Cite as: R. Li *et al*., *Science* 10.1126/science.abb3221 (2020).

Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2)

Ruiyun Li[1]*, Sen Pei[2]*†, Bin Chen[3]*, Yimeng Song[4], Tao Zhang[5], Wan Yang[6], Jeffrey Shaman[2]†

[1]MRC Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, School of Public Health, Faculty of Medicine, Imperial College London, London W2 1PG, UK. [2]Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY 10032, USA. [3]Department of Land, Air and Water Resources, University of California, Davis, Davis, CA 95616, USA. [4]Department of Urban Planning and Design, The University of Hong Kong, Hong Kong. [5]Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 10084, P. R. China. [6]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY 10032, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: sp3449@cumc.columbia.edu (S.P.); jls106@cumc.columbia.edu (J.S.)

Estimation of the prevalence and contagiousness of undocumented novel coronavirus (SARS-CoV2) infections is critical for understanding the overall prevalence and pandemic potential of this disease. Here we use observations of reported infection within China, in conjunction with mobility data, a networked dynamic metapopulation model and Bayesian inference, to infer critical epidemiological characteristics associated with SARS-CoV2, including the fraction of undocumented infections and their contagiousness. We estimate 86% of all infections were undocumented (95% CI: [82%–90%]) prior to 23 January 2020 travel restrictions. Per person, the transmission rate of undocumented infections was 55% of documented infections ([46%–62%]), yet, due to their greater numbers, undocumented infections were the infection source for 79% of documented cases. These findings explain the rapid geographic spread of SARS-CoV2 and indicate containment of this virus will be particularly challenging.

Downloaded from http://science.sciencemag.org/ on March 24, 2020

The novel coronavirus that emerged in Wuhan, China (SARS-CoV2) at the end of 2019 quickly spread to all Chinese provinces and, as of 1 March 2020, to 58 other countries (*1*, *2*). Efforts to contain the virus are ongoing; however, given the many uncertainties regarding pathogen transmissibility and virulence, the effectiveness of these efforts is unknown.

The fraction of undocumented but infectious cases is a critical epidemiological characteristic that modulates the pandemic potential of an emergent respiratory virus (*3*–*6*). These undocumented infections often experience mild, limited or no symptoms and hence go unrecognized, and, depending on their contagiousness and numbers, can expose a far greater portion of the population to virus than would otherwise occur. Here, to assess the full epidemic potential of SARS-CoV2, we use a model-inference framework to estimate the contagiousness and proportion of undocumented infections in China during the weeks before and after the shutdown of travel in and out of Wuhan.

We developed a mathematical model that simulates the spatiotemporal dynamics of infections among 375 Chinese cities (see supplementary materials). In the model, we divided infections into two classes: (i) documented infected individuals with symptoms severe enough to be confirmed, i.e., observed infections; and (ii) undocumented infected individuals. These two classes of infection have separate rates of transmission: β, the transmission rate due to documented infected individuals; and $\mu\beta$, the transmission rate due to undocumented individuals, which is β reduced by a factor μ.

Spatial spread of SARS-CoV2 across cities is captured by the daily number of people traveling from city j to city i and a multiplicative factor. Specifically, daily numbers of travelers between 375 Chinese cities during the Spring Festival period ("Chunyun") were derived from human mobility data collected by the Tencent Location-based Service during the 2018 Chunyun period (1 February–12 March 2018) (*7*). Chunyun is a period of 40 days—15 days before and 25 days after the Lunar New Year—during which there are high rates of travel within China. To estimate human mobility during the 2020 Chunyun period, which began 10 January, we aligned the 2018 Tencent data based on relative timing to the Spring Festival. For example, we used mobility data from 1 February 2018 to represent human movement on 10 January 2020, as these days were similarly distant from the Lunar New Year. During the 2018 Chunyun, a total of 1.73 billion travel events were captured in the Tencent data; whereas 2.97 billion trips are reported (*7*). To compensate for underreporting and reconcile these two numbers, a travel multiplicative factor, θ, which is greater than 1, is included (see supplementary materials).

To infer SARS-CoV2 transmission dynamics during the early stage of the outbreak, we simulated observations during 10–23 January 2020 (i.e., the period before the initiation of

travel restrictions, fig. S1) using an iterated filter-ensemble adjustment Kalman filter (IF-EAKF) framework (*8–10*). With this combined model-inference system, we estimated the trajectories of four model state variables (S_i, E_i, I_i^r, I_i^u: the susceptible, exposed, documented infected, and undocumented infected sub-populations in city i) for each of the 375 cities, while simultaneously inferring six model parameters (Z, D, μ, β, α, θ: the average latent period, the average duration of infection, the transmission reduction factor for undocumented infections, the transmission rate for documented infections; the fraction of documented infections, and the travel multiplicative factor).

Details of model initialization, including the initial seeding of exposed and undocumented infections, are provided in the supplementary materials. To account for delays in infection confirmation, we also defined a time-to-event observation model using a Gamma distribution (see supplementary materials). Specifically, for each new case in group I_i^r, a reporting delay t_d (in days) was generated from a Gamma distribution with a mean value of T_d. In fitting both synthetic and the observed outbreaks, we performed simulations with the model-inference system using different fixed values of T_d (6 days ≤ T_d ≤ 10 days) and different maximum seeding, $Seed_{max}$ (1500 ≤ $Seed_{max}$ ≤ 2500) (see supplementary materials, fig. S2). The best fitting model-inference posterior was identified by log-likelihood.

We first tested the model-inference framework versus alternate model forms and using synthetic outbreaks generated by the model in free simulation. These tests verified the ability of the model-inference framework to accurately estimate all six target model parameters simultaneously (see supplementary methods and figs. S3 to S14). Indeed, the system could identify a variety of parameter combinations and distinguish outbreaks generated with high α and low μ from low α and high μ. This parameter identifiability is facilitated by the assimilation of observed case data from multiple (375) cities into the model-inference system and the incorporation of human movement in the mathematical model structure (see supplementary methods and figs. S15 and S16).

We next applied the model-inference framework to the observed outbreak before the travel restrictions of 23 January—a total of 801 documented cases throughout China, as reported by 8 February 2020 (*1*). Figure 1, A to C, shows simulations of reported cases generated using the best-fitting model parameter estimates. The distribution of these stochastic simulations captures the range of observed cases well. In addition, the best-fitting model captures the spread of infections with the novel coronavirus (COVID-19) to other cities in China (fig. S17). Our median estimate of the effective reproductive number, R_e—equivalent to the basic reproductive number (R_0) at the beginning of the epidemic—is 2.38 (95% CI: 2.04–2.77), indicating a high capacity for sustained transmission of COVID-19 (Table 1 and Fig. 1D). This finding aligns with other recent estimates of the reproductive number for this time period (*6*, *11–15*). In addition, the median estimates for the latent and infectious periods are approximately 3.69 and 3.48 days, respectively. We also find that, during 10–23 January, only 14% (95% CI: 10–18%) of total infections in China were reported. This estimate reveals a very high rate of undocumented infections: 86%. This finding is independently corroborated by the infection rate among foreign nationals evacuated from Wuhan (see supplementary materials). These undocumented infections are estimated to have been half as contagious per individual as reported infections (μ = 0.55; 95% CI: 0.46–0.62). Other model fittings made using alternate values of T_d and $Seed_{max}$ or different distributional assumptions produced similar parameter estimates (figs. S18 to S22), as did estimations made using an alternate model structure with separate average infection periods for undocumented and documented infections (see supplementary methods, table S1). Further sensitivity testing indicated that α and μ are uniquely identifiable given the model structure and abundance of observations utilized (see supplementary methods and Fig. 1, E and F). In particular, Fig. 1F shows that the highest log-likelihood fittings are centered in the 95% CI estimates for α and μ and drop off with distance from the best fitting solution (α= 0.14 and μ = 0.55).

Using the best-fitting model (Table 1 and Fig. 1), we estimated 13,118 (95% CI: 2,974–23,435) total new COVID-19 infections (documented and undocumented combined) during 10–23 January in Wuhan city. Further, 86.2% (95% CI: 81.5%–89.8%) of all infections were infected from undocumented cases. Nationwide, the total number of infections during 10–23 January was 16,829 (95% CI: 3,797–30,271) with 86.2% (95% CI: 81.6%–89.8%) infected by undocumented cases.

To further examine the impact of contagious, undocumented COVID-19 infections on overall transmission and reported case counts, we generated a set of hypothetical outbreaks using the best-fitting parameter estimates but with μ = 0, i.e., the undocumented infections are no longer contagious (Fig. 2). We find that without transmission from undocumented cases, reported infections during 10–23 January are reduced 78.8% across all of China and 66.1% in Wuhan. Further, there are fewer cities with more than 10 cumulative documented cases: only 1 city with more than 10 documented cases versus the 10 observed by 23 January (Fig. 2). This finding indicates that contagious, undocumented infections facilitated the geographic spread of SARS-CoV2 within China.

We also modeled the transmission of COVID-19 in China after 23 January, when greater control measures were effected. These control measures included travel restrictions imposed between major cities and Wuhan; self-quarantine and contact precautions advocated by the government; and

Downloaded from http://science.sciencemag.org/ on March 24, 2020

more available rapid testing for infection confirmation (*11*, *12*). These measures along with changes in medical care-seeking behavior due to increased awareness of the virus and increased personal protective behavior (e.g., wearing of facemasks, social distancing, self-isolation when sick), likely altered the epidemiological characteristics of the outbreak after 23 January. To quantify these differences, we re-estimated the system parameters using the model-inference framework and city-level daily cases reported between 24 January and 8 February. As inter-city mobility was restricted after 23 January, we tested two altered travel scenarios: (i) scenario 1: a 98% reduction of travel in and out of Wuhan and an 80% reduction of travel between all other cities, as indicated by changes in the Baidu Mobility Index (*16*) (table S2); and (ii) scenario 2: a complete stoppage of inter-city travel (i.e., θ to 0) (see supplementary methods for more details).

The results of inference for the 24 January–8 February period are presented in Table 2, figs. S23 to S26, and table S3. As control measures have continually shifted, we present estimates for both 24 January–3 February (Period 1) and 24 January–8 February (Period 2). For both periods, the best-fitting model for Scenario 1 had a reduced reporting delay, T_d, of 6 days (vs. 10 days before 23 January), consistent with more rapid confirmation of infections. Estimates of both the latency and infectious periods were similar to those made for 10–23 January; however, α, β, and R_e all shifted considerably. The transmission rate of documented cases, β, dropped to 0.52 (95% CI: 0.39–0.71) during Period 1 and 0.35 (95% CI: 0.27–0.50) during Period 2, less than half the estimate prior to travel restrictions (Table 2). The fraction of all infections that were documented, α, was estimated to be 0.65 (95% CI: 0.60–0.69), i.e., 65% of infections were documented during Period 1, up from 14% prior to travel restrictions, and remained nearly the same for Period 2. The reproductive number was 1.36 (95% CI: 1.14–1.63) during Period 1 and 0.99 (95% CI: 0.76–1.33) during Period 2, down from 2.38 prior to travel restrictions. While the estimate for the relative transmission rate, μ, is lower than before 23 January, the contagiousness of undocumented infections, represented by $\mu\beta$, was substantially reduced, possibly reflecting that only very mild, less contagious infections remain undocumented or that individual protective behavior and contact precautions have proven effective. Similar parameter estimates are derived under Scenario 2 (no travel at all) (table S3). These inference results for both Period 1 and 2 should be interpreted with caution as care-seeking behavior and control measures were continually in flux at these times.

Overall, our findings indicate that a large proportion of COVID-19 infections were undocumented prior to the implementation of travel restrictions and other heightened control measures in China on 23 January, and that a large proportion of the total force of infection was mediated through these undocumented infections (Table 1). This high proportion of undocumented infections, many of whom were likely not severely symptomatic, appears to have facilitated the rapid spread of the virus throughout China. Indeed, suppression of the infectiousness of these undocumented cases in model simulations reduces the total number of documented cases and the overall spread of SARS-CoV2 (Fig. 2). In addition, the best-fitting model has a reporting delay of 9 days from initial infectiousness to confirmation; in contrast line-list data for the same 10–23 January period indicates an average 6.6 day delay from initial manifestation of symptoms to confirmation (*17*). This discrepancy suggests pre-symptomatic shedding may be typical among documented infections. The relative timing of viremia and shedding onset and peak versus symptom onset and peak has been shown to potentially affect outbreak control success (*18*).

Our findings also indicate that a radical increase in the identification and isolation of currently undocumented infections would be needed to fully control SARS-CoV2. Increased news coverage and awareness of the virus in the general population have already likely prompted increased rates of seeking medical care for respiratory symptoms. In addition, awareness among healthcare providers, public health officials and the availability of viral identification assays suggest that capacity for identifying previously missed infections has increased. Further, general population and government response efforts have increased the use of face masks, restricted travel, delayed school reopening and isolated suspected persons, all of which could additionally slow the spread of SARS-CoV2.

Combined, these measures are expected to increase reporting rates, reduce the proportion of undocumented infections, and decrease the growth and spread of infection. Indeed, estimation of the epidemiological characteristics of the outbreak after 23 January in China, indicate that government control efforts and population awareness have reduced the rate of spread of the virus (i.e., lower β, $\mu\beta$, R_e), increased the reporting rate, and lessened the burden on already overextended healthcare systems.

Importantly, the situation on the ground in China is changing day-to-day. New travel restrictions and control measures are being imposed on new populations in different cities, and these rapidly varying effects make certain estimation of the epidemiological characteristics for the outbreak difficult. Further, reporting inaccuracies and changing care-seeking behavior add another level of uncertainty to our estimations. While the data and findings presented here indicate that travel restrictions and control measures have reduced SARS-CoV2 transmission considerably, whether these controls are sufficient for reducing R_e below 1 for the length of time needed to eliminate the disease locally and prevent a

Downloaded from http://science.sciencemag.org/ on March 24, 2020

rebound outbreak once control measures are relaxed is unclear. Further, similar control measures and travel restrictions would have to be implemented outside China to prevent reintroduction of the virus.

The results for 10–23 January 2020 delineate the characteristics of the SARS-CoV2 moving through a developed society, China, without major restrictions or control. These findings provide a baseline assessment of the fraction of undocumented infections and their relative infectiousness for such an environment. However, differences in control activity, viral surveillance and testing, and case definition and reporting would likely impact rates of infection documentation. Thus, the key findings, that 86% of infections went undocumented and that, per person, these undocumented infections were 55% as contagious as documented infections, could shift in other countries with different control, surveillance and reporting practices.

Our findings underscore the seriousness and pandemic potential of SARS-CoV2. The 2009 H1N1 pandemic influenza virus also caused many mild cases, quickly spread globally, and eventually became endemic. Presently, there are four, endemic, coronavirus strains currently circulating in human populations (229E, HKU1, NL63, OC43). If the novel coronavirus follows the pattern of 2009 H1N1 pandemic influenza, it will also spread globally and become a fifth endemic coronavirus within the human population.

REFERENCES AND NOTES

1. National Health Commission of the People's Republic of China, Update on the novel coronavirus pneumonia outbreak; http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml [accessed 8 February 2020].
2. World Health Organization, Coronavirus disease (COVID-2019) situation reports; https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ [accessed 1 March 2020].
3. J. F. Chan, S. Yuan, K. H. Kok, K. K. To, H. Chu, J. Yang, F. Xing, J. Liu, C. C. Yip, R. W. Poon, H. W. Tsoi, S. K. Lo, K. H. Chan, V. K. Poon, W. M. Chan, J. D. Ip, J. P. Cai, V. C. Cheng, H. Chen, C. K. Hui, K. Y. Yuen, A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *Lancet* **395**, 514–523 (2020). doi:10.1016/S0140-6736(20)30154-9 Medline
4. P. Wu, X. Hao, E. H. Y. Lau, J. Y. Wong, K. S. M. Leung, J. T. Wu, B. J. Cowling, G. M. Leung, Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020. *Euro Surveill.* **25**, 2000044 (2020). doi:10.2807/1560-7917.ES.2020.25.3.2000044 Medline
5. V. J. Munster, M. Koopmans, N. van Doremalen, D. van Riel, E. de Wit, A novel coronavirus emerging in China - Key questions for impact assessment. *N. Engl. J. Med.* **382**, 692–694 (2020). doi:10.1056/NEJMp2000929 Medline
6. Z. Du, L. Wang, S. Cauchemez, X. Xu, X. Wang, B. J. Cowling, L. A. Meyers, Risk for transportation of 2019 novel coronavirus disease from Wuhan to other cities in China. *Emerg. Infect. Dis.* **26**, (2020). doi:10.3201/eid2605.200146 Medline
7. http://society.people.com.cn/n1/2018/0315/c1008-29869526.html [accessed 8 February 2020].
8. E. L. Ionides, C. Bretó, A. A. King, Inference for nonlinear dynamical systems. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18438–18443 (2006). doi:10.1073/pnas.0603181103 Medline
9. A. A. King, E. L. Ionides, M. Pascual, M. J. Bouma, Inapparent infections and cholera dynamics. *Nature* **454**, 877–880 (2008). doi:10.1038/nature07084 Medline
10. S. Pei, F. Morone, F. Liljeros, H. Makse, J. L. Shaman, Inference and control of the nosocomial transmission of methicillin-resistant *Staphylococcus aureus*. *eLife* **7**, e40977 (2018). doi:10.7554/eLife.40977 Medline
11. Health Commission of Hubei Province, The 8th Press Conference on the Prevention and Control of COVID-19; http://wjw.hubei.gov.cn/fbjd/dtyw/202001/t20200130_2016544.shtml.
12. Health Commission of Hubei Province, The 9th Press Conference on the Prevention and Control of COVID-19; http://wjw.hubei.gov.cn/fbjd/dtyw/202001/t20200131_2017018.shtml.
13. J. T. Wu, K. Leung, G. M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *Lancet* **395**, 689–697 (2020). doi:10.1016/S0140-6736(20)30260-9 Medline
14. J. Riou, C. L. Althaus, Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Euro Surveill.* **25**, 2000058 (2020). doi:10.2807/1560-7917.ES.2020.25.4.2000058 Medline
15. N. Imai, I. Dorigatti, A. Cori, C. Donnelly, S. Riley, N. M. Ferguson, Report 2: Estimating the potential total number of novel Coronavirus cases in Wuhan City, China (2020); https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/news--wuhan-coronavirus/.
16. Baidu Migration; https://qianxi.baidu.com/ [accessed 26 February 2020].
17. M. Kramer, D. Pigott, B. Xu, S. Hill, B. Gutierrez, O. Pybus, Epidemiological data from the nCoV-2019 Outbreak: Early Descriptions from Publicly Available Data; http://virological.org/t/epidemiological-data-from-the-ncov-2019-outbreak-early-descriptions-from-publicly-available-data/337) [accessed 24 February 2020].
18. C. Fraser, S. Riley, R. M. Anderson, N. M. Ferguson, Factors that make an infectious disease outbreak controllable. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6146–6151 (2004). doi:10.1073/pnas.0307506101 Medline
19. S. Pei, SenPei-CU/COVID-19: COVID-19, Version 1, Zenodo (2020); http://doi.org/10.5281/zenodo.3699624.
20. S. Pei, S. Kandula, W. Yang, J. Shaman, Forecasting the spatial transmission of influenza in the United States. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 2752–2757 (2018). doi:10.1073/pnas.1708856115 Medline
21. A. Rambaut, Phylodynamic Analysis | 129 genomes | 24 Feb 2020; http://virological.org/t/phylodynamic-analysis-129-genomes-24-feb-2020/356/ [accessed 25 February 2020].
22. China National Health Commission, Policy and regulatory documents; http://www.nhc.gov.cn/wjw/gfxwjj/list.shtml [accessed 14 February 2020].
23. SenPei-CU/COVID-19, Data and code posting; https://github.com/SenPei-CU/COVID-19.
24. Field Briefing, Diamond Princess COVID-19 Cases, 20 Feb Update; https://www.niid.go.jp/niid/en/2019-ncov-e.html [accessed 25 February 2020].
25. Tencent Big Data Platform; https://heat.qq.com [accessed 14 February 2020].
26. D. Zhu, Z. Huang, L. Shi, L. Wu, Y. Liu, Inferring spatial interaction patterns from sequential snapshots of spatial distributions. *Int. J. Geogr. Inf. Sci.* **32**, 783–805 (2018). doi:10.1080/13658816.2017.1413192
27. D. He, E. L. Ionides, A. A. King, Plug-and-play inference for disease dynamics: Measles in large and small populations as a case study. *J. R. Soc. Interface* **7**, 271–283 (2010). doi:10.1098/rsif.2009.0151 Medline
28. M. S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.* **50**, 174–188 (2002). doi:10.1109/78.978374
29. J. L. Anderson, An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.* **129**, 2884–2903 (2001). doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2
30. C. Snyder, T. Bengtsson, P. Bickel, J. Anderson, Obstacles to high-dimensional particle filtering. *Mon. Weather Rev.* **136**, 4629–4640 (2008). doi:10.1175/2008MWR2529.1
31. W. Yang, A. Karspeck, J. Shaman, Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics. *PLOS Comput. Biol.* **10**, e1003583 (2014). doi:10.1371/journal.pcbi.1003583 Medline

Downloaded from http://science.sciencemag.org/ on March 24, 2020

32. J. Shaman, W. Yang, S. Kandula, Inference and forecast of the current West African ebola outbreak in Guinea, sierra leone and liberia. *PLOS Curr.* **6**, 10.1371/currents.outbreaks.3408774290b1a0f2dd7cae877c8b8ff6 (2014). Medline
33. N. B. DeFelice, E. Little, S. R. Campbell, J. Shaman, Ensemble forecast of human West Nile virus cases and mosquito infection rates. *Nat. Commun.* **8**, 14592 (2017). doi:10.1038/ncomms14592 Medline
34. J. Reis, J. Shaman, Retrospective parameter estimation and forecast of respiratory syncytial virus in the United States. *PLOS Comput. Biol.* **12**, e1005133 (2016). doi:10.1371/journal.pcbi.1005133 Medline
35. O. Diekmann, J. A. P. Heesterbeek, J. A. Metz, On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382 (1990). doi:10.1007/BF00178324 Medline
36. O. Diekmann, J. A. P. Heesterbeek, M. G. Roberts, The construction of next-generation matrices for compartmental epidemic models. *J. R. Soc. Interface* **7**, 873–885 (2010). doi:10.1098/rsif.2009.0386 Medline
37. S. Lai, I. Bogoch, N. Ruktanonchai, A. Watts, Y. Li, J. Yu, X. Lv, W. Yang, H. Yu, K. Khan, Z. Li, Assessing spread risk of Wuhan novel coronavirus within and beyond China, January-April 2020: a travel network-based modelling study. medRxiv 2020.02.04.20020479 [Preprint]. 5 February 2020. https://doi.org/10.1101/2020.02.04.20020479.

ACKNOWLEDGMENTS

Funding: This work was supported by US NIH grants GM110748 and AI145883. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences, the National Institute for Allergy and Infectious Diseases, or the National Institutes of Health. **Author contributions:** R.L., S.P., B.C., W.Y. and J.S. conceived the study. R.L., B.C., Y.S. and T.Z. curated data. S.P. performed the analysis. R.L., S.P., W.Y. and J.S. wrote the first draft of the manuscript. B.C, Y.S. and T.Z. reviewed and edited the manuscript. **Competing interests:** J.S. and Columbia University disclose partial ownership of SK Analytics. J.S. also reports receiving consulting fees from Merck and BNI. All other authors declare no competing interests. **Data and materials availability:** All code and data are available in the supplementary materials and posted online at https://github.com/SenPei-CU/COVID-19 and (*19*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/cgi/content/full/science.abb3221/DC1
Materials and Methods
Figs. S1 to S26
Tables S1 to S3
References (*20*–*37*)
MDAR Reproducibility Checklist
Data S1

15 February 2020; accepted 12 March 2020
Published online 16 March 2020
10.1126/science.abb3221

Downloaded from http://science.sciencemag.org/ on March 24, 2020

Table 1. Best-fit model posterior estimates of key epidemiological parameters for simulation with the full metapopulation model during 10–23 January 2020 ($Seed_{max}$ = 2000, T_d = 9 days).

Parameter	Median (95% CIs)
Transmission rate (β, days^{-1})	1.12 (1.04, 1.18)
Relative transmission rate (μ)	0.55 (0.46, 0.62)
Latency period (Z, days)	3.69 (3.28, 4.03)
Infectious period (D, days)	3.48 (3.18, 3.74)
Reporting rate (α)	0.14 (0.10, 0.18)
Basic reproductive number (R_e)	2.38 (2.04, 2.77)
Mobility factor (θ)	1.36 (1.28, 1.43)

Table 2. Best-fit model posterior estimates of key epidemiological parameters for simulation of the model during 24 January–3 February and 24 January–8 February ($Seed_{max}$ = 2000 on 10 January, T_d = 9 days before 24 January, T_d = 6 days between 24 January and 8 February). Travel to and from Wuhan is reduced by 98%, and other inter-city travel is reduced by 80%.

Parameter	24 January–3 February [Median (95% CIs)]	24 January–8 February [Median (95% CIs)]
Transmission rate (β, days^{-1})	0.52 (0.39, 0.71)	0.35 (0.27, 0.50)
Relative transmission rate (μ)	0.49 (0.37, 0.69)	0.44 (0.29, 0.66)
Latency period (Z, days)	3.60 (3.41, 3.91)	3.44 (3.26, 4.06)
Infectious period (D, days)	3.13 (2.74, 3.76)	3.30 (2.81, 4.36)
Reporting rate (α)	0.65 (0.60, 0.69)	0.69 (0.62, 073)
Effective reproductive number (R_e)	1.36 (1.14, 1.63)	0.99 (0.76, 1.33)

Downloaded from http://science.sciencemag.org/ on March 24, 2020

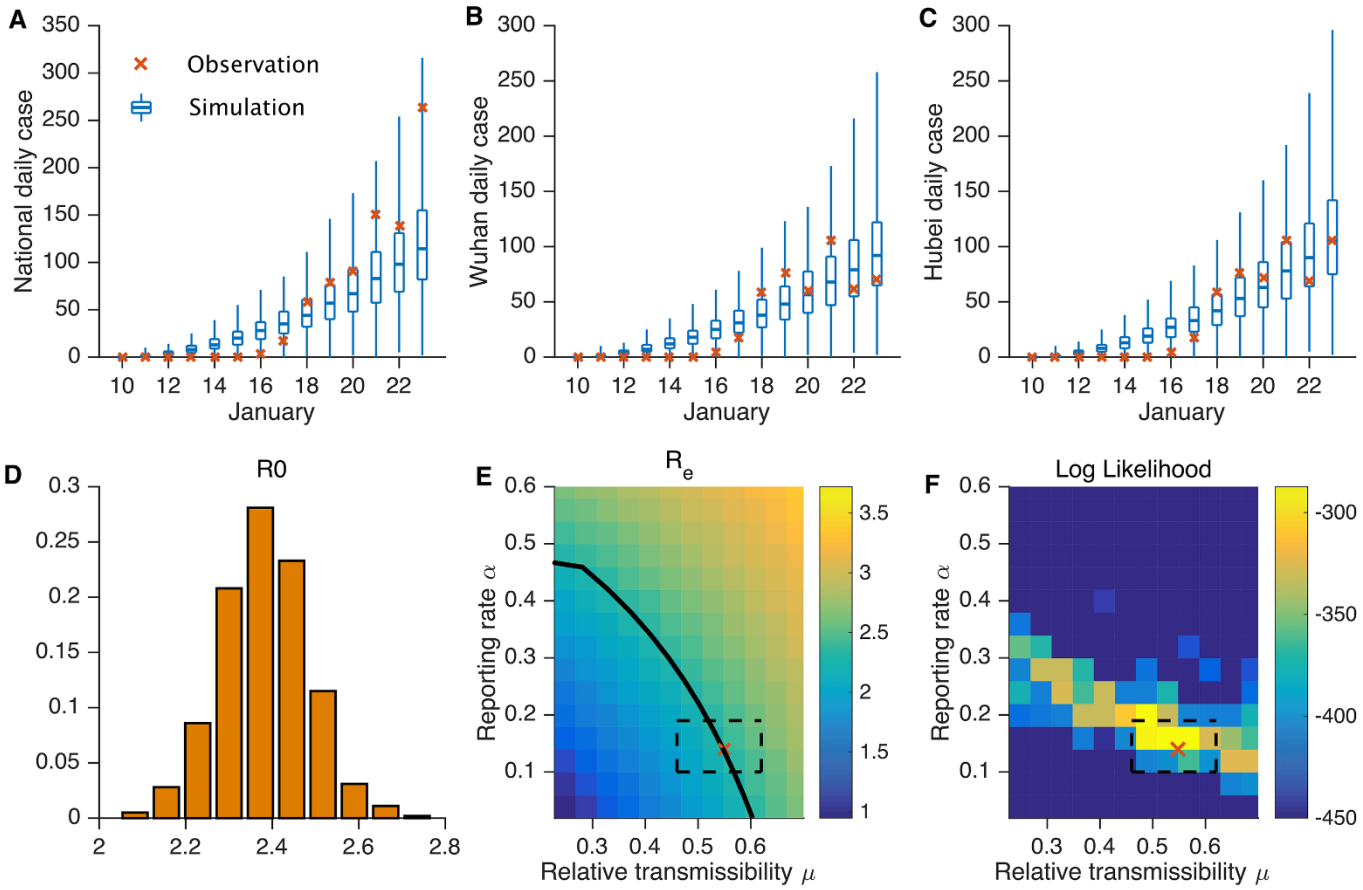

Fig. 1. Best-fit model and sensitivity analysis. Simulation of daily reported cases in all cities (**A**), Wuhan city (**B**) and Hubei province (**C**). The blue box and whiskers show the median, interquartile range, and 95% credible intervals derived from 300 simulations using the best-fit model (Table 1). The red x's are daily reported cases. The distribution of estimated R_e is shown in (**D**). The impact of varying α and μ on R_e with all other parameters held constant at Table 1 mean values (**E**). The black solid line indicates parameter combinations of (α,μ) yielding R_e = 2.38. The estimated parameter combination α = 0.14 and μ = 0.55 is shown by the red x; the dashed box indicates the 95% credible interval of that estimate. Log-likelihood for simulations with combinations of (α,μ) and all other parameters held constant at Table 1 mean values (**F**). For each parameter combination, 300 simulations were performed. The best-fit estimated parameter combination α = 0.14 and μ = 0.55 is shown by the red x (note that the x is plotted at the lower left corner of its respective heat map pixel, i.e., the pixel with the highest log likelihood); the dashed box indicates the 95% credible interval of that estimate.

Downloaded from http://science.sciencemag.org/ on March 24, 2020

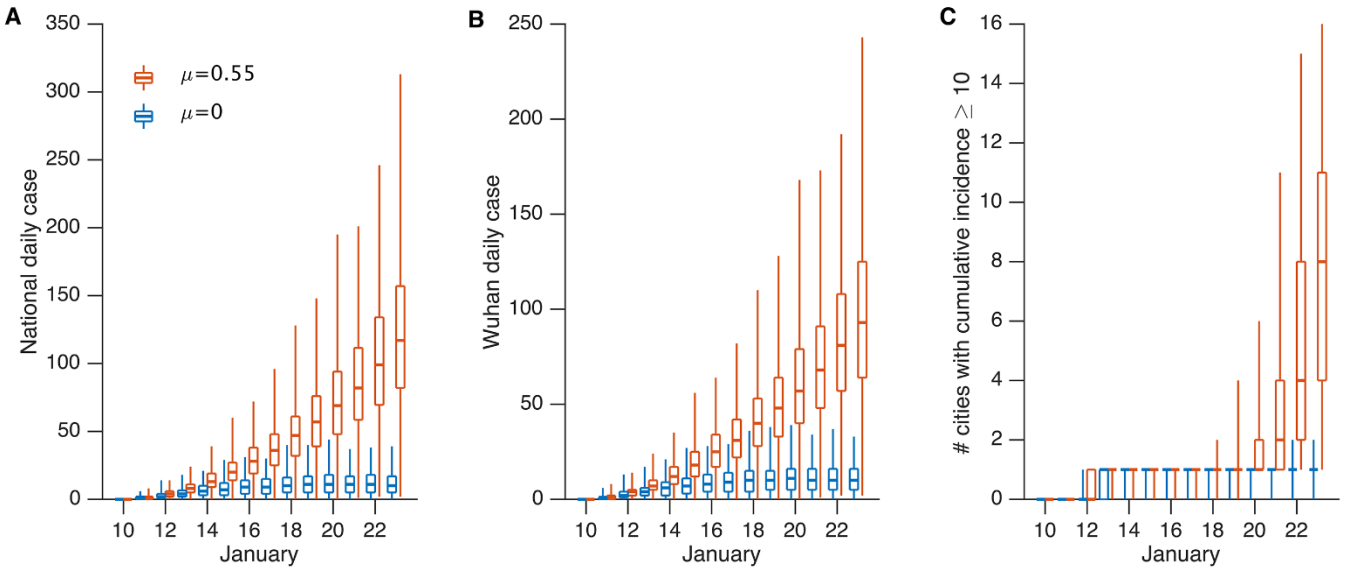

Fig. 2. Impact of undocumented infections on the transmission of SARS-CoV2. Simulations generated using the parameters reported in Table 1 with μ = 0.55 (red) and μ = 0 (blue) showing daily documented cases in all cities (**A**), daily documented cases in Wuhan city (**B**) and the number of cities with ≥ 10 cumulative documented cases (**C**). The box and whiskers show the median, interquartile range, and 95% credible intervals derived from 300 simulations.

Downloaded from http://science.sciencemag.org/ on March 24, 2020

Science

Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2)

Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang and Jeffrey Shaman

published online March 16, 2020

ARTICLE TOOLS http://science.sciencemag.org/content/early/2020/03/13/science.abb3221

SUPPLEMENTARY MATERIALS http://science.sciencemag.org/content/suppl/2020/03/13/science.abb3221.DC1

REFERENCES This article cites 22 articles, 3 of which you can access for free
http://science.sciencemag.org/content/early/2020/03/13/science.abb3221#BIBL

PERMISSIONS http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2020, American Association for the Advancement of Science

Downloaded from http://science.sciencemag.org/ on March 24, 2020