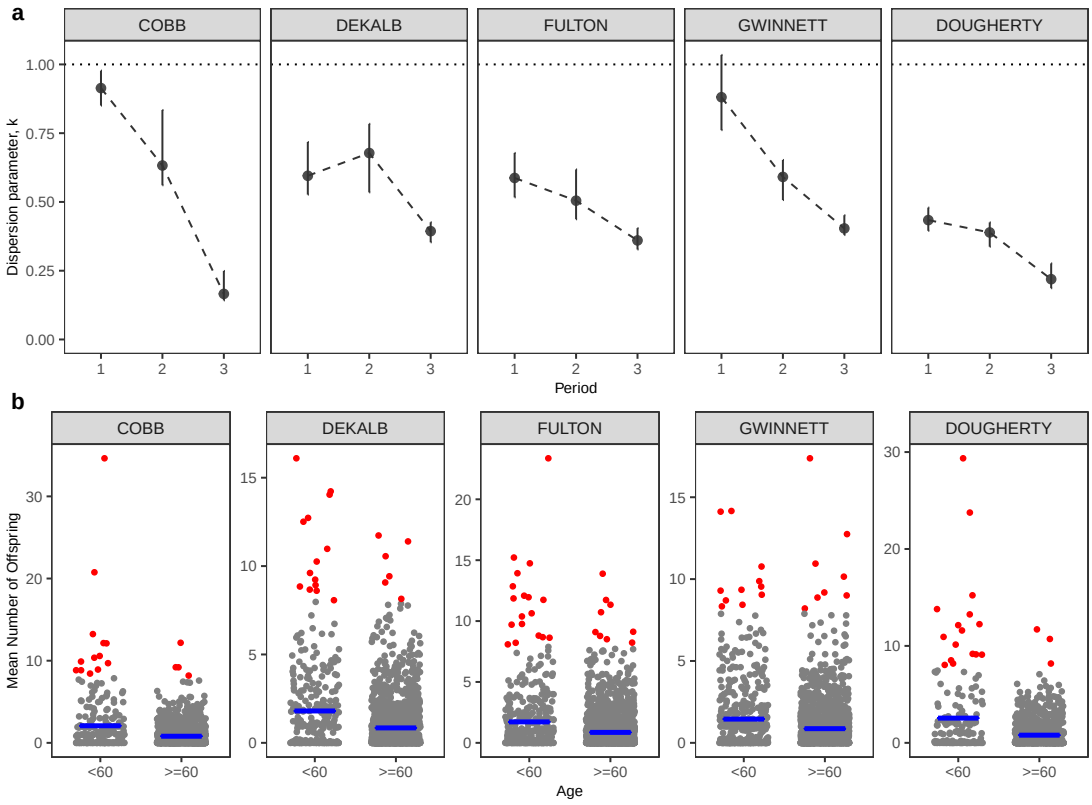

Figure 2: (a) Degree of super-spreading quantified by the dispersion parameter k (where $k < 1$ indicates significant super-spreading) during different periods of the outbreak. Let T be the day of announcing shelter-in-place order: period 1 is time $t < T$, period 2 is $[T, T + 14)$ and finally period 3 is $t > T + 14$. Overall, about top 2% of cases (that have highest mean number of offsprings) directly infected 20% of the total infections. (b) Mean number of offsprings generated by cases in each age group. Red dots represent those case have mean offspring >=8. The younger age group (<60) tends to have more cases that produce extreme number of offsprings, and also a larger average (blue lines) of the mean number of offsprings.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.20.20130476; this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

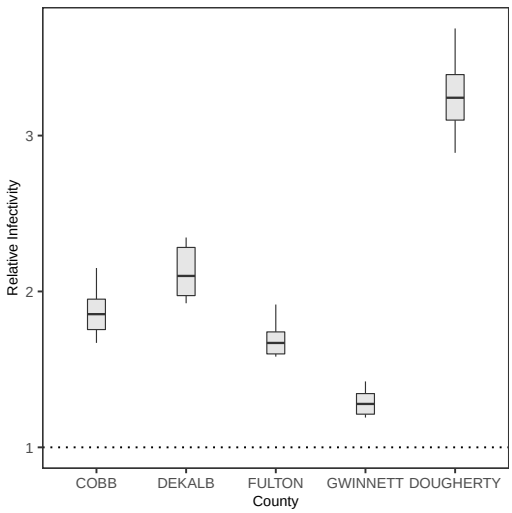

Figure 3: Infectivity of younger patients (<60) relative to the older patients (>=60).

3.1 Age-specific Infectivity

A markedly low proportion of cases among the younger children population is observed in the current COVID-19 pandemic, indicating heterogeneity of susceptibility among different age groups [3, 18, 19]. Much is unknown about the variation of infectivity among different age groups [1, 2, 3]. Our results suggest that the younger patients (children and adults who are <60) may be overall 2.38 [1.30, 3.51] times more transmissible than elderly patients (>=60). Due to the very small number of reported cases in children (e.g., <15), we do not consider a finer age stratification (see also *Discussion*). We also test the robustness of these results towards to under-reporting and take into account the discrepancy in the reporting rates of different age groups (see next section *Sensitivity Analysis*).

3.2 Sensitivity Analysis: Effect of Under-reported Cases

Under-reporting is ubiquitous feature of epidemiological data, and is particularly so for COVID-19 due to, in particular, a substantial number of asymptomatic cases and the lack of testings at the earlier stages of the pandemic. In particular, older people may tend to be more susceptible and develop severe symptoms, and hence have a higher probability of being reported [3, 18, 19]. Such a discrepancy in reporting rates may potentially affect our estimation of age-specific infectivity. We explore the effect of such under-reporting on our results under these probable scenarios: younger case (<60) has a probability 0.1 of being reported in March before the intervention, and this probability increases to 0.2 in April; the reporting probabilities for an elderly case are 0.3 and 0.6 in March and April respectively. Details of how to include under-reported cases are given in *Materials and Methods*. Figure 4 shows that the younger age group remains to be more infectious than the older age group. The estimated impact of super-spreading also appears to be robust: estimated overall dispersion parameter k is 0.38 for Cobb county, 0.49 for Dekalb, 0.47 for Fulton, 0.42 for Gwinnett, and 0.32 for Dougherty.

4 Discussion

Transmission dynamics of infectious diseases are often non-linear and heterogeneous over space and time. It is important to exploit available data that are relevant to describing and estimating such complex processes. A key step is to enable individual-level model inference that is able to statistically synthesize these datastreams, beyond aggregate-level dynamics [2, 11, 12]. In this paper, we incorporated multiple valuable datastreams including formal surveillance data into our individual-level spatio-temporal transmission modelling framework, achieving a more granular mechanistic understanding of the dynamics and heterogeneities in the transmission of COVID-19.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.20.20130476; this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

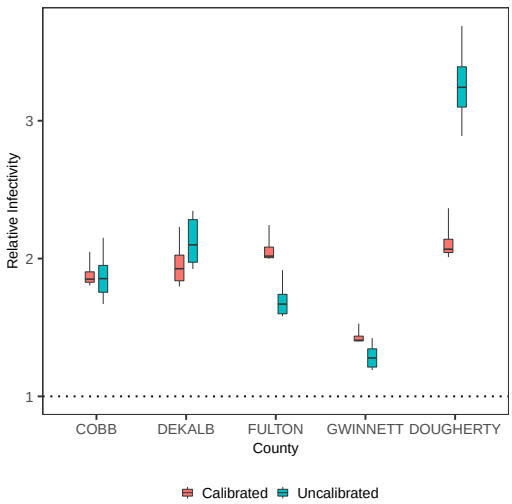

Figure 4: Infectivity of younger patients (<60) relative to the older patients (>=60), calibrated for under-reported cases.

Our results give similar estimates of important population-level epidemiological parameters such R_0 found in the literature, and reinforce the conclusion from most studies that social distancing measures are effective. This paper also advances our understanding of individual-level heterogeneities in the transmission of COVID-19, which is crucial for informing optimal interventions. We show that super-spreading is an important and ubiquitous feature throughout the pandemic, and it may have a pivotal role in driving the (large) outbreak in rural areas. The increasing significance of super-spreading over time also highlights the importance of maintaining social distancing measures that may curtail close contacts (e.g. gatherings with densely packed crowds). We also find that infected children and younger adults (<60) tend to be more transmissible and to promote super-spreading. Our results have important implications for designing more effective control measures − particularly, highlight the importance of more targeted interventions.

Our study has a number of limitations. First of all, due to the lack of widely available testings, the under-reporting rate was almost surely high during earlier phases of the pandemic. Also, severity of symptoms (and hence reporting rates) may vary among different age groups. We explore the robustness of our main results towards these possible under-reporting scenarios in the *Sensitivity Analysis*. Reassuringly, our main conclusions appears to be largely robust. Second, we only consider modelling age-specific infectivity in two age groups (<60 and >=60). The model would tend to be over-parameterized if we further break down the age groups that we have considered (e.g. by having a group for younger than 15), mainly due to imbalance of reported number of cases between age groups (particularly, markedly low reported numbers in very young population). Our current binning of age is still useful as the first group tends to be more socially active and is useful for informing the design of social distancing measures. Finally, although our analysis reveals the importance of age as a demographic risk factor of super-spreading, future work in linking them with biological factors (e.g., age-specific viral loads) may shed further light.

5 Materials and Methods

5.1 Spatio-temporal Transmission Process Model

We formulate an age-specific spatio-temporal transmission modelling framework that allows us to infer the unobserved infection times and transmission tree among cases, integrating detailed individual-level surveillance data, geo-spatial location data and highly-resolved population density (grid-)data, and aggregate mobility data . Our framework represents an extension of the models we developed *et al.* [8, 9], which were validated generally and applied successfully to dissect the transmission dynamics

medRxiv preprint doi: https://doi.org/10.1101/2020.06.20.20130476; this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

of the Ebola outbreak in Western Africa between 2014-2016. This approach also allows us to infer explicitly the distribution of the *offspring* (i.e. number of secondary cases generated) of each case, hence informing the degree of super-spreading.

We estimate $\mathbf{\Theta}$ (i.e. the parameter vector) in the Bayesian framework by sampling it from the posterior distribution $P(\mathbf{\Theta}|\mathbf{z})$ where $\mathbf{z}$ are the data. Denoting the likelihood by $L(\boldsymbol{\theta};\mathbf{z})$, the posterior distribution of $\mathbf{\Theta}$ is $P(\mathbf{\Theta}|\mathbf{z}) \propto L(\mathbf{\Theta};\mathbf{z})\pi(\mathbf{\Theta})$, where $\pi(\mathbf{\Theta})$ is prior distribution for $\mathbf{\Theta}$. Non-informative uniform priors for parameters in $\mathbf{\Theta}$ are used. Markov chain Monte Carlo (MCMC) techniques are used to obtain the posterior distribution. The unobserved infection times and transmission network and missing symptom onset dates are also imputed in the MCMC procedure.

5.2 Sensitivity analysis: Effect of Under-reported Cases

The number of total under-reported cases m for a particular age group during a particular period is calculated as $m = n/p - n$, where n is the reported number of cases and p the probability of being reported. We consider these scenarios: a younger case (<60) has a probability 0.1 of being reported in March before the intervention, and this probability increases to 0.2 in April; the reporting probabilities for an elderly case are 0.3 and 0.6 in March and April respectively. The m cases are then assigned infection times and spatial locations according to the temporal and spatial distributions of observed cases in the time period, before merging with the observed data. The infection times and sources of infections of the m cases are also treated as unknown and and are inferred in the data-augmentation procedure. Our main focus is to test how the potential discrepancy in reporting rate between age groups may impact our estimation of age-specific infectivity.

Acknowledgement

We thank Dr. Laura Edison and Michael Bryan from the Georgia Department of Public Health for their efforts in leading and coordinating the research partnership between GDPH and Emory University. We also would like to acknowledge Nishant Kishore and the COVID-19 Mobility Network for their support accessing and analyzing Facebook mobility data. We also thank Professor Gavin Gibson for his helpful discussion during the preparation of this manuscript. Finally, we thank Google-For-Education program (*https://edu.google.com)* for donating free cloud credits which are used for geocoding involved in the research.

References

[1] Qifang Bi, Yongsheng Wu, Shujiang Mei, Chenfei Ye, Xuan Zou, Zhen Zhang, Xiaojian Liu, Lan Wei, Shaun A. Truelove, Tong Zhang, Wei Gao, Cong Cheng, Xiujuan Tang, Xiaoliang Wu, Yu Wu, Binbin Sun, Suli Huang, Yu Sun, Juncen Zhang, Ting Ma, Justin Lessler, and Tiejian Feng. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases*, 0(0), April 2020. Publisher: Elsevier.

[2] Kiesha Prem, Yang Liu, Timothy W Russell, Adam J Kucharski, Rosalind M Eggo, Nicholas Davies, Stefan Flasche, Samuel Clifford, Carl A B Pearson, James D Munday, Sam Abbott, Hamish Gibbs, Alicia Rosello, Billy J Quilty, Thibaut Jombart, Fiona Sun, Charlie Diamond, Amy Gimma, Kevin van Zandvoort, Sebastian Funk, Christopher I Jarvis, W John Edmunds, Nikos I Bosse, Joel Hellewell, Mark Jit, and Petra Klepac. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *The Lancet Public Health*, 5(5):e261–e270, May 2020.

[3] Nicholas G Davies, Petra Klepac, Yang Liu, Kiesha Prem, Mark Jit, CMMID COVID-19 working group, and Rosalind M Eggo. Age-dependent effects in the transmission and control of COVID-19 epidemics. preprint, Epidemiology, March 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.20.20130476; this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[4] Shin Young Park, Young-Man Kim, Seonju Yi, Sangeun Lee, Baeg-Ju Na, Chang Bo Kim, Jung-il Kim, Hea Sook Kim, Young Bok Kim, Yoojin Park, In Sil Huh, Hye Kyung Kim, Hyung Jun Yoon, Hanaram Jang, Kyungnam Kim, Yeonhwa Chang, Inhye Kim, Hyeyoung Lee, Jin Gwack, Seong Sun Kim, Miyoung Kim, Sanghui Kweon, Young June Choe, Ok Park, Young Joon Park, and Eun Kyeong Jeong. Early Release - Coronavirus Disease Outbreak in Call Center, South Korea - Volume 26, Number 8—August 2020 - Emerging Infectious Diseases journal - CDC. 2020.

[5] Akira Endo, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Sam Abbott, Adam J. Kucharski, and Sebastian Funk. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Research*, 5(67):67, April 2020.

[6] Dillon Adam, Peng Wu, Jessica Wong, Eric Lau, Tim Tsang, Simon Cauchemez, Gabriel Leung, and Benjamin Cowling. Clustering and superspreading potential of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections in Hong Kong. preprint, In Review, May 2020.

[7] Caroline O. Buckee, Satchit Balsari, Jennifer Chan, Mercè Crosas, Francesca Dominici, Urs Gasser, Yonatan H. Grad, Bryan Grenfell, M. Elizabeth Halloran, Moritz U. G. Kraemer, Marc Lipsitch, C. Jessica E. Metcalf, Lauren Ancel Meyers, T. Alex Perkins, Mauricio Santillana, Samuel V. Scarpino, Cecile Viboud, Amy Wesolowski, and Andrew Schroeder. Aggregated mobility data could help fight COVID-19. *Science (New York, N.Y.)*, 368(6487):145–146, 2020.

[8] Max S. Y. Lau, Benjamin Douglas Dalziel, Sebastian Funk, Amanda McClelland, Amanda Tiffany, Steven Riley, C. Jessica E. Metcalf, and Bryan T. Grenfell. Spatial and temporal dynamics of superspreading events in the 2014–2015 West Africa Ebola epidemic. *Proceedings of the National Academy of Sciences*, 114(9):2337–2342, February 2017. Publisher: National Academy of Sciences Section: Biological Sciences.

[9] Max S. Y. Lau, Gavin J. Gibson, Hola Adrakey, Amanda McClelland, Steven Riley, Jon Zelner, George Streftaris, Sebastian Funk, Jessica Metcalf, Benjamin D. Dalziel, and Bryan T. Grenfell. A mechanistic spatio-temporal framework for modelling individual-to-individual transmission—With an application to the 2014-2015 West Africa Ebola outbreak. *PLOS Computational Biology*, 13(10):e1005798, October 2017. Publisher: Public Library of Science.

[10] Max S. Y. Lau, Alexander D. Becker, Hannah M. Korevaar, Quentin Caudron, Darren J. Shaw, C. Jessica E. Metcalf, Ottar N. Bjørnstad, and Bryan T. Grenfell. A competing-risks model explains hierarchical spatial coupling of measles epidemics en route to national elimination. *Nature Ecology & Evolution*, pages 1–6, April 2020. Publisher: Nature Publishing Group.

[11] Moritz U. G. Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M. Pigott, Open COVID-19 Data Working Group†, Louis du Plessis, Nuno R. Faria, Ruoran Li, William P. Hanage, John S. Brownstein, Maylis Layan, Alessandro Vespignani, Huaiyu Tian, Christopher Dye, Oliver G. Pybus, and Samuel V. Scarpino. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 368(6490):493–497, May 2020. Publisher: American Association for the Advancement of Science Section: Research Article.

[12] Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*, March 2020. Publisher: American Association for the Advancement of Science Section: Research Article.

[13] Paige Maas. Facebook Disaster Maps: Aggregate Insights for Crisis Response & Recovery. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 3173–3173, Anchorage AK USA, July 2019. ACM.

[14] Stephen A. Lauer, Kyra H. Grantz, Qifang Bi, Forrest K. Jones, Qulu Zheng, Hannah R. Meredith, Andrew S. Azman, Nicholas G. Reich, and Justin Lessler. The Incubation Period of Coronavirus

medRxiv preprint doi: https://doi.org/10.1101/2020.06.20.20130476; this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*, 172(9):577–582, March 2020. Publisher: American College of Physicians.

[15] Thomas V. Inglesby. Public Health Measures and the Reproduction Number of SARS-CoV-2. *JAMA*, 323(21):2186–2187, June 2020. Publisher: American Medical Association.

[16] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, November 2005. Number: 7066 Publisher: Nature Publishing Group.

[17] Ellen Barry. Days After a Funeral in a Georgia Town, Coronavirus 'Hit Like a Bomb'. *The New York Times*, March 2020.

[18] Kaiyuan Sun, Jenny Chen, and Cécile Viboud. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. *The Lancet Digital Health*, 2(4):e201–e208, April 2020. Publisher: Elsevier.

[19] Eunha Shim, Amna Tariq, Wongyeong Choi, Yiseul Lee, and Gerardo Chowell. Transmission potential and severity of COVID-19 in South Korea. *International Journal of Infectious Diseases*, 93:339–344, April 2020. Publisher: Elsevier.

[20] US Census Bureau. 2010 Census Urban and Rural Classification and Urban Area Criteria, 2010. Library Catalog: www.census.gov Section: Government.

[21] Robert Verity, Lucy C. Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick G. T. Walker, Han Fu, Amy Dighe, Jamie T. Griffin, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Anne Cori, Zulma Cucunubá, Rich FitzJohn, Katy Gaythorpe, Will Green, Arran Hamlet, Wes Hinsley, Daniel Laydon, Gemma Nedjati-Gilani, Steven Riley, Sabine van Elsland, Erik Volz, Haowei Wang, Yuanrong Wang, Xiaoyue Xi, Christl A. Donnelly, Azra C. Ghani, and Neil M. Ferguson. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*, 20(6):669–677, June 2020. Publisher: Elsevier.