Crowding and the epidemic intensity of COVID-19 transmission

Benjamin Rader[1,2], Samuel V. Scarpino[3,$], Anjalika Nande[4], Alison L. Hill[4], Benjamin D. Dalziel[5,6], Robert C. Reiner[7,8], David M. Pigott[7,8], Bernardo Gutierrez[9,10], Munik Shrestha[3], open COVID-19 data working group[#], John S. Brownstein[1,11], Marcia C. Castro[12], Huaiyu Tian[13], Bryan T. Grenfell[14,15], Oliver G. Pybus[9,16,$], C. Jessica E. Metcalf[14,15], Moritz UG Kraemer[1,9,11,$]

1. Computational Epidemiology Lab, Boston Children's Hospital, Boston, United States
2. Department of Epidemiology, Boston University School of Public Health, Boston, United States
3. Network Science Institute, Northeastern University, Boston, United States
4. Program for Evolutionary Dynamics, Harvard University, Cambridge, United States
5. Department of Integrative Biology, Oregon State University, Corvallis, United States
6. Department of Mathematics, Oregon State University, Corvallis, United States
7. Department of Health Metrics, University of Washington, Seattle, United States
8. Institute for Health Metrics and Evaluation, University of Washington, Seattle, United States
9. Department of Zoology, University of Oxford, Oxford, United Kingdom
10. School of Biological and Environmental Sciences, Universidad San Francisco de Quito USFQ, Quito, Ecuador
11. Harvard Medical School, Boston, United States
12. Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, United States
13. State Key Laboratory of Remote Sensing Science, College of Global Change and Earth System Science, Beijing Normal University, Beijing, China
14. Department of Ecology and Evolutionary Biology, Princeton University, Princeton, United States
15. Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, United States
16. Department of Pathobiology and Population Science, The Royal Veterinary College, London, United Kingdom

[$]correspondence should be addressed to moritz.kraemer@zoo.ox.ac.uk and s.scarpino@northeastern.edu and oliver.pybus@zoo.ox.ac.uk

[#]Members of the working group are listed at the end of the manuscript

Summary

The COVID-19 pandemic is straining public health systems worldwide and major non-pharmaceutical interventions have been implemented to slow its spread[1–4]. During the initial phase of the outbreak the spread was primarily determined by human mobility[5,6]. Yet empirical evidence on the effect of key geographic factors on local epidemic spread is lacking[7]. We analyse highly-resolved spatial variables for cities in China together with case count data in order to investigate the role of climate, urbanization, and variation in interventions across China. Here we show that the epidemic intensity of COVID-19 is strongly shaped by crowding, such that epidemics in dense cities are more spread out through time, and denser cities have larger total incidence. Observed differences in epidemic intensity are well captured by a metapopulation model of COVID-19 that explicitly accounts for spatial hierarchies. Densely-populated cities worldwide may experience more prolonged epidemics. Whilst stringent interventions can shorten the time length of these local epidemics, although these may be difficult to implement in many affected settings.

Main text

Predicting the epidemiology of the COVID-19 pandemic is a central priority for guiding epidemic responses around the world. China has undergone its first epidemic wave and, remarkably, cities across the country are now reporting few or no locally-acquired cases[8]. Analyses have indicated that that the spread of COVID-19 from Hubei to the rest of China was driven primarily by human mobility[6] and the stringent measures to restrict human movement and public gatherings within and among cities in China have been associated with bringing local epidemics under control[5]. Key uncertainties remain as to which geographic factors drive local transmission dynamics and affect the intensity of transmission of COVID-19. For respiratory pathogens, "epidemic intensity" (*i.e.,* the peakedness of the number of cases through time, or the shortest period during which the majority of cases are observed) varies with increased indoor crowding, and socio-economic and climatic factors[9–14]. Epidemic intensity is minimized when incidence is spread evenly across weeks and increases as incidence becomes more focused in particular days (**Figure 1C**, see a detailed description of how epidemic intensity is defined in Ref. [9]). In any given location, higher epidemic intensity requires a larger surge capacity in the public health system[15], especially for an emerging respiratory pathogen such as COVID-19[16].

China provides richly detailed epidemiological time series[2,17,18] across a wide range of geographic contexts, hence the epidemic there provides an opportunity to evaluate the role of factors in shaping the intensity of local epidemics. We use detailed line-list data from Chinese cities[19,20], climate and population data, local human mobility data from Baidu, and timelines of outbreaks responses and interventions to

identify drivers of local transmission in Chinese cities, with a focus on epidemic intensity among provinces in China.

To explore the impact of urbanization, temperature, and humidity, we used daily incidence data of confirmed COVID-19 cases (date of onset) aggregated at the prefectural level (n = 293) in China. Prefectures are administrative units that typically have one urban center (**Figure 1**). We aggregate individual level data that were collected from official government reports[18]. Epidemiological data in each prefecture were truncated to exclude dates before the first and after the last day of reported cases. The shape of epidemic curves varied between prefectures with some showing rapid rises and declines in cases and others showing more prolonged epidemics (**Figure 1A**). We estimate epidemic intensity for each prefecture from these data by calculating the inverse Shannon entropy of the distribution of incident cases[9]. We define the incidence distribution p_{ij} for a given city to be the proportion of COVID-19 cases during epidemic wave j that occurred on day i. The inverse Shannon entropy of incidence for a given prefecture and year is then given by $v_j = \left(-\sum_i p_{ij} \log p_{ij}\right)^{-1}$. Because v_j is a function of the disease incidence curve in each location, rather than of absolute incidence values, it is invariant under differences in overall reporting rates among cities or overall attack rates. Population counts for each prefecture were extracted from a 1 km x 1 km gridded surface of the world utilizing administrative-2 level cartographic boundaries.

Within each prefecture, we calculate Lloyd's index of mean crowding[9,21] treating the population count of each pixel as an individual unit (**Methods**, **Figure 1B and C**). The term 'mean crowding' used here is a specific metric that summarizes both, population density and how density is distributed across a prefecture (patchiness). Values on the resulting index above the mean pixel population count within each prefecture suggest a spatially-aggregated population structure (**Methods**). For example, Guangzhou has high values of crowding whilst Quzhou which has a more evenly distributed population in its prefecture (**Figure 1B and C**). Using the centroid of each prefecture we calculate daily mean temperature and specific humidity; these values are subsequently averaged over each prefecture's reporting period (**Methods**). We performed log-linear regression modeling to determine the association between epidemic intensity with the socio-economic and environmental variables (**Methods**).

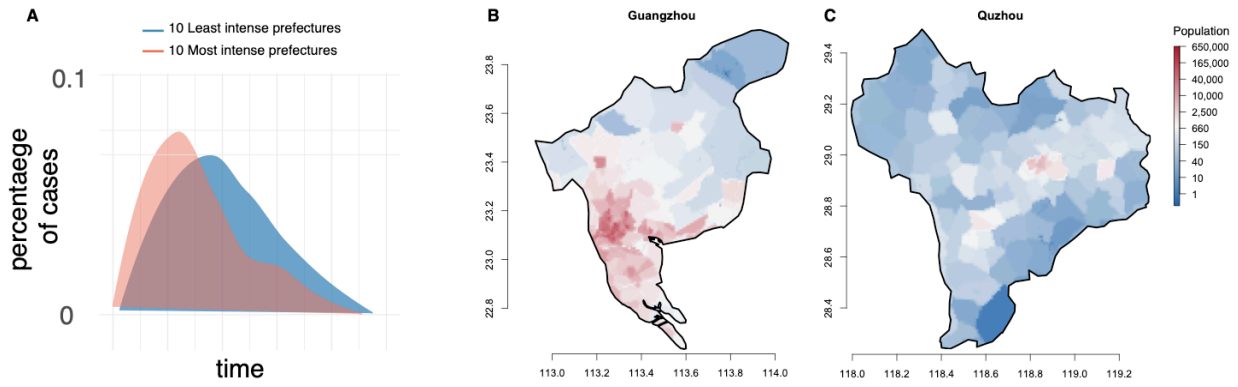


Figure 1: Maps of crowding in prefectures in China. *A) shows epidemic curves that are normalized to show the percentage of cases that are occurring at each given day. The 10 most intense (red) prefectures are shown versus the 10 least intense (blue). B) An example of a prefecture with high levels of crowding (Guangzhou, Guangdong Province), versus (C) a prefecture where populations are more equally distributed across the prefecture (Quzhou, Zhejiang Province). The colour scale illustrates the number of inhabitants per grid cell (1km x 1km).*

We found that epidemic intensity is significantly negatively correlated with mean population crowding and varies widely across the country (**Figure 2**, Extended Data Table 1, p-value < 0.001). Our observation contrasts those expected from simple and classical epidemiological models where it would be expected to see more intensity in crowded areas[22,23]. We hypothesize that the mechanism that underlies the more crowded cities experience less intense outbreaks because crowding enables more widespread and sustained transmission between households leading incidence to be more widely distributed in time (see section below for detailed simulation, **Methods**). Population size, mean temperature, and mean specific humidity were all significant but their correlation coefficients were much smaller (**Extended Data Table 1**). A multivariate-model was able to explain a large fraction of the variation in epidemic intensity across Chinese cities ($R^2 = 0.54$). We perform sensitivity analysis to account for potential noise in the city level incidence distribution (Extended Data Fig. 1).

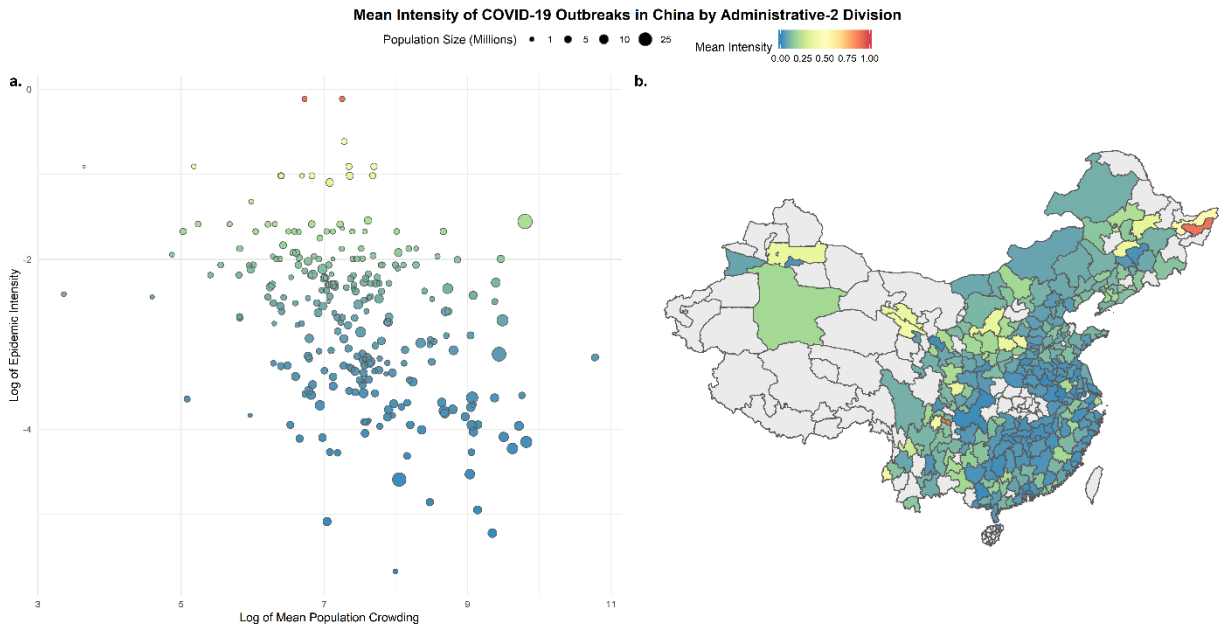


Figure 2: Crowding and the intensity of transmission of COVID-19 in China. a) *negative association between log of epidemic intensity, as measured by inverse Shannon entropy (Methods), and log population crowding, as measure by Lloyd's mean crowding (Methods). Lower intensity and therefore prolonged epidemics in larger cities. The size of the points indicate the size of the population in each city,* ***b)*** *Map of epidemic intensity in China at the prefecture level. Darker colours indicate lower intensity and lighter colours higher intensity. Grey prefectures had not enough reported cases, no cases or were not included in analyses (Hubei Province).*

One key uncertainty in previous applications of models of epidemic intensity was the contribution of disease importation(s) on the shape of the epidemic[9]. Due to the unprecedented scale of human mobility restrictions imposed in China, the fact that the early epidemic was effectively from a single source, coupled with the availability of real-time data on mobility, we can evaluate the impact of these restrictions on the epidemic intensity relative to the local dynamics. To do so, we performed a univariate analysis (**Extended Data Table 1**) and found that human mobility explained 14% of the variation in epidemic intensity. This further supports earlier findings that COVID-19 had already spread throughout much of China prior to the cordon sanitaire of Hubei province and that the pattern of seeding potentially modulates epidemic intensity[6,24]. These findings are also in agreement with previous work on other pathogens (measles, influenza) which showed that once local epidemics are established case importation becomes less important in determining epidemic intensity[25].

To evaluate the potential impact of variability of intensity on the peak incidence and total incidence we performed a simple linear regression. We found that peak incidence was correlated with epidemic intensity (locations that had high intensity also had more cases at the peak). Total incidence, however, was larger in areas with lower estimated intensity, which is intuitive as crowded areas have longer epidemics that affect more people (**Extended Data Table 2**). This suggests that measures taken to mitigate the epidemic may need to be enforced more strictly in smaller cities to lower the peak incidence (flatten the curve) but conversely may not need to be implemented as long. Furthermore, with lower total incidence in small cities, the risk of resurgence may be elevated due to lower population immunity. There is urgent need to collect serological evidence to provide a full picture of attack rates across the world[26].

Using our model trained on cities in China we extrapolated epidemic intensity to cities across the world (**Figure 3**). Figure 3 shows the distribution of epidemic intensity in 380 urban centers. Cities in yellow are predicted to have higher epidemic intensity relative to those in blue (a full list is provided in **Extended Data Table 3**). Small inland cities in sub-Saharan Africa had high predicted epidemic intensity and may be particularly prone to experience large surge capacity in the public health system[27]. In general, coastal cities had lower predicted intensity and larger and more prolonged predicted epidemics. Global predictions of epidemic intensity in cities rely on fitted relationships of the first epidemic curve from Chinese prefectures and therefore need to be interpreted with extreme caution.

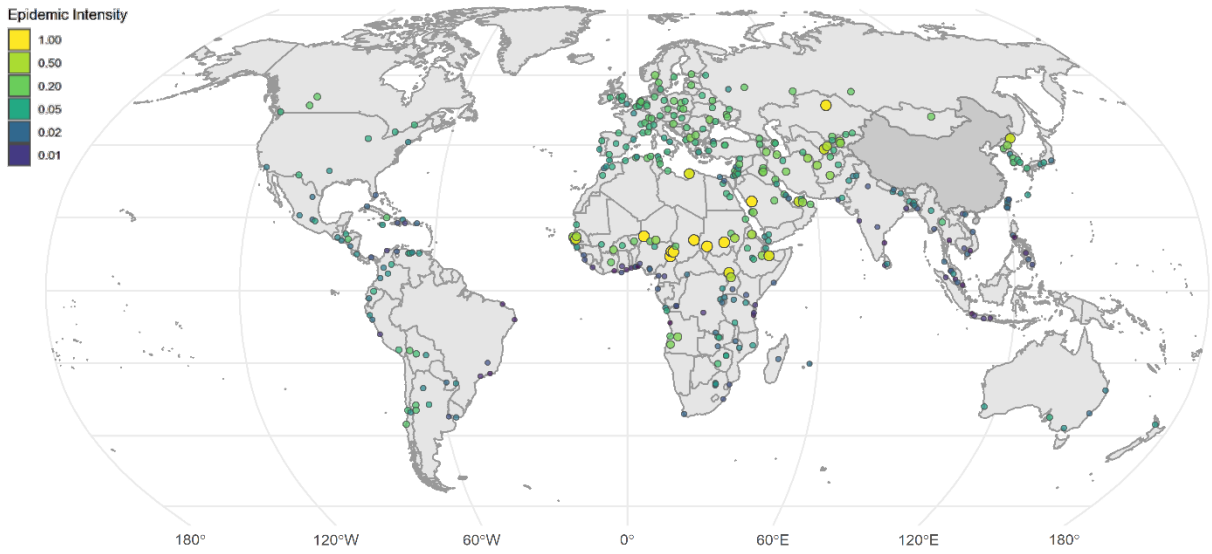


Figure 3: Predicted epidemic intensities vary across 380 global cities. *Darker colours represent low epidemic intensity and lighter colours represent high epidemic intensity. Estimates were generated using the full model (Model 5) fitted to epidemic curves in Chinese cities (Extended Data Table 1). A full list of*

epidemic intensities can be found in the Extended Data Table 3. Epidemic intensity is a measure of peakedness of epidemics and does not reflect the expected number of cases (Methods).

To understand the mechanism responsible for our finding that outbreaks in crowded cities were lower intensity (*i.e.* more spread out in time), we simulated stochastic epidemic dynamics in different types of populations. Simple, well-mixed transmission models where contact rates are higher in crowded regions were not consistent with our findings, since they predict crowded regions would have more intense and higher-peaked outbreaks. To capture more realistic contact patterns, we created hierarchically-structured populations[28] where individuals had high rates of contact within their households (households are defined broadly and could represent care homes, hospitals, prisons, etc.), lower rates with individuals from other households but within the same "neighborhoods", and relatively rare contact with other individuals in the same prefecture (**Figure 4A**). Assumptions are consistent with reports that the majority of onward transmission occurred in households[2,29]. We assumed spread between prefectures was negligible once an outbreak started. In this scenario, "sparse" prefectures often had more intense, short-term outbreaks that were isolated to certain neighborhoods, while "crowded" prefectures could have drawn-out, low intensity outbreaks that jumped between the more highly-connected "neighborhoods" (**Figures 4B and C**). These outbreaks had larger final size than those in less-crowded areas (**Figure 4C**) which likely is related to large overdispersion in the reproduction number of COVID-19[30,31] where local outbreaks can reach their full potential due to the availability of contacts. We also considered outbreak dynamics in sparse and crowded prefectures under strong social distancing measures, which is likely to be the scenario occurring across China during most of the time captured by our study and certainly after January 23, 2020[2]. If social distancing reduces non-household contacts by the same relative amount in all prefectures, there will be more contacts remaining in crowded areas, since baseline contact rates are higher. In this case, it may take much longer for the infection to die out post-intervention in crowded areas (**Figure 4D**), leading to a lower intensity outbreak with larger final size, as seen in our data (**Figure 1C**).

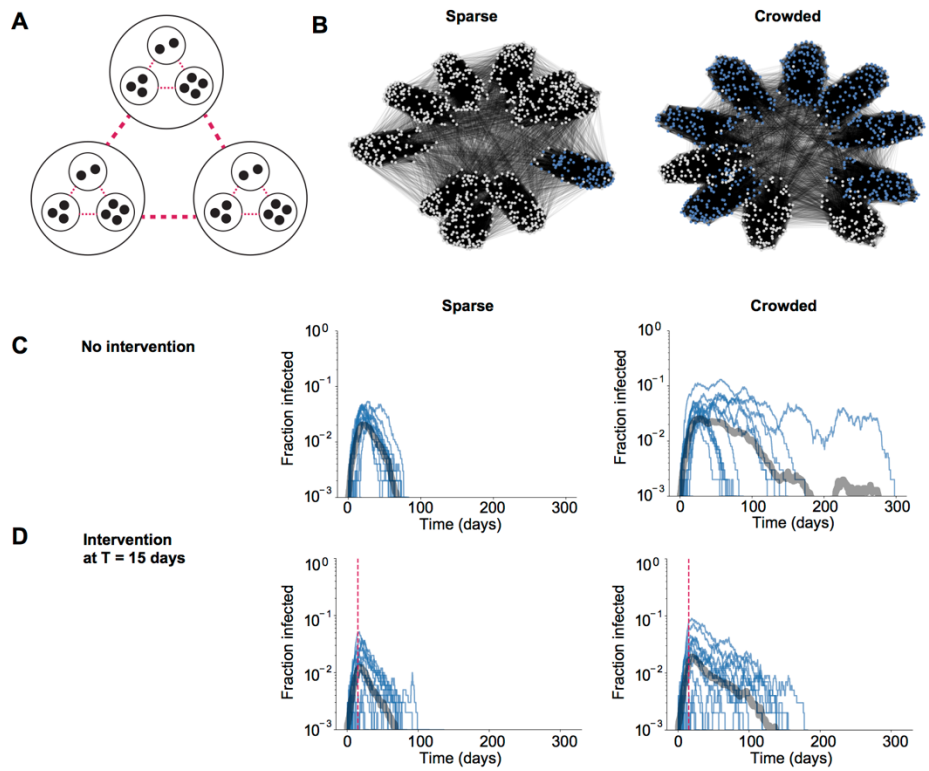


Figure 4: Mechanisms generating less intense epidemics in crowded populations. *A) Schematic of a hierarchically-structured population model consisting of households and "neighborhoods" within a prefecture. Transmission is more likely among contacts connected at lower spatial levels. Crowded populations have stronger connections outside the household, and interventions reduce the strength of these connections in both populations (pink lines). B) - C) Simulated outbreak dynamics in the absence of interventions in crowded vs sparse populations. For the networks in (B), blue nodes are individuals who were eventually infected by the end of the outbreak. In (C), individual realizations are shown with thin blue lines and the average in the thick grey line. D) Simulated outbreak dynamics in the presence of strong social distancing measures in crowded vs. sparse populations. The intervention was implemented at day 15 (pink line) and led to a 75% reduction in contacts.*

Spatial covariates and particularly crowding are important parameters to consider in the assessment of epidemics across the world. Crowded cities tend to be more prolonged due to increased crowding and the higher potential for transmission chains to persist (*i.e.,* in denser environments there is higher potential for two randomly selected hosts in a population to attain spatiotemporal proximity sufficient for COVID-19 transmission). Indeed, that epidemic intensity is higher in comparatively low density areas is consistent with observations in the most affected areas in Italy (e.g., Bergamo)[32]. Our findings confirm previous work on epidemic intensity of transmission of influenza in cities[9] albeit by a different mechanism:

influenza is likely driven by the accumulation of immunity rather than the specific network structure of individuals. More generally, our work provides empirical support for the role of spatial organization in determining infectious disease dynamics and the limited capacity of *cordon sanitaires* to control local epidemics[28,33]. We were unable to test more specific hypotheses about which interventions may have impacted the intensity of transmission within and between cities. Further, even though humidity was negatively associated with epidemic intensity it did not explain the majority of the variation and more work will be needed to find causal evidence for the effect of humidity on transmission dynamics of COVID-19. Therefore, maps showing epidemic intensity in cities outside China (**Figure 3**) should be interpreted with particular caution.

Currently, non-pharmaceutical interventions are the primary control strategy for COVID-19. As a result, public health measures are often focused on 'flattening the curve' to lower the risk of essential services running out of capacity. We show that spatial context, especially crowding, can result in a higher risk of intensive epidemics in less crowded, comparatively rural or suburban areas. Therefore, it will be critical to view non-pharmaceutical interventions through the perspective of crowding (*i.e.*, how does an intervention reduce the circle of contacts of an average individual) in cities across the world. Specifically, cities in sub-Saharan Africa have high predicted epidemic intensities that will likely overwhelm already stressed health care systems.

References

1. Fraher, E. P. *et al.* Ensuring and Sustaining a Pandemic Workforce. *N. Engl. J. Med.* NEJMp2006376 (2020). doi:10.1056/NEJMp2006376
2. Leung, K., Wu, J. T., Liu, D. & Leung, G. M. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. *Lancet* **6736**, (2020).
3. Ji, Y., Ma, Z., Peppelenbosch, M. P. & Pan, Q. Potential association between COVID-19 mortality and health-care resource availability. *Lancet Glob. Heal.* **8**, e480 (2020).
4. Rosenbaum, L. Facing Covid-19 in Italy — Ethics, Logistics, and Therapeutics on the Epidemic's Front Line. *N. Engl. J. Med.* NEJMp2005492 (2020). doi:10.1056/NEJMp2005492
5. Tian, H. *et al.* An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* eabb6105 (2020). doi:10.1126/science.abb6105
6. Kraemer, M. U. G. *et al.* The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **21**, eabb4218 (2020).
7. Lipsitch, M., Swerdlow, D. L. & Finelli, L. Defining the Epidemiology of Covid-19 — Studies

Needed. *N. Engl. J. Med.* **382**, 1194–1196 (2020).

8. World Health Organization (WHO). Coronavirus disease 2019 (COVID-19) Situation Report - 71. (2020).

9. Dalziel, B. D. *et al.* Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities. *Science* **362**, 75–79 (2018).

10. Shaman, J., Pitzer, V. E., Viboud, C., Grenfell, B. T. & Lipsitch, M. Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol.* **8**, (2010).

11. Gog, J. R. *et al.* Spatial transmission of 2009 pandemic influenza in the US. *PLoS Comput. Biol.* **10**, e1003635 (2014).

12. Shaman, J. & Kohn, M. Absolute humidity modulates influenza survival, transmission, and seasonality. *Proc. Natl. Acad. Sci.* **106**, 3243–3248 (2009).

13. Chetty, R. *et al.* The association between income and life expectancy in the United States, 2001-2014. *JAMA - J. Am. Med. Assoc.* **315**, 1750–1766 (2016).

14. Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H. & Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **21**, 1–9 (2020).

15. Crawford, J. M. *et al.* Laboratory Surge Response to Pandemic (H1N1) 2009 Outbreak, New York City Metropolitan Area, USA. *Emerg. Infect. Dis.* **16**, 8–13 (2010).

16. Grasselli, G., Pesenti, A. & Cecconi, M. Critical Care Utilization for the COVID-19 Outbreak in Lombardy, Italy. *JAMA* **19**, NEJMoa2002032 (2020).

17. Li, Q. *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N. Engl. J. Med.* NEJMoa2001316 (2020). doi:10.1056/NEJMoa2001316

18. Xu, B. *et al.* Epidemiological data from the COVID-19 outbreak, real-time case information. *Sci. Data* **7**, (2020).

19. Xu, B. *et al.* Epidemiological data from the COVID-19 outbreak, real-time case information. *figshare* (2020). doi:10.6084/m9.figshare.11949279

20. Xu, B. & Kraemer, M. U. G. Open access epidemiological data from the COVID-19. *Lancet Infect. Dis.* **3099**, 30119 (2020).

21. Lloyd, M. `Mean Crowding'. *J. Anim. Ecol.* **36**, 1 (1967).

22. May, R. M. & Anderson, R. M. Spatial heterogeneity and the design of immunization programs. *Math. Biosci.* **72**, 83–111 (1984).

23. Anderson, R. M. & May, R. M. *Infectious diseases of humans: dynamics and control*. (Oxford University Press, 1991).

24. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (COVID-19). *Science* **3221**, 2020.02.14.20023127 (2020).

25. Bjørnstad, O. N. & Grenfell, B. T. Hazards, spatial transmission and timing of outbreaks in epidemic metapopulations. *Environ. Ecol. Stat.* **15**, 265–277 (2008).
26. Lipsitch, M., Swerdlow, D. L. & Finelli, L. Defining the Epidemiology of Covid-19 — Studies Needed. *N. Engl. J. Med.* NEJMp2002125 (2020). doi:10.1056/NEJMp2002125
27. Martinez-Alvarez, M. *et al.* COVID-19 pandemic in west Africa. *Lancet Glob. Heal.* **2019**, 2019–2020 (2020).
28. Watts, D. J., Muhamad, R., Medina, D. C. & Dodds, P. S. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proc. Natl. Acad. Sci.* **102**, 11157–11162 (2005).
29. Aylward, Bruce (WHO); Liang, W. (PRC). Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). **2019**, 16–24 (2020).
30. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005).
31. Kucharski, A. J. *et al.* Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* **3099**, 1–7 (2020).
32. Senni, M. COVID-19 experience in Bergamo, Italy. *Eur. Heart J.* 1–2 (2020). doi:10.1093/eurheartj/ehaa279
33. Sattenspiel, L. Simulating the Effect of Quarantine on the Spread of the 1918–19 Flu in Central Canada. *Bull. Math. Biol.* **65**, 1–26 (2003).
34. Ramshaw, R. E. *et al.* A database of geopositioned Middle East Respiratory Syndrome Coronavirus occurrences. *Sci. Data* **6**, 318 (2019).
35. Doxsey-Whitfield, E. *et al.* Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4. *Pap. Appl. Geogr.* **1**, 226–234 (2015).
36. Reiczigel, J., Lang, Z., Rózsa, L. & Tóthmérész, B. Properties of crowding indices and statistical tools to analyse parasite crowding data. *J. Parasitol.* **91**, 245–252 (2005).
37. Wade, M. J., Fitzpatrick, C. L. & Lively, C. M. 50-year anniversary of Lloyd's "mean crowding": Ideas on patchy distributions. *J. Anim. Ecol.* **87**, 1221–1226 (2018).
38. Florczyk, A. *et al.* GHS-UCDB R2019A - GHS Urban Centre Database 2015, multitemporal and multidimensional attributes. *Eur. Comm. Jt. Res. Cent.* (2019).

Methods

Epidemiological data

No officially reported line list was available for cases in China. We use a standardised protocol[34] to extract individual level data from December 1st, 2019 - March 30th, 2020. Sources are mainly official reports from provincial, municipal or national health governments. Data included basic demographics (age, sex), travel histories and key dates (dates of onset of symptoms, hospitalization, and confirmation). Data were entered by a team of data curators on a rolling basis and technical validation and geo-positioning protocols were applied continuously to ensure validity. A detailed description of the methodology is available[18]. Lastly, total numbers were matched with officially reported data from China and other government reports.

Estimating epidemic intensity

Epidemic intensity was estimated for each prefecture by calculating the inverse Shannon entropy of the distribution of COVID-19 cases. Shannon entropy was used to fit time series of other respiratory infections (influenza)[9]. The Shannon entropy of incidence for a given prefecture and year is then given by $v_j = \left(-\sum_i p_{ij} \log p_{ij}\right)^{-1}$. Because v_j is a function of incidence distribution in each location rather than raw incidence it is invariant under differences in overall reporting rates between cities or attack rates. We then assessed how mean intensity $v \propto \sum_j v_j$ varied across geographic areas in China.

Proxies for COVID-19 interventions

Real-time measures of human mobility were extracted from the Baidu Qianxi web platform to estimate the proportion of daily movement between the city of Wuhan to Hubei and 30 other Chinese provinces. Relative mobility volume was available from January 2, 2020 to January 25, 2020 and averaged across these dates to create a single measure of relative flows from Wuhan. This data was only available at the province level, so each individual prefecture inherited the relative mobility of its higher-level province. Baidu's mapping service is estimated to have a 30% market share in China and more data can be found[5,6].

Data on drivers of transmission of COVID-19

Prefecture-specific population counts and densities were derived from the 2020 Gridded Population of The World, a modeled continuous surface of population estimated from national census data and the United Nations World Population Prospectus[35]. Population counts are defined at a 30 arc-second resolution (approximately 1 km x 1 km at the equator) and extracted within administrative-2 level cartographic boundaries defined by the National Bureau of Statistics of China. Lloyd's mean crowding,

$\frac{[\sum_i (q_i - 1) q_i]}{\sum_i q_i}$, was estimated for each prefecture where q_i represents the population count of each non-zero pixel within a prefecture's boundary and the resulting value estimates an individual's mean number of expected neighbors[9,36].

Daily temperature (°F), relative humidity (%) and atmospheric pressure (Pa) at the centroid of each prefecture was provided by The Dark Sky Company via the Dark Sky API and aggregated across a variety of data sources. Specific humidity (kg/kg) was then calculated using the R package, `humidity`[12]. Meteorological variables for each prefecture were then averaged across the entirety of the study period.

Statistical analysis

We normalized the values of epidemic intensity between 0 and 1, and for all non-zero values fit a Generalized Linear Model (GLM) of the form:

$$\log(Y_j) \sim \beta_0 + \beta_1 \log(C_j) + \beta_2 q_j + \beta_3 \log(P_j) + \beta_3 f_j + \beta_4 R_j$$

where for each prefecture j, Y is the scaled Shannon-diversity measure of epidemic intensity derived from the COVID-19 time series, C is Lloyd's Index of Mean Crowding[21,37], q is the mean specific humidity over the reporting period in kg/kg, P is the estimated population count and f is the relative population flows from Wuhan to each prefecture's higher level province. To account for the length of the epidemic period in each city we calculate R as the number of reporting days.

Projecting epidemic intensity in cities around the world

We selected 380 urban centers from the European Commission Global Human Settlement Urban Centre Database and their included cartographic boundaries[38]. To ensure global coverage, up to the five most populous cities in each country were selected from the 1,000 most populous urban centers recorded in the database. Population count, crowding, and meteorological variables were then estimated following identical procedures used to calculate these variables in the Chinese prefectures. Weather measurements were averaged over the 2-month period starting on February 1, 2020.

The parameters from the model of epidemic intensity predicted by humidity, crowding and population size (see Table 1, Model 6) were used to estimate relative intensity in the 380 urban centers. Predicted values of epidemic intensity that fell outside the original covariate space [0,1] (n=7) were set to 1. A full list of predicted epidemic intensities can be found in the Supplementary Information.

372

373 Sensitivity analyses

374 The inverse Shannon entropy metric may be sensitive to noise in incidence distribution. For example, the
375 noisier the incidence distribution the higher the epidemic intensity. To the extent that noise is elevated in
376 small populations (due to demographic stochasticity for instance) intensity also tends to be higher in
377 smaller populations, even if they have the same underlying shape to their epidemic curve. Lloyd's mean
378 crowding also varies strongly with population size. Therefore, some of the observed relationship between
379 intensity and crowding may be due to (possibly independent) statistical scaling of both intensity and
380 crowding with population size. We therefore perform sensitivity analysis to test if cities that are more
381 crowded than expected for their size have lower epidemic intensities than expected for their size. We
382 calculate 'excess intensity' as the residuals on a regression of log(epidemic intensity) ~ log(pop); 'excess
383 crowing' as the residuals on a regression of log(crowding) ~ log(pop) and plot the relationship between
384 excess intensity and excess crowding' (Extended Data Figure 1).

385

386 Simulating epidemic dynamics

387 We simulated a simple stochastic SIR model of infection spread on weighted networks created to
388 represent hierarchically-structured populations. Individuals were first assigned to households using the
389 distribution of household sizes in China (data from UN Population Division, mean 3.4 individuals).
390 Households were then assigned to "neighborhoods" of ~100 individuals, and all neighborhood members
391 were connected with a lower weight. A randomly-chosen 10% of individuals were given "external"
392 connections to individuals outside the neighborhood. The total population size was N=1000. Simulations
393 were run for 300 days and averages were taken over 20 iterations. The SIR model used a per-contact
394 transmission rate of β=0.15/day and recovery rate γ=0.1/day. For the simulations without interventions,
395 the weights were $w_{HH} = 1$, $w_{NH} = 0.01$, and $w_{EX} = 0.001$ for the "crowded" prefecture and $w_{EX} = 0.0001$ for
396 the "sparse" prefecture. For the simulations with interventions, the household and neighborhood weights
397 were the same but we used $w_{EX} = 0.01$ for the "crowded" prefecture and $w_{EX} = 0.001$ for the "sparse"
398 prefecture. The intervention reduced the weight of all connections outside the household by 75%.

399

400

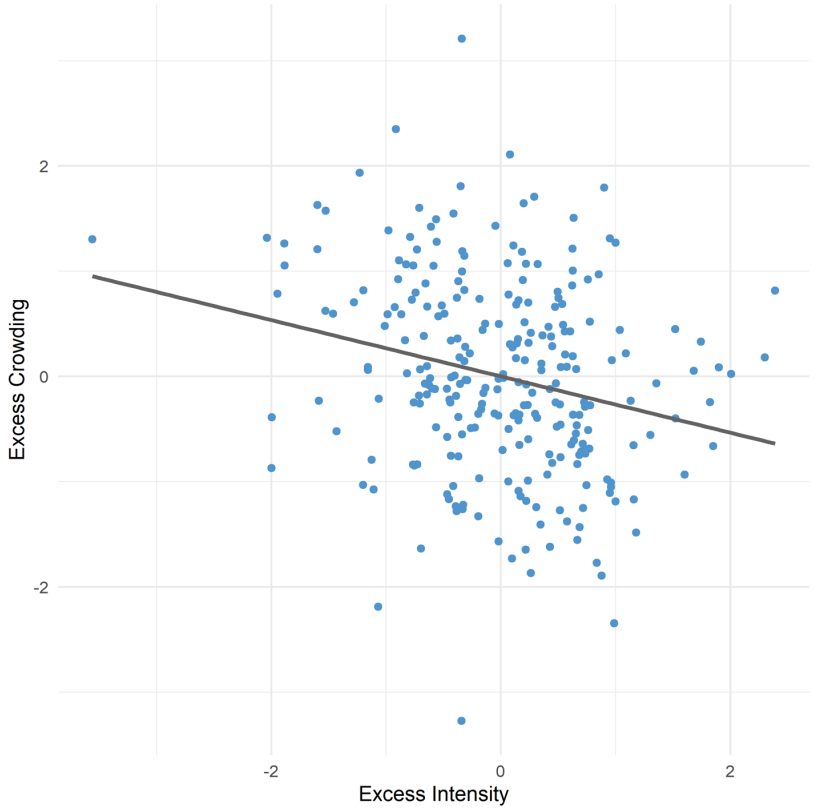
Acknowledgements: The authors thank Kathryn Cordiano for her statistical assistance. BR acknowledges funding from Google.org. MUGK acknowledges funding from European Commission H2020 program (MOOD project) and a Branco Weiss Fellowship. OGP and HT acknowledge funding from the Oxford Martin School. ALH and AN acknowledge funding from the US National Institutes of Health (DP5OD019851). The funding bodies had no role in study design, data collection and analysis, preparation of the manuscript, or the decision to publish. All authors have seen and approved the manuscript.

Author contributions MUGK, SVS, CJEM, OGP conceived the research. BR, ALH, AN, BA, SVS, MUGK analysed the data. MUGK wrote the first draft of the manuscript. All authors contributed to interpretation of results and manuscript writing.

Competing interests: The authors declare no competing interests.

Data availability: We collated epidemiological data from publicly available data sources (news articles, press releases and published reports from public health agencies) which are described in full here[18]. All the epidemiological information that we used is documented in the main text, the extended data, and supplementary tables.

Code availability: The code is available from this link: tbc and the simulation code is available from here: https://github.com/alsnhll/SIRNestedNetwork



Extended Data Figure 1: *Relationship between excess crowding and excess epidemic intensity.*

Extended Data Table 1: *Regression model results of variables predicting epidemic intensity (log scale).*

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
(Intercept)	-1.75 ***	-2.09 ***	-0.12	-1.91 ***	-2.31 ***	-0.19	0.06
	[-2.03, -1.48]	[-2.28, -1.90]	[-0.84, 0.61]	[-2.13, -1.69]	[-2.47, -2.16]	[-0.86, 0.48]	[-0.53, 0.65]
Temperature (° F)	-0.03 ***						
	[-0.03, -0.02]						
Specific Humidity		-154.31 ***				-127.35 ***	-81.12 ***
		[-195.57, -113.06]				[-164.57, -90.13]	[-117.06, -45.19]
Log Crowding			-0.35 ***			-0.21 ***	-0.18 ***
			[-0.45, -0.25]			[-0.30, -0.11]	[-0.26, -0.09]
Log Population Size				-0.59 ***		-0.36 ***	-0.14 *
				[-0.74, -0.45]		[-0.51, -0.21]	[-0.29, -0.00]
Mobility Flows					-25.23 ***		-12.00 ***
					[-32.98, -17.47]		[-18.62, -5.37]
# of Reporting Days							-0.04 ***
							[-0.05, -0.03]
N	262	262	262	262	262	262	262
AIC	657.84	655.81	659.51	648.64	666.94	595.17	529.32
Pseudo R2	0.18	0.18	0.17	0.21	0.15	0.38	0.54

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

430 ***Extended Data Table 2:*** *Relationship between total incidence, peak incidence and epidemic intensity.*

	Model 1
(Intercept)	-0.42 ***
	[-0.59, -0.24]
Incidence Peak	0.03 ***
	[0.03, 0.04]
Log Total Incidence	-0.89 ***
	[-0.96, -0.82]
N	262
AIC	344.43
Pseudo R2	0.81

*** p < 0.001; ** p < 0.01; * p < 0.05.

431

432 ***Extended Data Table 3:*** *380 global cities and their predicted epidemic intensities.*

433