Reconstruction of Transmission Pairs for COVID-19 in family clusters by Crowdsourcing data

Xiao-Ke XU

Dalian Minzu Univerisity

July 30, 2020

Crowdsourcing data of COVID-19

Starting from mid-January, most of the provincial and urban level health authorities in China began to report detailed information regarding COVID-19 cases.

岳某某，女，42岁，田家庵区洞山街道阳光国际城西区C栋，1月20日由合肥返淮，1月23-30日在凤台县新集镇期间与付某（确诊病例）、岳某（确诊病例）有密切接触史。2月1日，患者出现发热症状，2月4日入朝阳医院隔离治疗，2月5日初步诊断为新型冠状病毒感染的肺炎疑似病例，2月6日确诊，目前患者在定点收治医院隔离治疗，病情平稳。

Patient Yue, female, 42, resident in Building C, XX housing estate, Tianjia'an District, Huainan City, Anhui Province. Arrived in Huainan from Hefei on 20 January. Had close contacts with confirmed cases Fu and Yue in Fengtai County during 23-30 January. Showed fever symptom on 1 February, admitted to Chaoyang Hospital on 4 February, tentatively diagnosed with COVID-19 on 5 February, confirmed on 6 February. Patient is currently quarantined and treated in the destined hospital, with mild clinical symptoms.

Crowdsourcing data of COVID-19 outside Hubei Province

The reports were published via official websites and/or social media platforms.

We collected all of the case disclosure manually.

9936 records, from 27 provincial and 264 urban health authorities in China outside Hubei province, from January 19, 2020 to March 5.

Five categories of detailed information can be extracted from the above disclosure, i.e., demographic information, past mobility trace, exposure to the virus and/or potential contact with known infection(s), the timeline of symptom onset, hospitalisation, infection confirmation, and clinical symptoms.

Encoding process: 30 research assistants worked ten hours a day for three weeks.

Thanks a lot for their hard work!

Data fields and their content formats

Category	Data field	Data format
Demographic information	ID	Province_City-SerialNumber
	Age	Number
	Gender	Male/Female
	Occupation	Text
	Place_Residency	Text (Province_City)
Mobility trace	Place_Departure	Text (Province_City/"local")
	Place_Transit	Text (Province_City)
	Place_Destination	Text (Province_City)
	Arrival_Date	Date
Exposure to virus	Earliest_Possible_Date	Date
	Latest_Possible_Date	Date
	Place_and_Event	Text
	Venue	Text
	With_Whom	Text
	Contact_ID_Relationship	Text (ID(relationship)&. . .)

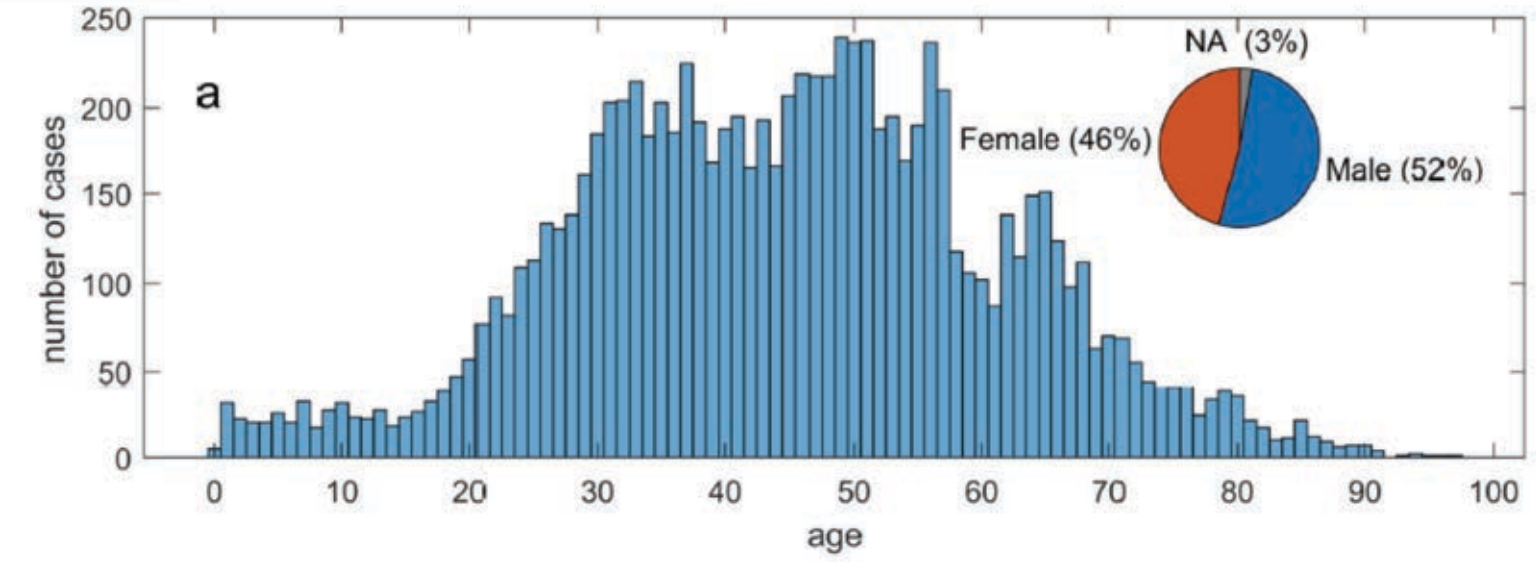
Data fields and their content formats

Epidemiological timelines	Place_Admission	Text (Province_City)
	Method_Discovery	"Self" / "Passive" / "Quarantine"
	Date_Quarantine	Date
	Place_Quarantine	"Quarantine camp"/ "Home" / "Hospital"
	Date_Symptom_Onset	Date
	Date_Hospitalisation	Date
	Place_Hospitalisation	Text
	Date_Confirmation	Date
	Date_Disclose	Date
Clinical symptoms	Symptom	Text
	Symptom category	Text
	Symptom_Severity	"Stable" / "Mild" / "Severe"
Other information	Original_Text_CN	Text
	Original_Text_EN	Text

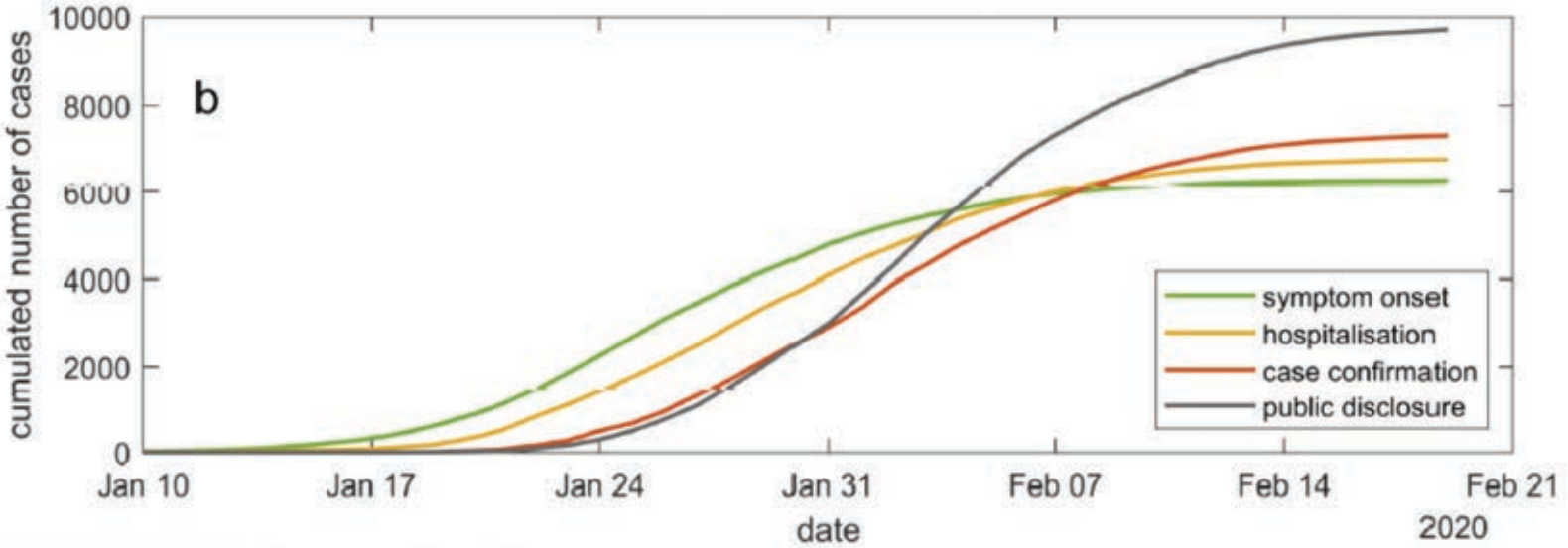
Big data：A large and very accurate case dataset

ID	Age	Gender	Occupation	Place of Residency	Place of Departure	Place of Transit	Place of Destination	Arrival Date
Anhui_Wuhu-1	48	Male	Canteen Boss	Hubei_Wuhan	Hubei_Wuhan	NA	Anhui_Wuhu	1/19/2020
Anhui_Wuhu-10	25	Female	NA	Shaanxi_Xian	Shaanxi_Xian	NA	Anhui_Wuhu	1/22/2020
Anhui_Wuhu-11	37	Female	NA	Anhui_Wuhu	Hubei_Wuhan	NA	Anhui_Wuhu	1/22/2020
Anhui_Wuhu-12	23	Male	NA	NA	Hubei_Wuhan	NA	Anhui_Wuhu	1/23/2020
Anhui_Wuhu-13	57	Male	Farmer	Tianjin	NA	NA	NA	NA
Anhui_Wuhu-14	29	Male	Worker	Hubei_Wuhan	Hubei_Wuhan	NA	Anhui_Wuhu	1/23/2020
Anhui_Wuhu-15	35	Male	NA	Hubei_Wuhan	Hubei_Wuhan	NA	Anhui_Wuhu	1/20/2020
Anhui_Wuhu-16	56	Female	NA	Hubei_Wuhan	Hubei_Wuhan	NA	Anhui_Wuhu	1/22/2020
Anhui_Wuhu-17	34	Male	NA	NA	NA	NA	Anhui_Wuhu	1/26/2020
Anhui_Wuhu-18	50	Female	NA	NA	NA	NA	Anhui_Wuhu	1/26/2020
Anhui_Wuhu-19	36	Male	NA	Hubei_Wuhan	Hubei_Wuhan	NA	Anhui_Wuhu	1/20/2020
Anhui_Wuhu-2	57	Male	Teacher	Hubei_Wuhan	Hubei_Wuhan	NA	Anhui_Wuhu	1/21/2020
Anhui_Wuhu-20	8	Female	NA	NA	Hubei_Wuhan	NA	Anhui_Wuhu	1/22/2020

Statistical validation of several data fields



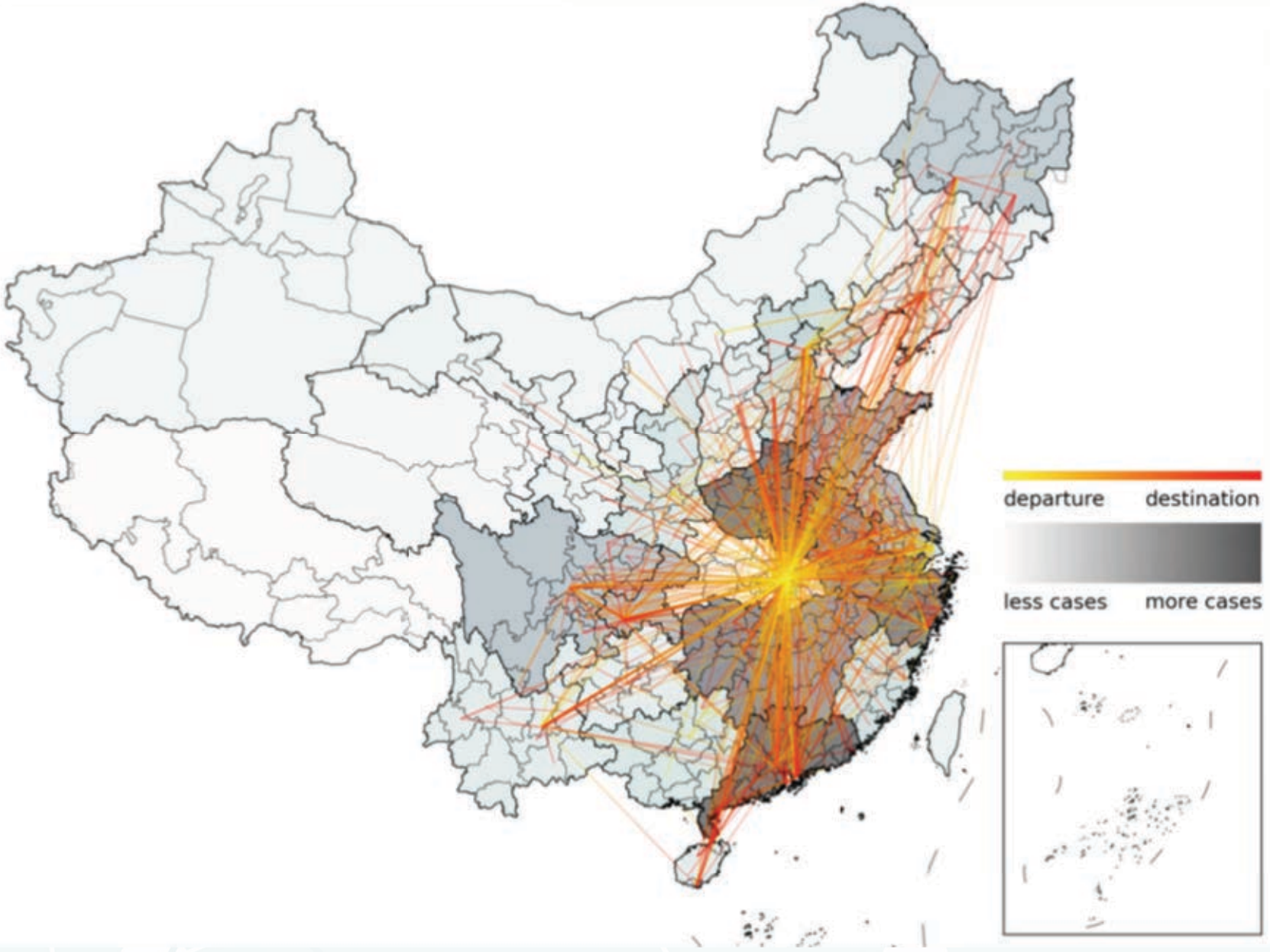
(a) The age and gender distributions of the cases.



(b) The cumulative numbers of symptom onset, hospitalisation, infection confirmation, and public disclosure between 10 January and 20 February.

Cao Wen-Jing, Liu Xiao-Fei, Han Zhuo, Feng Xin, Zhang Lin, Liu Xiao-Fan, **Xu Xiao-Ke**, Wu Ye. Statistical analysis and autoregressive modeling of confirmed coronavirus disease 2019 epidemic cases. Acta Physica Sinica, 2020, 69(9): 090203. (In Chinese)

Mobility traces reported



Lines with gradient color indicate the cases' movements before arriving at the destination cities.

Provinces in China are colored by their numbers of collected case reports.

Hubei Province is not colored.

Possible downstream tasks

- **Reconstruction of transmission events**
- Reconstruction of epidemiological timelines
- Spatial-temporal spreading trajectories
- Emergency monitoring with artificial intelligence

Xiao Fan Liu, **Xiao-Ke Xu***, Ye Wu*, Mobility, exposure, and epidemiological timelines of COVID-19 infections in China outside Hubei Province, **Scientific Data**, submitted

Reconstruction of transmission event and cluster

- Based on 9,936 confirmed COVID-19 cases, we identified 1407 transmission pairs with a strong evidence that **the infectee was infected by the infector**.

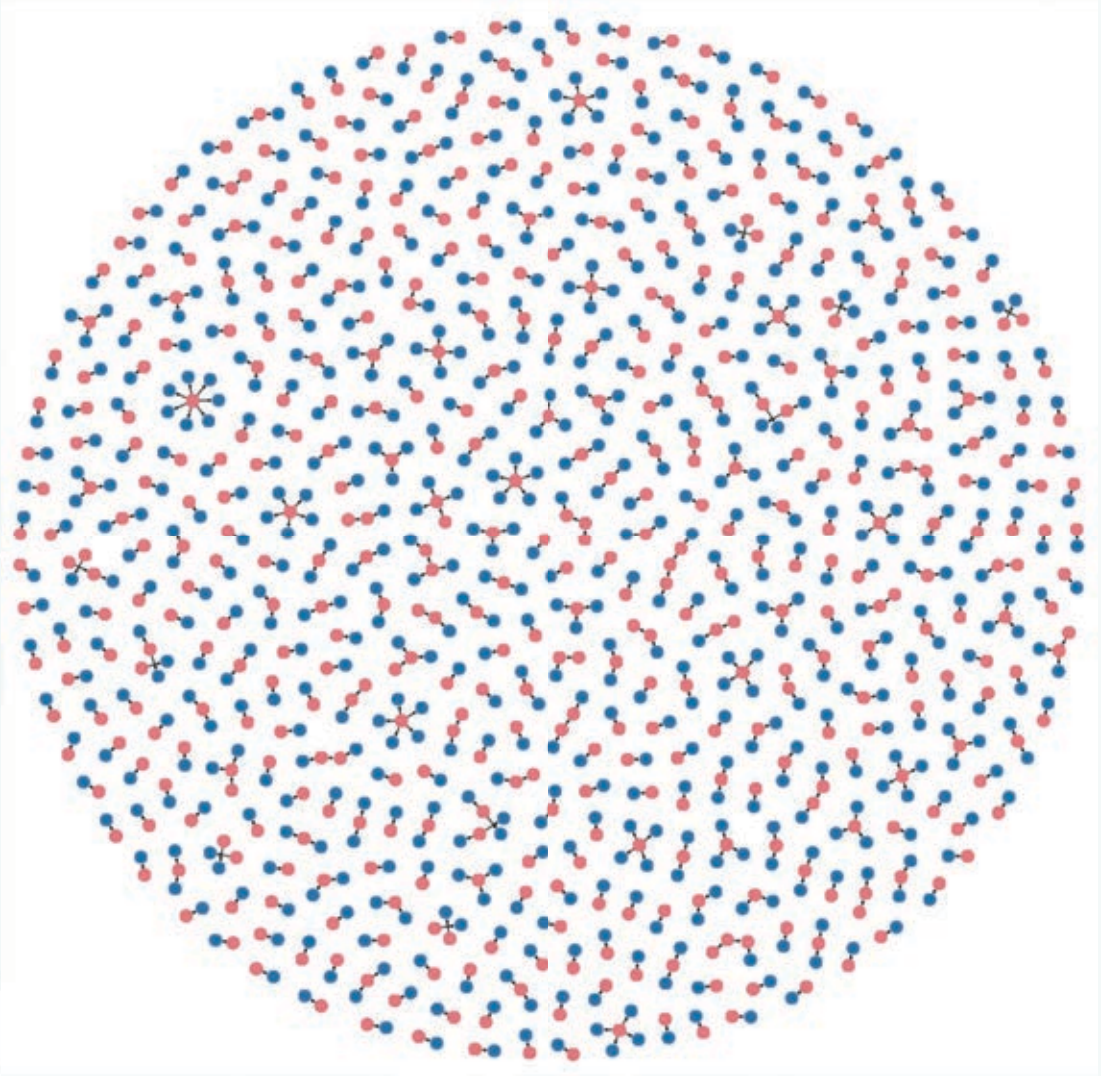
- Each transmission pair is established according to the **travel history**, **social relationships**, and **epidemiological timelines** provided by public case reports, which was cross-validated by two RAs.
- If an infectee had multiple possible exposures, the infector of this transmission pair is considered as the one reported the earliest symptom onset.

- For each transmission pair, we term the infector as “the primary case” and the infectee as “the secondary case”.

- We also consider connected chains of confirmed cases, in which we term the original case the index and the entire chain of cases, the transmission cluster.

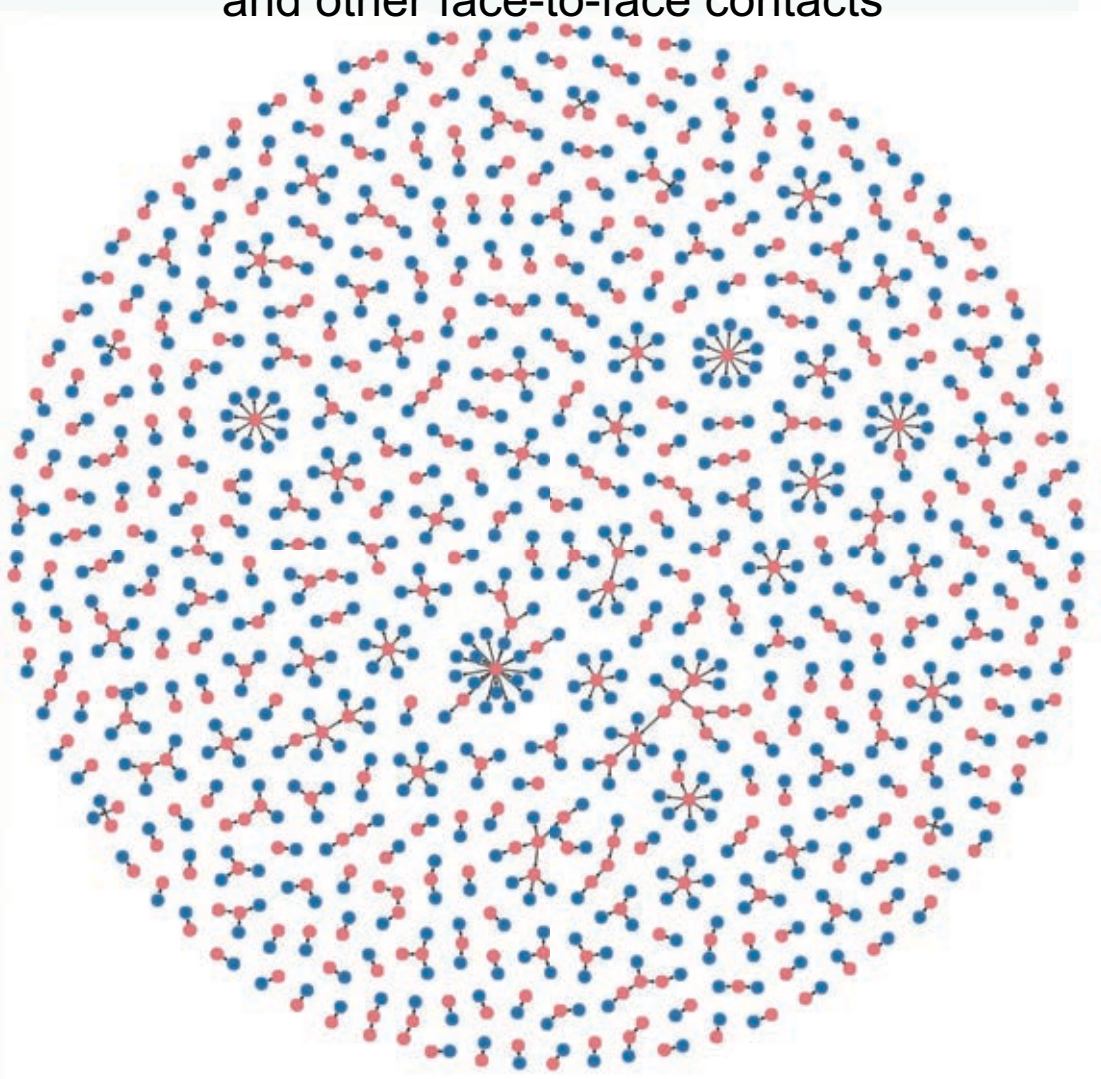
Comparison of transmission clusters in household and non-household

familial members of the same household



household transmission clusters

non-household relatives, colleagues, classmates, friends, and other face-to-face contacts



Non-household transmission clusters

Hazard of infection stratified by **age** for the household relative to non-household transmission

		Secondary cases				
		0~17	18~49	50~64	65+	Total
Primary cases	0~17	0.0	0.8	0.8	1.1	0.7
	18~49	6.3	0.7	0.9	2.0	1.1
	50~64	1.7	0.9	0.7	0.6	0.8
	65+	2.3	1.4	0.6	2.1	1.3
	Total	3.5	0.8	0.8	1.4	1.0

Red or blue shades indicate an increased or decreased hazard of infection within households relative to outside of households, respectively.

The hazard of being infected within households is higher for age groups of young (<18) and elderly (>65) people, whereas the hazard of being infected outside of households is higher for age groups between 18 and 64 years.

Primary cases of elderly (>65) people are more prone to cause household infections.

Hazard of infection stratified by **gender** for the household relative to non-household transmission

Table 1b		Secondary cases		
		Male	Female	Total
Primary cases	Male	0.6	1.6	1.0
	Female	1.2	0.7	0.9
	Total	0.8	1.2	1.0

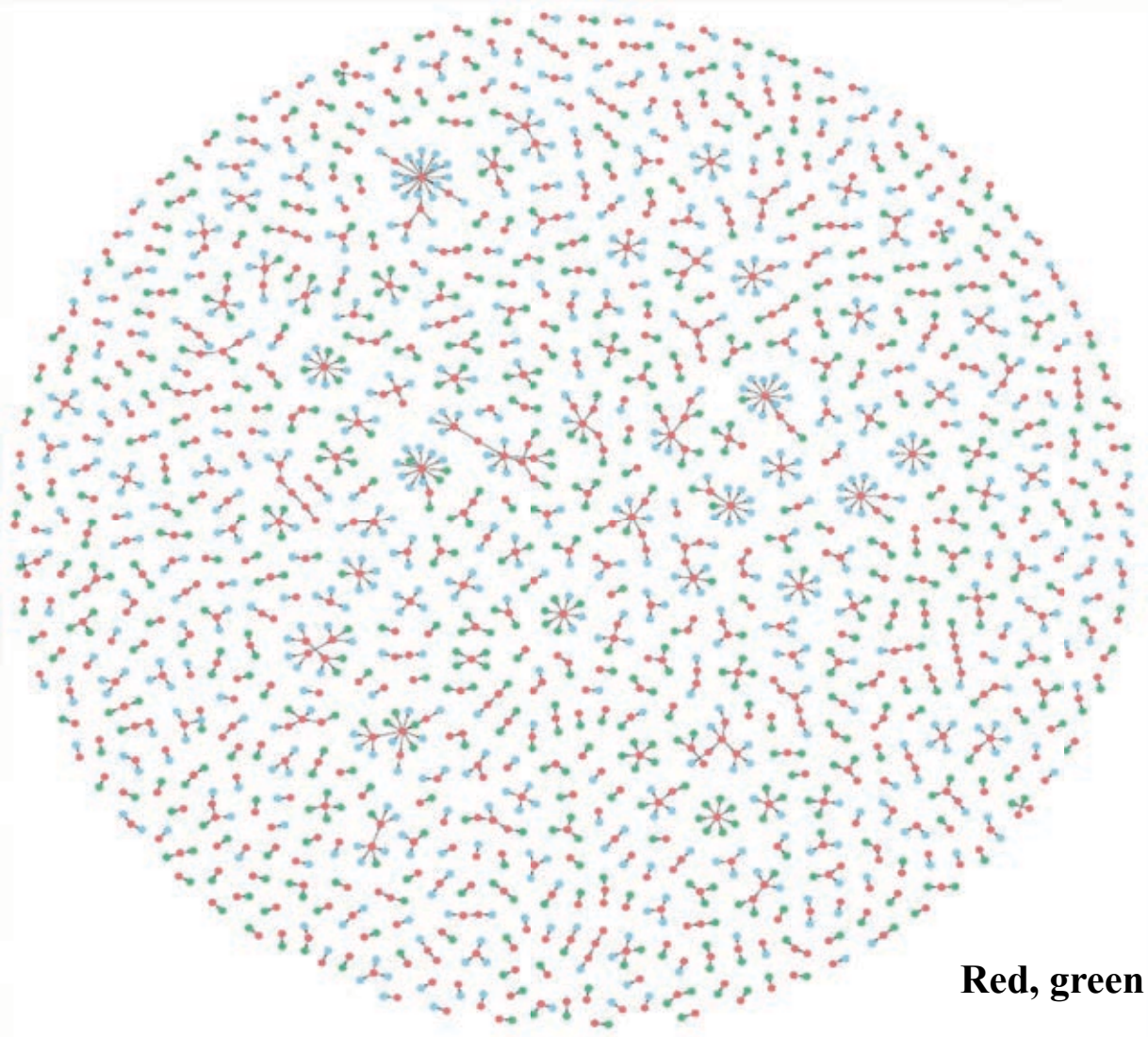
Red or blue shades indicate an increased or decreased hazard of infection within households relative to outside of households, respectively.

Hazard of infection between different genders is higher for households than non-household transmission

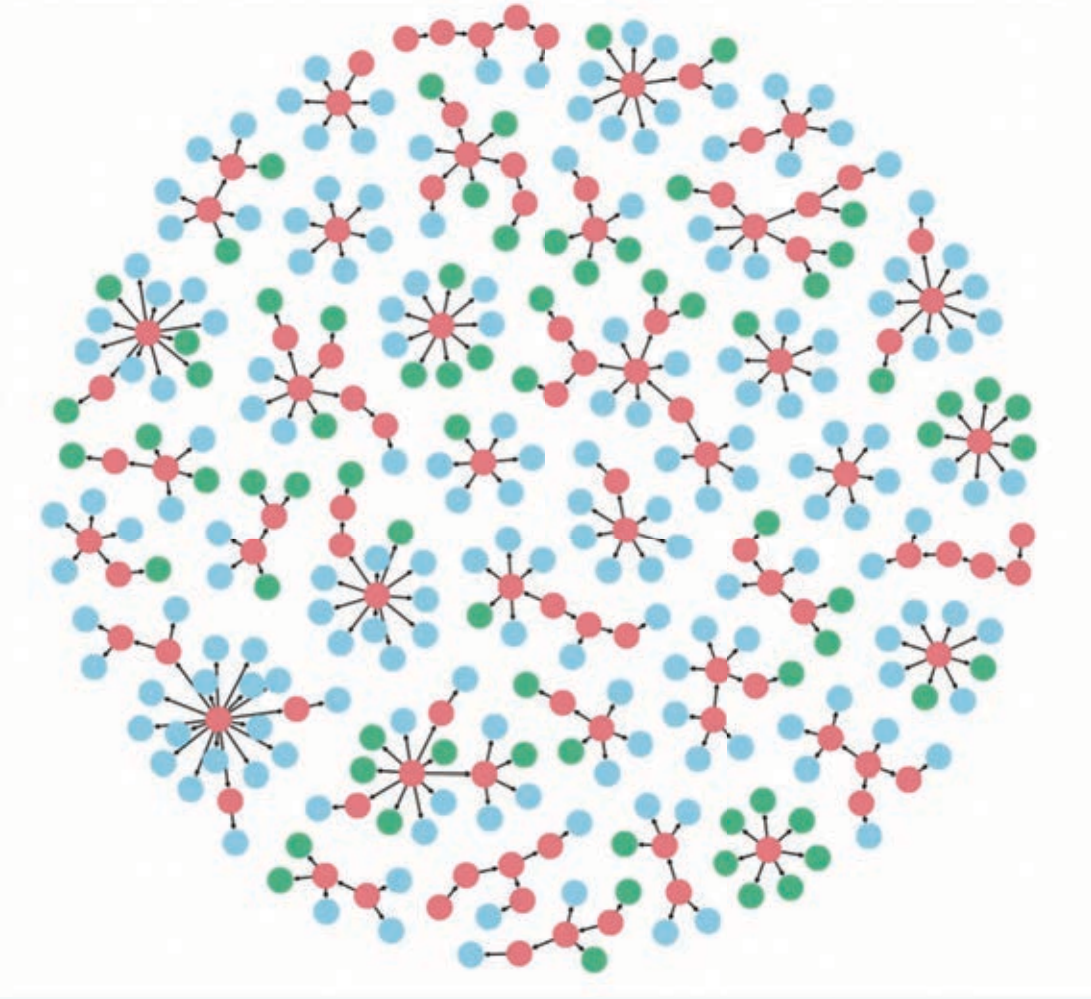
Xiao-Ke Xu#, Xiao-Fan Liu#, Ye Wu#, Sheikh Taslim Ali#, Zhanwei Du#, Paolo Bosetti, Eric H Y Lau, Benjamin J Cowling, Lin Wang*. Reconstruction of Transmission Pairs for novel Coronavirus Disease 2019 in mainland China: Estimation of Super-spreading Events, Serial Interval, and Hazard of Infection. Clinical Infectious Diseases, (2020) doi:10.1093/cid/ciaa790

Household and non-household transmission are coupled to form larger transmissions

643 transmission clusters

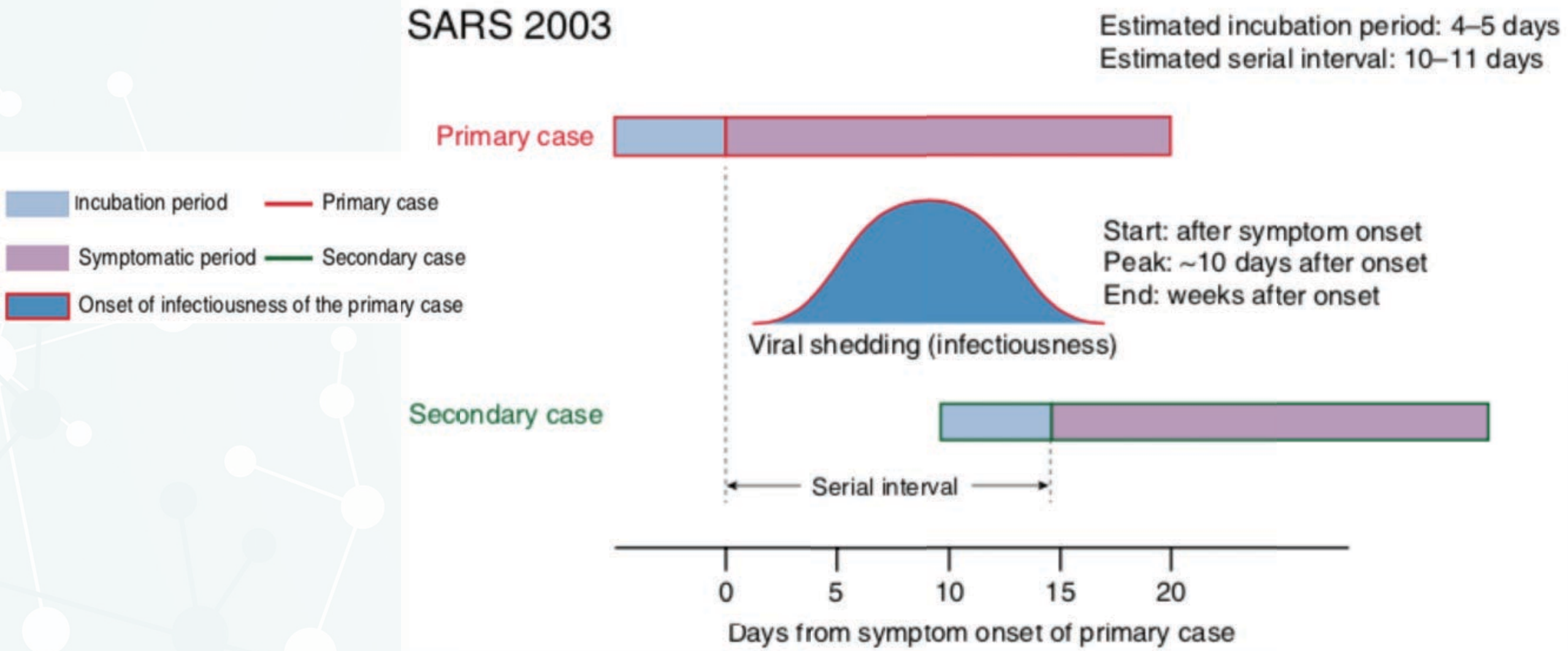


Superspreading: The size is larger than 6 nodes



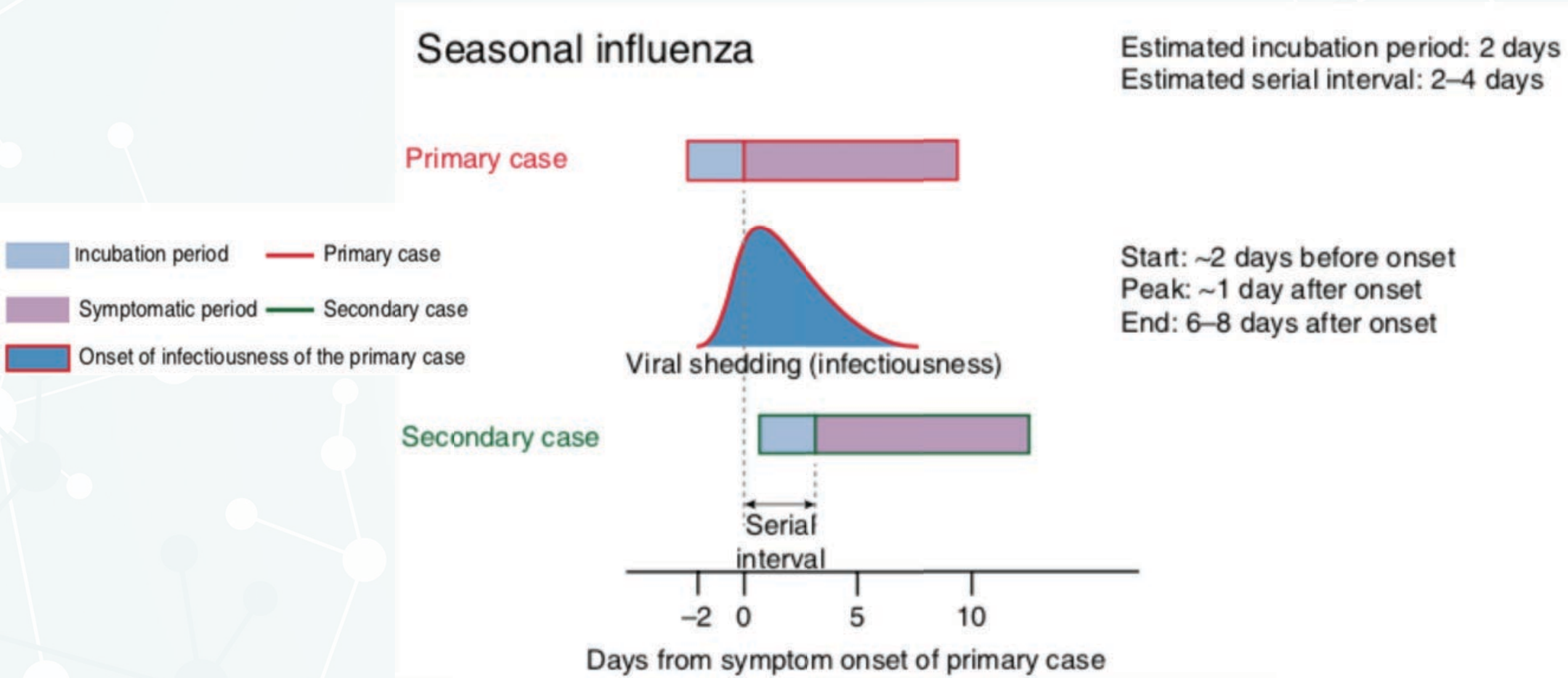
Red, green and blue nodes denote primary cases, household secondary cases and non-household secondary cases, respectively.

The relation between different time periods in disease transmission



He, X., Lau, E.H.Y., Wu, P. et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med 26, 672–675 (2020).

The relation between different time periods in disease transmission



He, X., Lau, E.H.Y., Wu, P. et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med 26, 672–675 (2020).

Serial Interval of COVID-19 among Publicly Reported Confirmed Cases

Pre-smptomatic transmission definitely makes containment more difficult

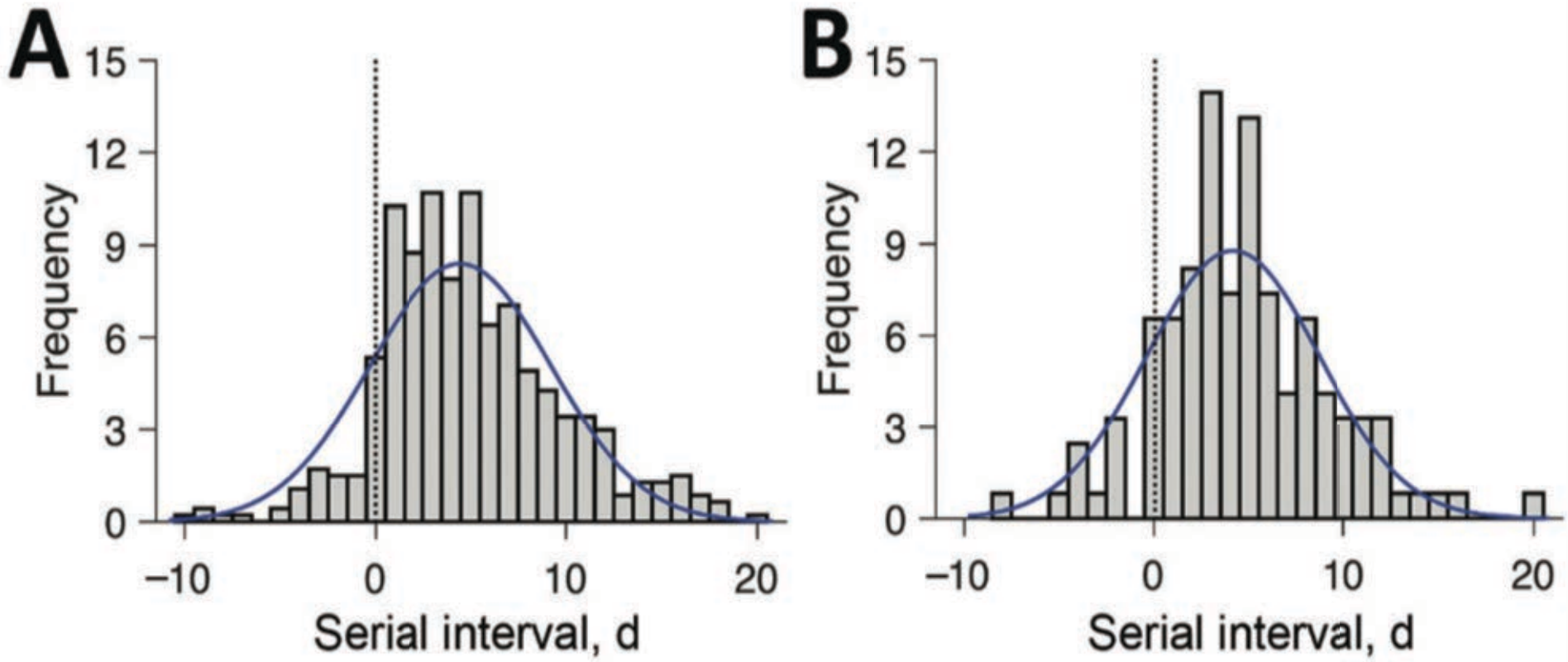
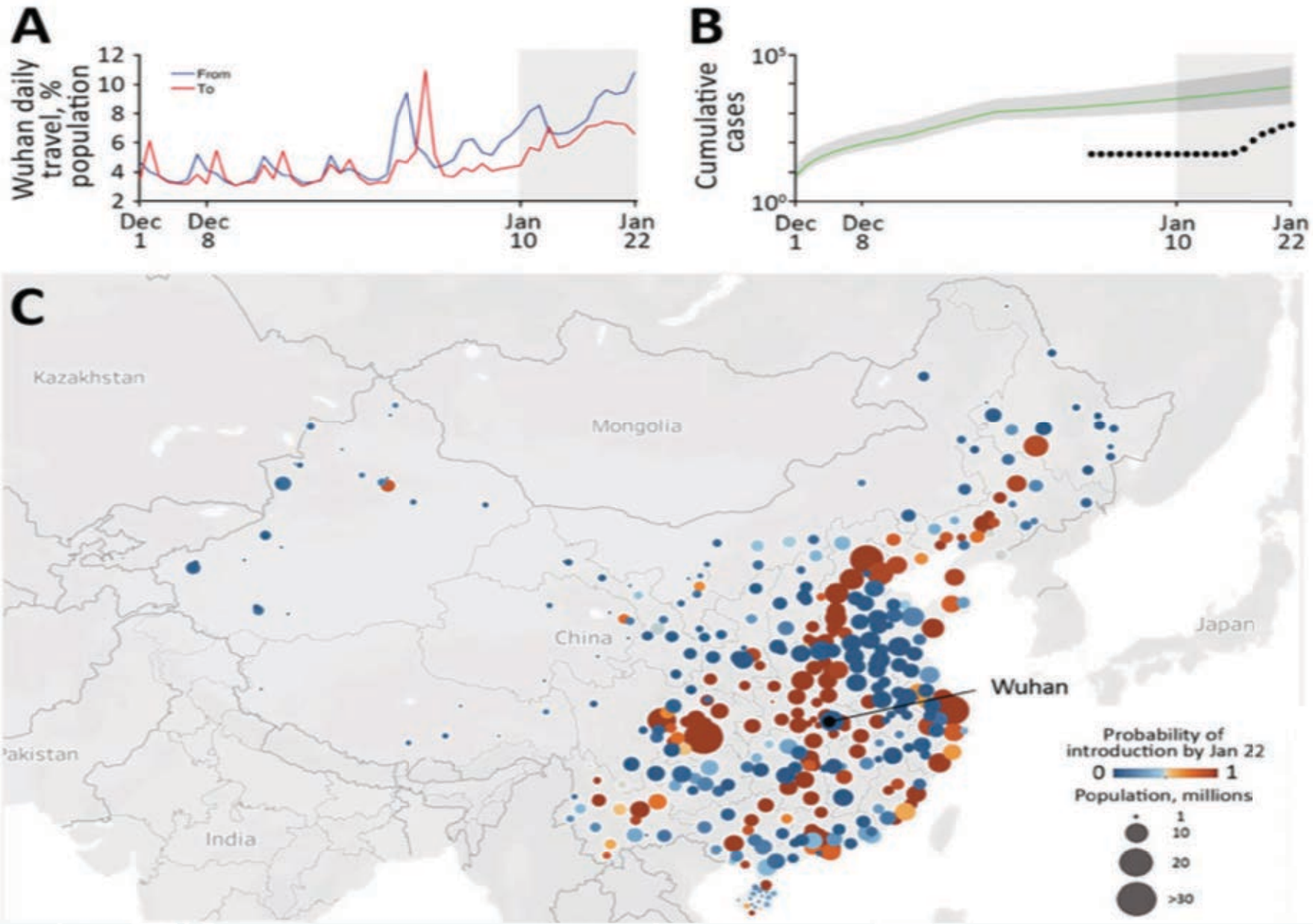


Figure. Estimated serial interval distribution for coronavirus disease (COVID-19) based on 468 reported transmission events, China, January 21–February 8, 2020. A) All infection events (N = 468) reported across 93 cities of mainland China as of February 8, 2020; B) the subset infection events (n = 122) in which both the infector and infectee were infected in the reporting city (i.e., the index patient's case was not an importation from another city). Gray bars indicate the number of infection events with specified serial interval, and blue lines indicate fitted normal distributions. Negative serial intervals (left of the vertical dotted lines) suggest the possibility of COVID-19 transmission from asymptomatic or mildly symptomatic case-patients.

Zhanwei Du#, **Xiao-Ke Xu#,** Ye Wu#, Lin Wang, Benjamin J. Cowling, and Lauren Ancel Meyers*, The serial interval of COVID-19 from publicly reported confirmed cases, Emerging Infectious Diseases. 2020, 26(6):1341-1343.

Risk for Transportation of COVID-19 from Wuhan to Other Cities



A) Daily travel volume to and from Wuhan, given as a percentage of the Wuhan population.

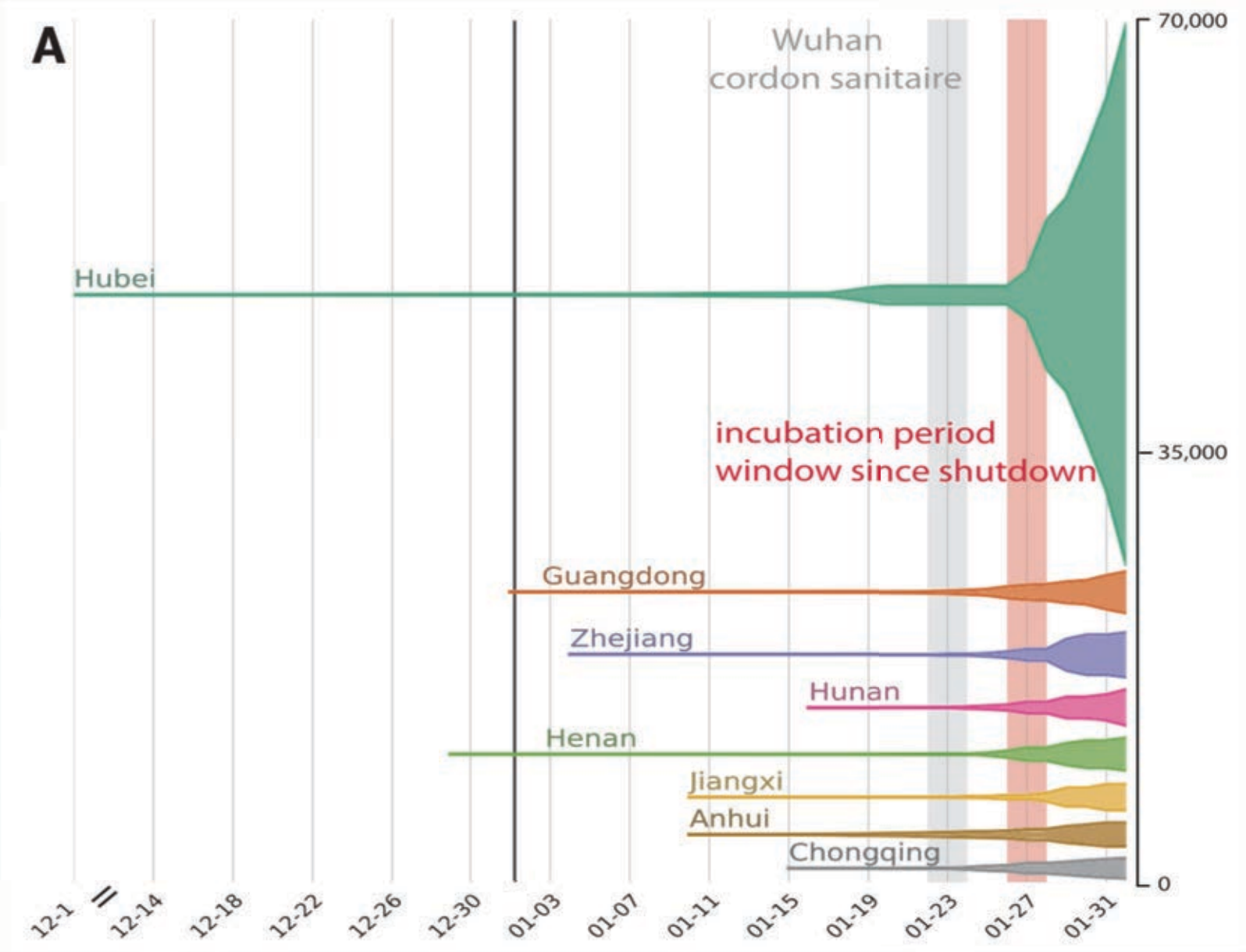
C) Probability that >1 COVID-19 case infected in Wuhan traveled to cities in China by January 22, 2020. The 131 cities with a risk threshold >50% are indicated in shades of orange; 239 cities below that threshold are indicated in shades of blue.

Submitted on January 27

Published on February 13

Zhanwei Du[#], Lin Wang[#], Simon Cauchemez, **Xiaoke Xu**, Xianwen Wang, Benjamin J. Cowling, and Lauren Ancel Meyers[*], Risk for transportation of 2019 novel coronavirus disease from Wuhan to other cities in China. Emerg Infect Dis. 2020 May.

Number of cases and key dates during the epidemic



M. U. G. Kraemer et al., The effect of human mobility and control measures on the COVID-19 epidemic in China. Science, eabb4218 (2020).

The cities deployed different types of interventions

(1) bans on entertainment and public gatherings;

(2) broad restrictions on public service including healthcare, schooling, shopping and restaurants;

(3) initiation of a level-1 response entailing systematic testing and isolation of confirmed cases;

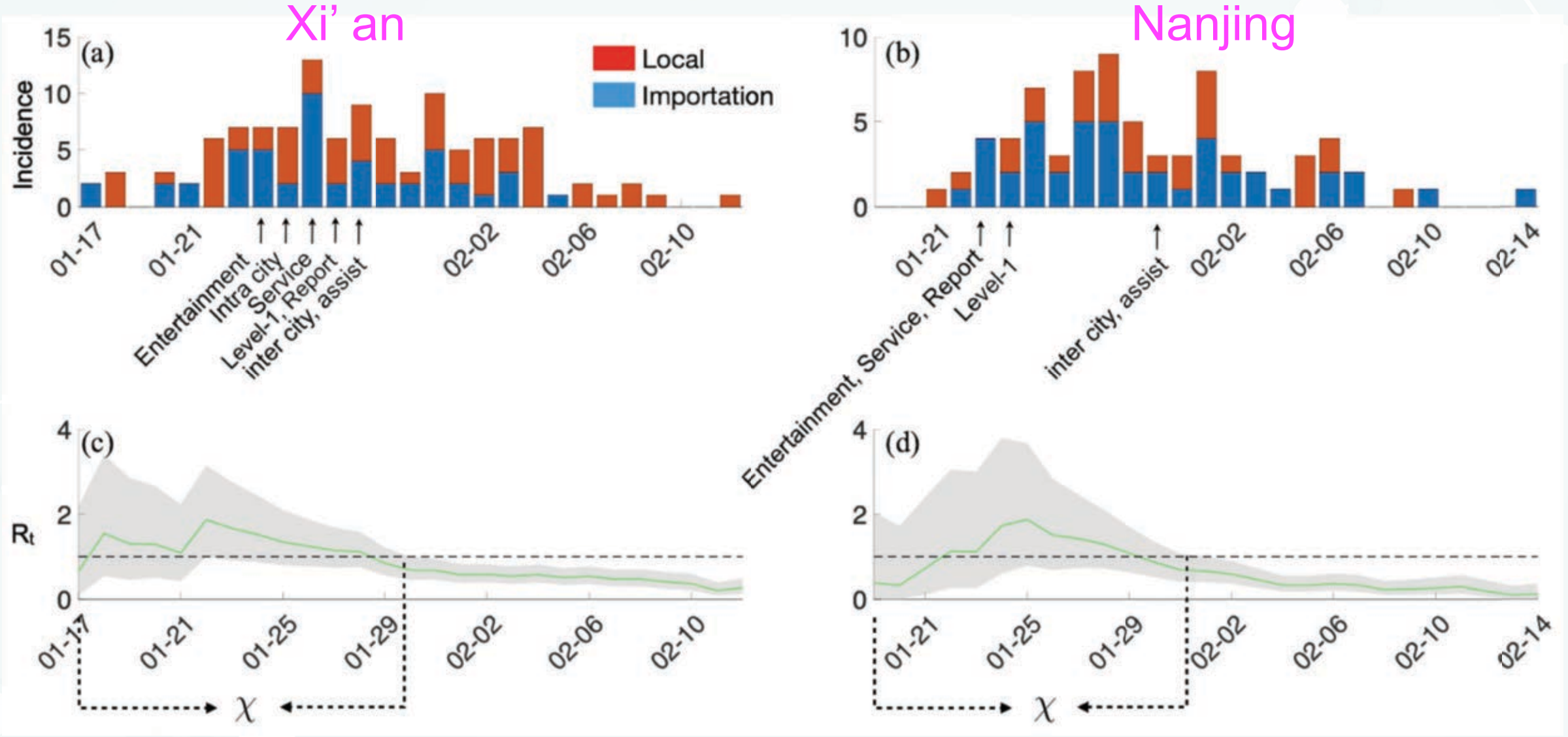
(4) suspension of intra-city public transport;

(5) suspension of travel between cities;

(6) reporting of confirmed cases;

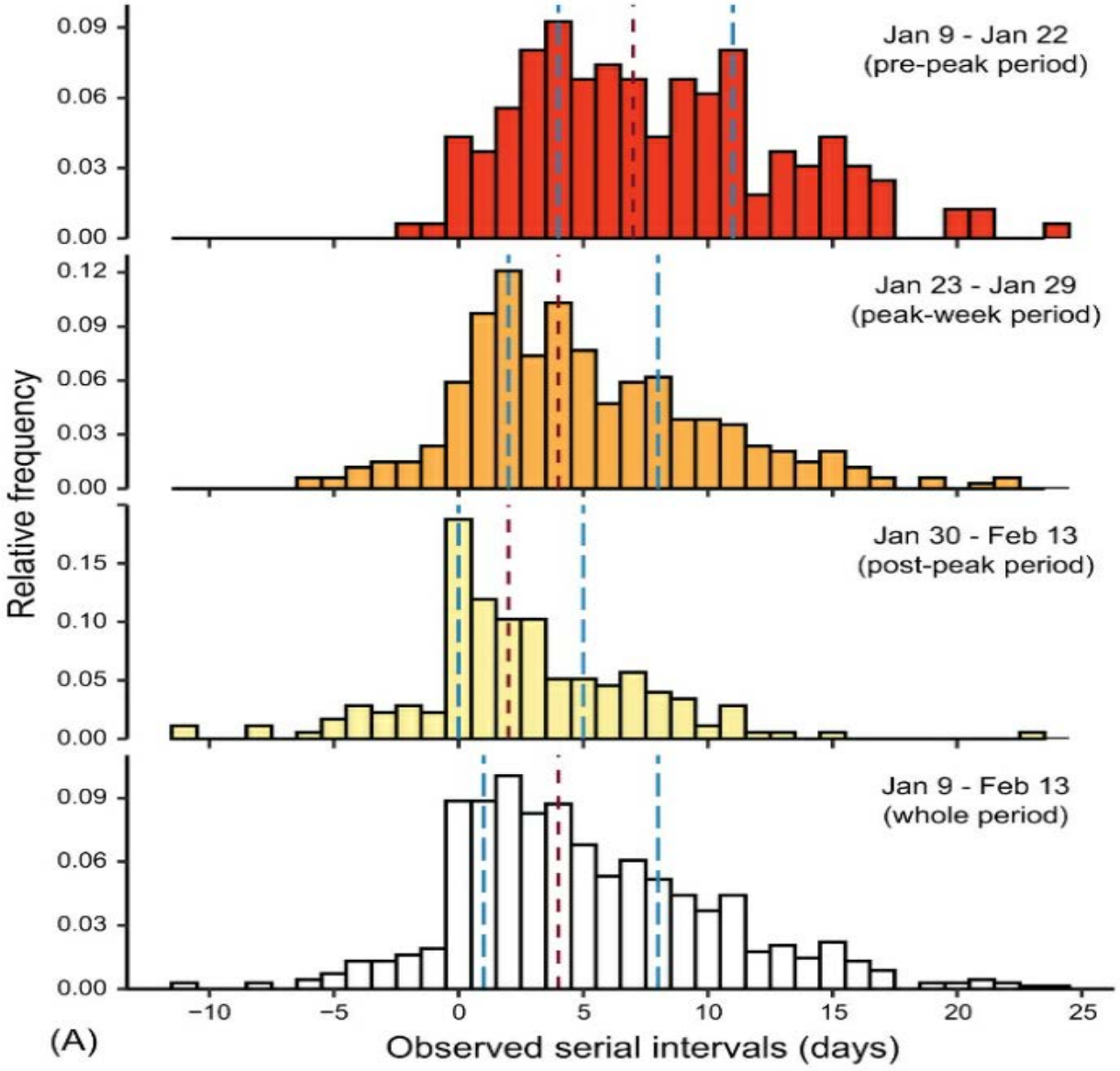
(7) recruitment of governmental staff and volunteers to enforce quarantine and social distancing.

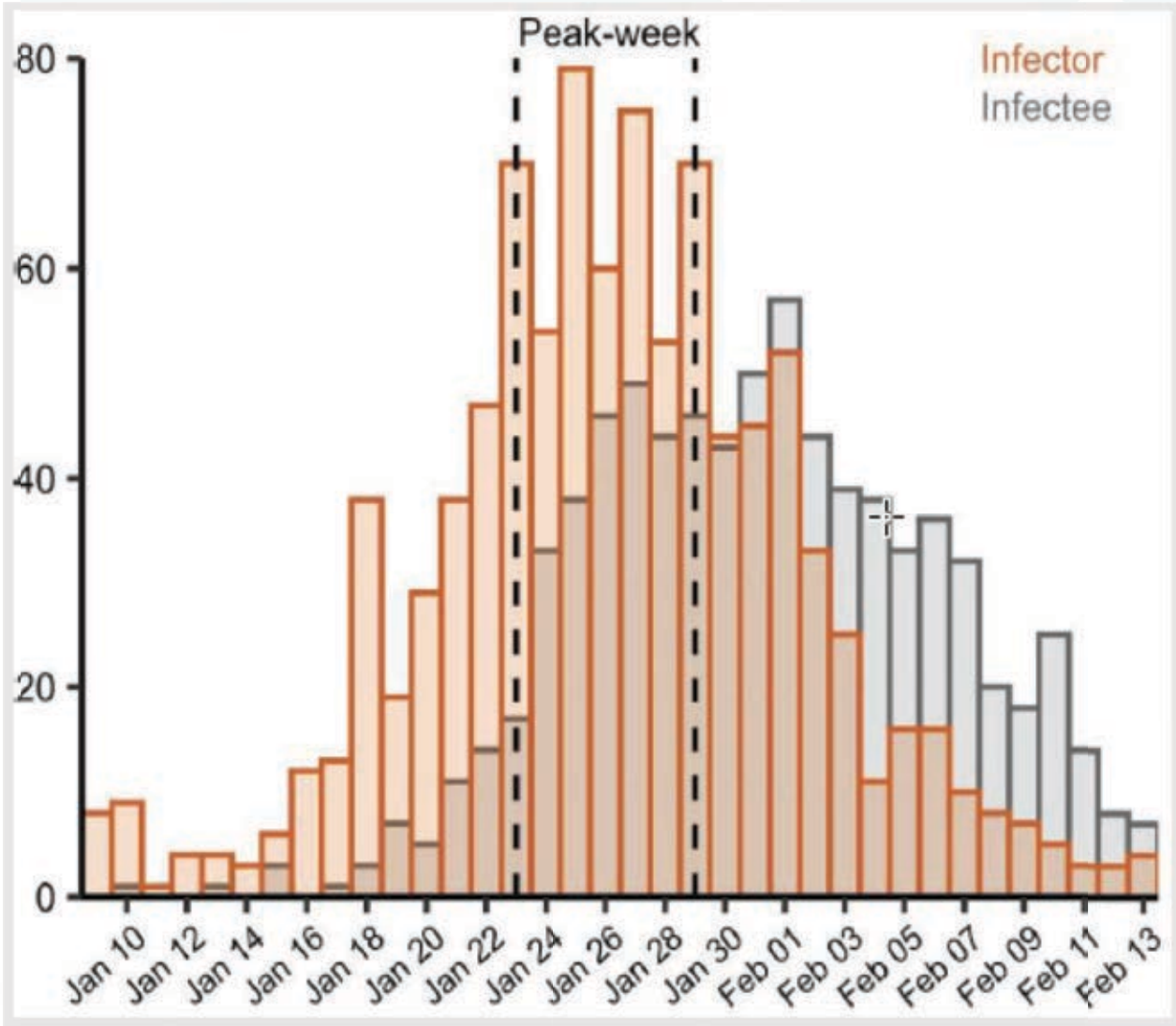
Proactive social distancing mitigates COVID-19 outbreaks across 58 cities



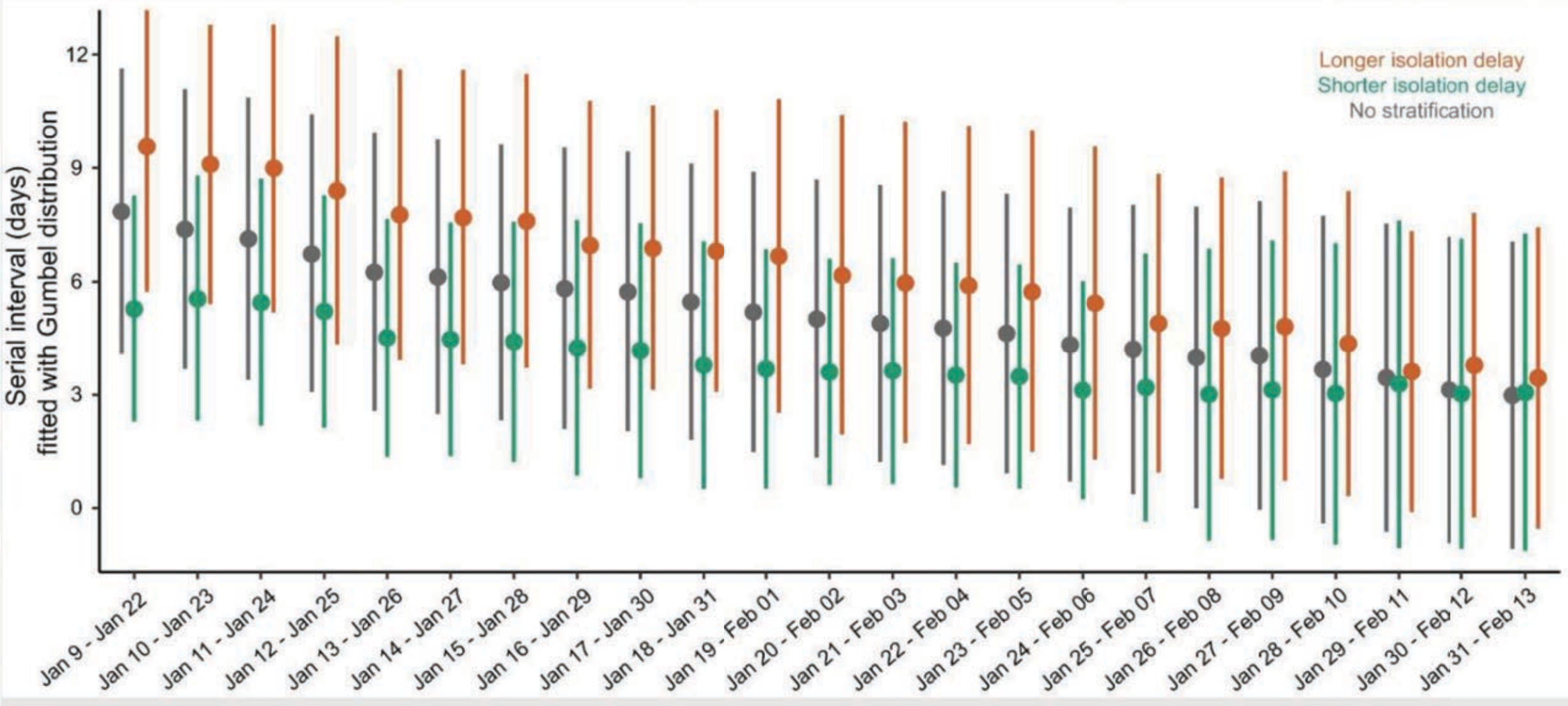
Zhanwei Du, **Xiao-Ke Xu**, Lin Wang, Spencer J. Fox, Benjamin J. Cowling, Alison P. Galvani, and Lauren Ancel Meyers*. Effects of proactive social distancing on COVID-19 outbreaks in 58 cities, China. Emerging Infectious Diseases. 2020, https://doi.org/10.3201/eid2609.201932

Temporal change of the time delay in isolating COVID-19 infectors from their symptom onset (i.e., isolation delay) in mainland China



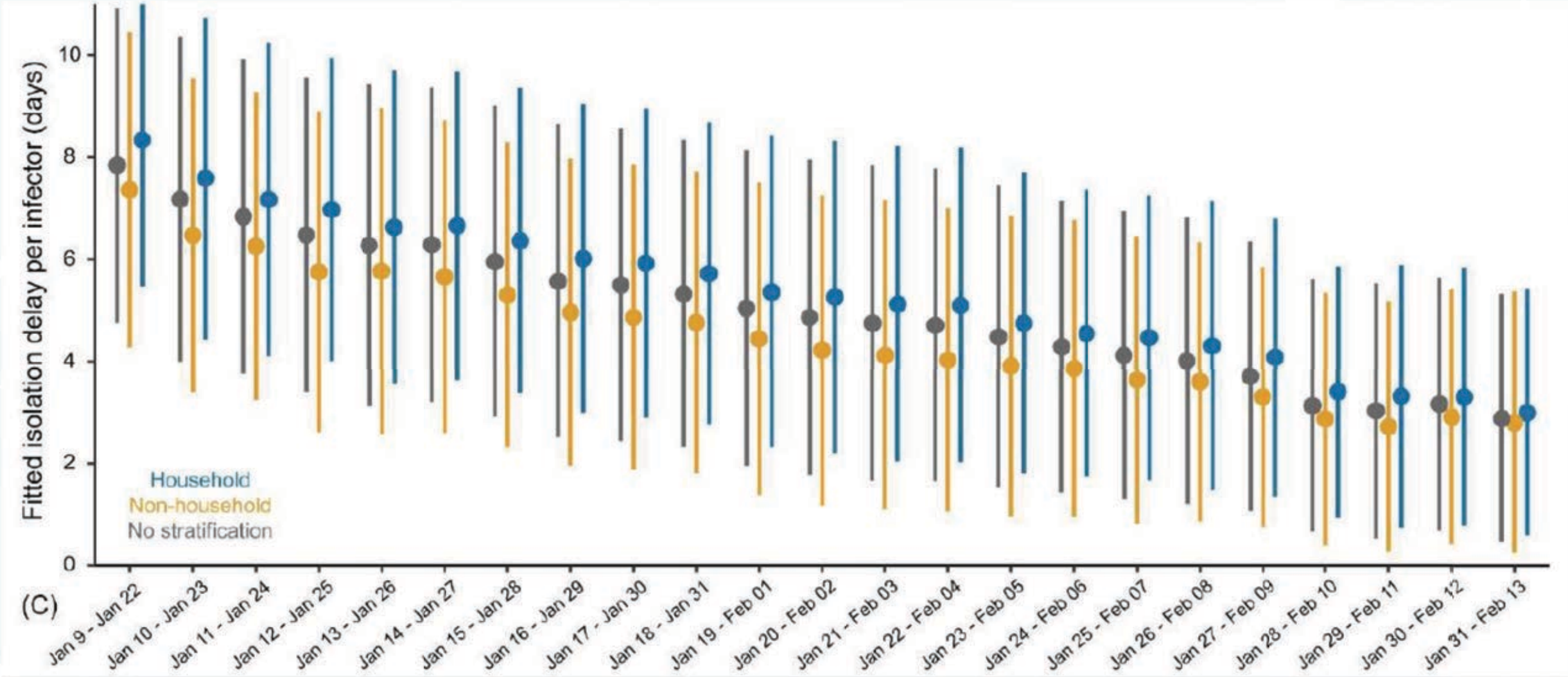


Evolution of effective serial interval of SARS-CoV-2 by non-pharmaceutical interventions



Sheikh Taslim Ali#, Lin Wang#, Eric HY Lau#, **Xiao-Ke Xu**, Zhanwei Du, Ye Wu, Gabriel M. Leung, Benjamin J. Cowling*, Evolution of effective serial interval of SARS-CoV-2 by non-pharmaceutical interventions, SCIENCE, 10.1126/science.abc9004 (2020)

Serial intervals of COVID-19 have shortened substantially from 7.8 days to 2.6 days within a month



Sheikh Taslim Ali#, Lin Wang#, Eric HY Lau#, **Xiao-Ke Xu**, Zhanwei Du, Ye Wu, Gabriel M. Leung, Benjamin J. Cowling*, Evolution of effective serial interval of SARS-CoV-2 by non-pharmaceutical interventions, SCIENCE, 10.1126/science.abc9004 (2020)

Possible tasks in future

- **Reconstruction of transmission events**
- Reconstruction of epidemiological timelines
- **Spatial-temporal spreading trajectories**
- Emergency monitoring with artificial intelligence

Now we provide this dataset to some interested researchers

Xiao Fan Liu, **Xiao-Ke Xu***, Ye Wu*, Mobility, exposure, and epidemiological timelines of COVID-19 infections in China outside Hubei Province, **Scientific Data**, submitted

Thanks to my collaborators!

xuxiaoke@foxmail.com