Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA

Max S. Y. Lau[a,1], Bryan Grenfell[b,c], Michael Thomas[d], Michael Bryan[d], Kristin Nelson[e], and Ben Lopman[e]

[a]Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA 30322; [b]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544; [c]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892; [d]Georgia Department of Public Health, Atlanta, GA 30303; and [e]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA 30322

Edited by Andrea Rinaldo, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, and approved August 6, 2020 (received for review June 13, 2020)

It is imperative to advance our understanding of heterogeneities in the transmission of SARS-CoV-2 such as age-specific infectiousness and superspreading. 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 spatiotemporal mechanistic framework to integrate individual surveillance data with geolocation data and aggregate mobility data, enabling a more granular understanding of the transmission dynamics of SARS-CoV-2. We analyze reported cases, between March and early May 2020, in five (urban and rural) counties in the state of Georgia. First, our results show that the reproductive number reduced to below one in about 2 wk after the shelter-in-place order. Superspreading appears to be widespread across space and time, and it may have a particularly important role in driving the outbreak in rural areas and an increasing importance toward later stages of outbreaks in both urban and rural settings. Overall, about 2% of cases were directly responsible for 20% of all infections. We estimate that the infected nonelderly cases (<60 y) may be 2.78 [2.10, 4.22] times more infectious than the elderly, and the former tend to be the main driver of superspreading. Our results improve our understanding of the natural history and transmission dynamics of SARS-CoV-2. More importantly, we reveal the roles of age-specific infectiousness and characterize systematic variations and associated risk factors of superspreading. These have important implications for the planning of relaxing social distancing and, more generally, designing optimal control measures.

SARS-CoV-2 | age-specific infectiousness | superspreading | social distancing | mobility data

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 infectiousness among different age groups (1–3). Also, while superspreading events have been documented, their impact and variation over space and time and associated risk factors have not yet been systematically characterized (1, 4–6).

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 SARS-CoV-2. For example, shelter-in-place orders likely change the movement patterns of a population, by reducing distance of travel. 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 (8–10). Deidentified 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). Geospatial location data and detailed data on spatial distribution of population are also important for capturing heterogeneous mixing in space (8–10). A key step is to enable individual-level model inference that can properly synthesize these data streams, 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 modeling Ebola outbreaks in western Africa (8, 9), we formulate an individual-level space–time stochastic model that describes the transmission of SARS-CoV-2 and captures the impact of statewide social distancing measure in the state of Georgia. Our model mechanistically integrates detailed individual-level surveillance data, geospatial 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

Significance

There is still considerable scope for advancing our understanding of the epidemiology and ecology of COVID-19. In particular, much is unknown about individual-level transmission heterogeneities such as superspreading and age-specific infectiousness. We statistically synthesize multiple valuable data streams, including surveillance data and mobility data, that are available during the current COVID-19 pandemic. We show that age is an important factor in the transmission of the virus. Superspreading is ubiquitous over space and time, and has particular importance in rural areas and later stages of an outbreak. Our results improve our understanding of the natural history of the virus and have important implications for designing optimal control measures.

Author contributions: M.S.Y.L. designed research; M.S.Y.L. performed research; M.S.Y.L. analyzed data; and M.S.Y.L., B.G., M.T., M.B., K.N., and B.L. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: msy.lau@emory.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2011802117/-/DCSupplemental.

First published August 20, 2020.

modeling framework also allows us to compute population-level epidemiological parameters such as the basic reproductive number R_0 and quantify the degree of superspreading over space and time.

Study Data

We analyzed a rich set of COVID-19 surveillance data collected by the Georgia Department of Public Health (GDPH), between March 1, 2020 and May 3, 2020, in five counties which had the largest numbers of cases. These counties include four (Cobb, DeKalb, Gwinnett, and Fulton) in the metro Atlanta area and one (Dougherty) in rural southwest Georgia. This dataset contains demographic information of 9,559 symptomatic cases which includes age, sex, and race, and symptom onset times. It also contains geolocation of the residences of cases. The GDPH Institutional Review Board has determined that this analysis is exempt from the requirement for IRB review and approval and informed consent was not required. Highly resolved population density data over 100 m $\times$ 100 m 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 the average change of movement distance within a county before and after the implementation of statewide social distancing measures. 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-h period and ending at the location where the same individual spent most of their time during the current 8-h period.

Results

Natural History Parameters and Effectiveness of Social Distancing. We estimate that the median value of R_0 across five counties was, overall, 3.30 with 95% CI [2.34, 5.2] before the shelter-in-place order. Dougherty County in the rural area had the largest prior-intervention R_0 (5.19 [5.01, 5.31]) and time-varying effective reproductive number R_{eff} at the earlier stage in March 2020 (Fig. 1). Our results suggest the shelter-in-place order was effective, and, in all of the counties, the R_{eff} declined below 1 in about 1 wk to 3 wk after the intervention. This is consistent with a recent study which shows that the effect of changes of mobility on reducing transmission may become noticeable after 9 d to 12 d (14). It is also worth noting that R_{eff} appears to begin to decline 1 wk to 2 wk before the 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 of 6.94 d [5.30, 7.37]. These estimates are largely consistent with the literature (15, 16).

Systematic Characterization of Superspreading. Superspreading refers to a phenomenon where certain individuals disproportionately infect a large number of secondary cases relative to an "average" infectious individual (whose infectiousness may be well represented by R_0). This phenomenon plays a key role in driving the spread of many pathogens, including Middle East Respiratory Syndrome and Ebola (8, 17). A common measure of the degree of superspreading is the dispersion parameter k, assuming that the distribution of the offspring (i.e., number of

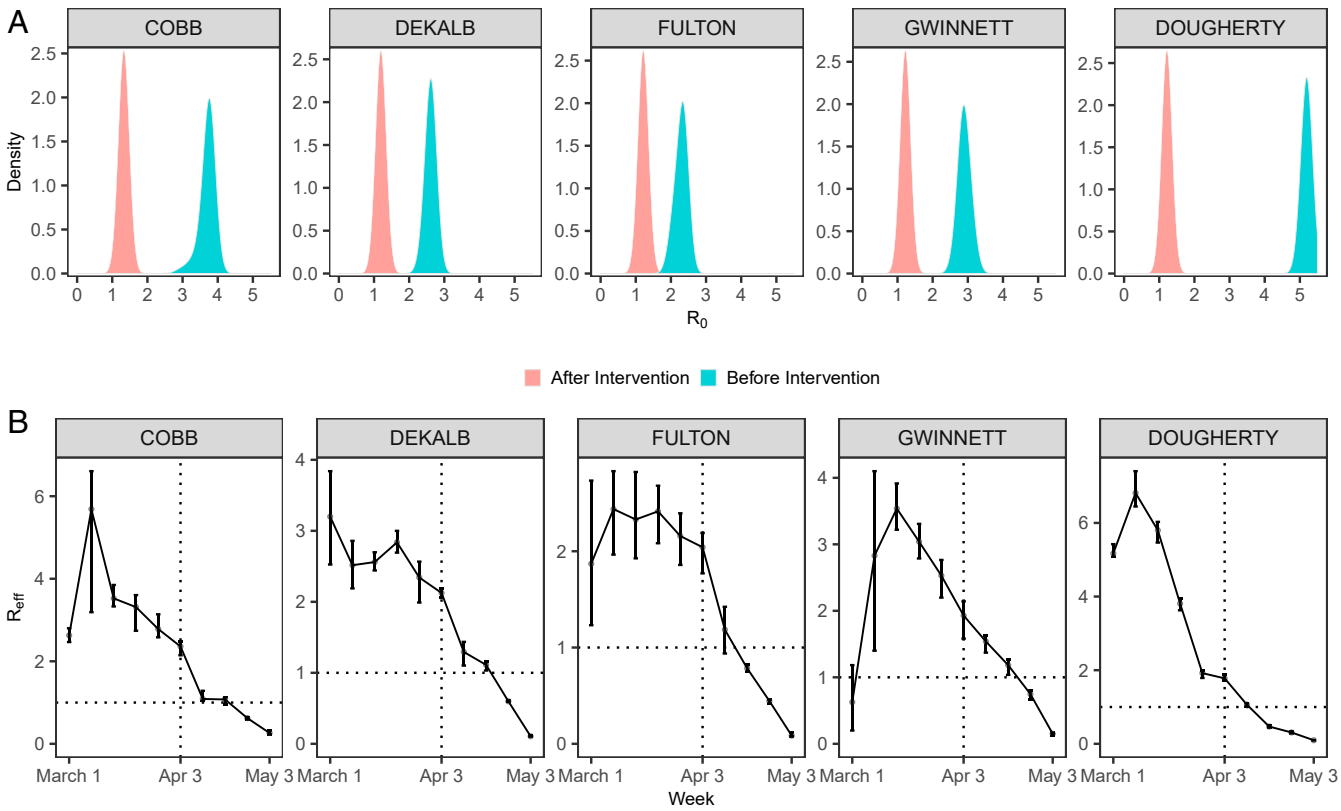

Fig. 1. Posterior distribution of (*A*) basic reproductive number R_0 and (*B*) the effective reproductive number R_{eff}, before and after the implementation of a statewide shelter-in-place order on April 3, 2020. Error bars represent 95% credible interval.

secondary cases generated) is a negative binomial with variance $\sigma^2 = \mu(1 + \mu/k)$, where μ is the mean (17). Generally speaking, a lower k corresponds to a higher degree of superspreading, and k less than 1 implies substantial superspreading. Our framework infers the transmission paths among all cases and therefore naturally generates the offspring distribution of each case (see *Materials and Methods*).

While superspreading of COVID-19 has been observed (1, 4–6), systematic characterization of its impact and variation (e.g., over time and space) and associated risk factors is lacking. Our results (Fig. 2*A*) suggest that superspreading is a ubiquitous feature during different periods (before and after the shelter-in-place order) of the outbreak. Superspreading may have a major impact for the rural area (Dougherty) among all counties (i.e., overall $k = 0.27$ for Dougherty, which is the lowest among all counties). Dougherty county has a disproportionately large outbreak compared to other more populated counties—having about only one-eighth 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 superspreading and large (prior intervention) R_0 in Dougherty (Fig. 1*A*). This is also consistent with the evidence of superspreading events due to a funeral in the area (18). The increasing significance of superspreading over time also highlights the importance of maintaining social distancing measures that may curtail close contacts (e.g., gatherings with densely packed crowds). Overall, the top 2% of cases (that generate highest number of infections) are responsible directly for about 20% of the total infections. Our results also show that younger infectees (<60 y) tend to be the main drivers of superspreading (Fig. 2*B*); infectiousness in this age group is also higher (see *Age-Specific Infectiousness*).

Age-Specific Infectiousness. The current COVID-19 pandemic indicates heterogeneity of susceptibility among different age groups (3, 19, 20). Much is unknown about the variation of infectiousness among different age groups (1–3). Our results (Fig. 3) suggest that the younger cases (<60 y) may be, overall, 2.78 [2.10, 4.22] times more infectious than elderly cases (≥60 y). Due to the very small number of reported cases in children (e.g., <15 y), we do not consider a finer age stratification (see also *Discussion*). We also test the robustness of

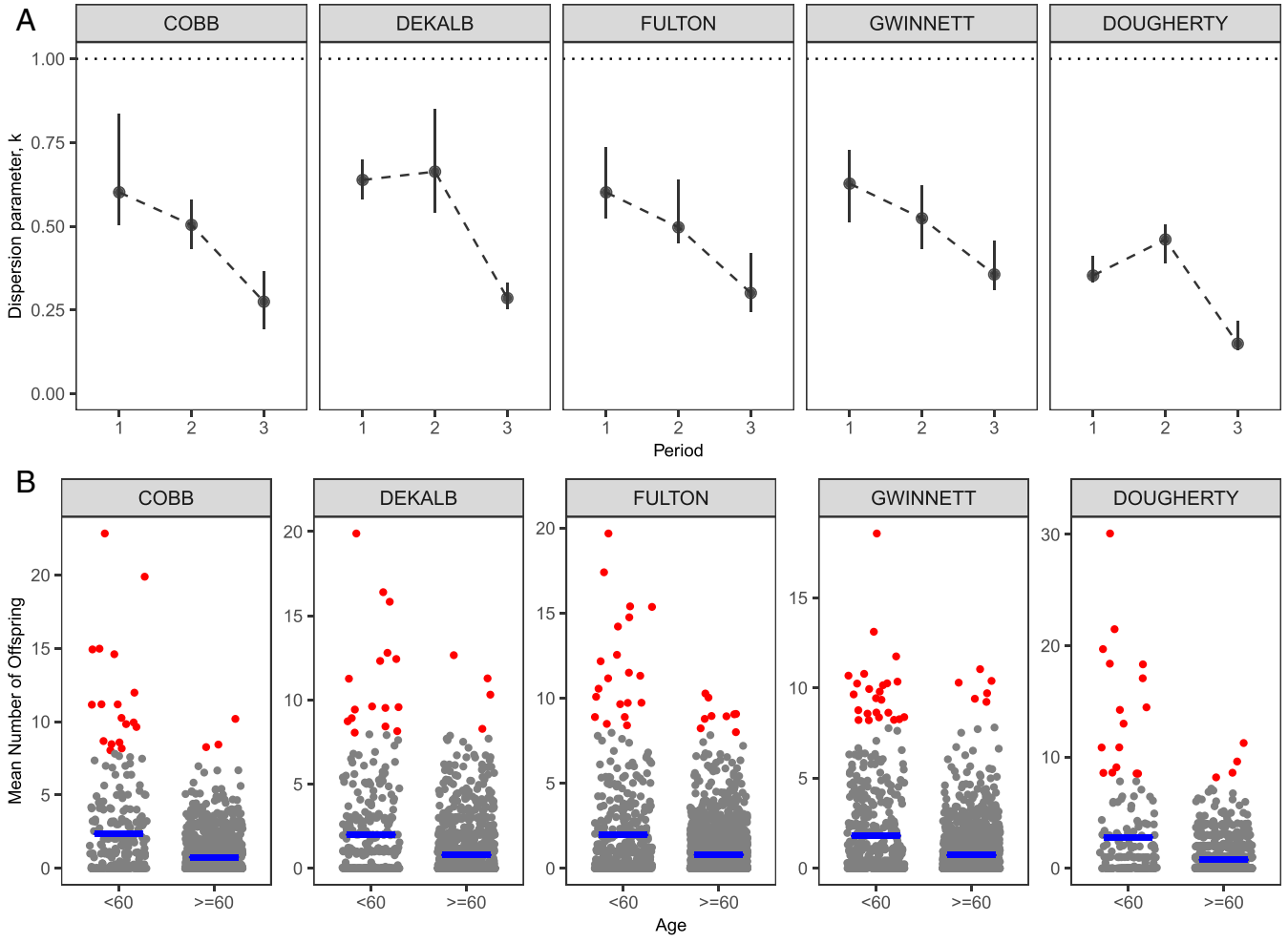

Fig. 2. (*A*) Degree of superspreading quantified by the dispersion parameter k (where $k < 1$ indicates significant superspreading) during different periods of the outbreak. Let T be the day of announcing the shelter-in-place order: Period 1 is time $t < T$, period 2 is $[T, T + 14)$, and, finally, period 3 is $t > T + 14$. Overall, about the top 2% of cases (that have highest mean number of offspring) directly infected 20% of the total infections. (*B*) Mean number of offspring generated by cases in each age group. Red dots represent those cases that have mean offspring ≥ 8. The younger age group (<60 y) tends to have more cases that produce an extreme number of offspring, and also a larger average (blue lines) of the mean number of offspring. Overall means of k also tend to be similar or smaller among the younger group: 0.53 for the younger group versus 0.82 for the older group in Cobb County, 0.57 versus 0.54 in DeKalb, 0.46 versus 0.64 in Fulton, 0.6 versus 0.6 in Gwinnett, and 0.39 versus 0.62 in Dougherty.

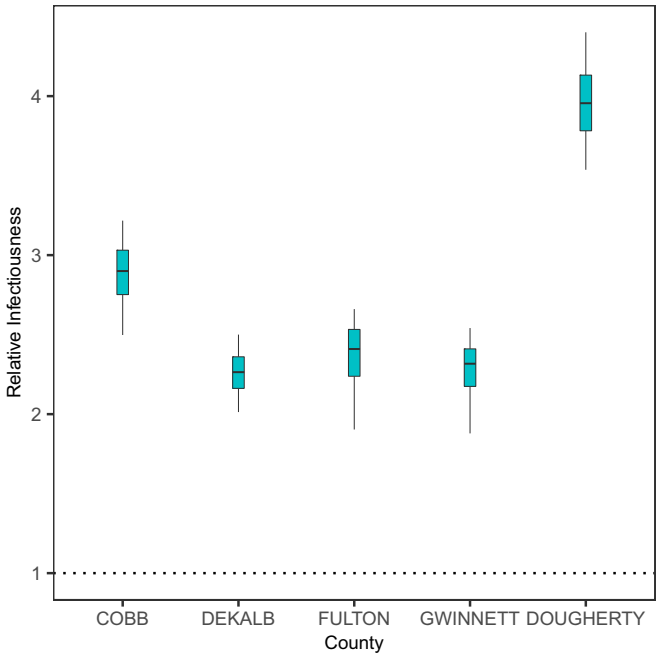

Fig. 3. Infectiousness of younger patients (<60 y) relative to the older patients (≥60 y).

these results to underreporting and take into account the discrepancy in the reporting rates of different age groups (see *Sensitivity Analysis*). We also test the robustness of our results to assumptions concerning relative susceptibility of the younger age group (see *Materials and Methods*). It is worth noting that we are not explicitly modeling biological factors such as viral loads that may potentially affect infectiousness. Instead, our model captures how likely it is that a case may generate secondary cases (see *Materials and Methods*), which, collectively and implicitly, accounts for multiple factors including viral loads and contact rates.

Sensitivity Analysis. Underreporting is a 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 testing 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, 19, 20). Such a discrepancy in reporting rates may potentially affect our estimation of age-specific infectiousness. We explore the effect of such underreporting on our results under these probable scenarios: We assume that, in March, probabilities of being reported for younger case (<60 y) and older cases are, respectively, 0.1 and 0.2, and, to account for increased testing capacity, these probabilities in April increase to 0.3 and 0.6, respectively. Details of how to include underreported cases are given in *Materials and Methods*. Fig. 4 shows that the younger age group remains to be more infectious than the older age group. The estimated impact of superspreading also appears to be robust: Estimated overall dispersion parameter k is 0.45 for Cobb County, 0.43 for Dekalb, 0.39 for Fulton, 0.49 for Gwinnett, and 0.32 for Dougherty.

We assumed that the older age group is twice as susceptible as the younger group (see *Materials and Methods*). To explore whether this assumption has an effect on the estimate of the relative infectiousness of the younger group, we fit the model by assuming equal susceptibility between the two age groups. The estimated infectiousness of the younger group relative to the older group is 2.84 [1.65, 3.60], which is similar to the estimate when we assume nonuniform susceptibility.

Discussion

Transmission dynamics of infectious diseases are often nonlinear and heterogeneous over space and time. It is important to exploit available data that are relevant to describing and estimating such complex processes. For COVID-19, a key step is to enable individual-level model inference that is able to statistically synthesize these data streams, beyond aggregate-level dynamics (2, 11, 12). In this paper, we incorporated multiple valuable data streams including formal surveillance data into our individual-level spatiotemporal transmission modeling framework, achieving a more granular mechanistic understanding of the dynamics and heterogeneities in the transmission of SARS-CoV-2.

Our results give similar estimates of important population-level epidemiological parameters such as 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 SARS-CoV-2, which is crucial for informing optimal interventions. We show that superspreading is an important and ubiquitous feature throughout the pandemic, and it may have a pivotal role in driving large outbreaks in rural areas. The increasing significance of superspreading 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 younger cases (<60 y) tend to be more infectious and to promote superspreading. Our results have important implications for designing more effective control measures—particularly, they highlight the importance of more targeted interventions.

Our study has a number of limitations. First of all, due to the lack of widely available testing, the underreporting 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 toward these possible underreporting scenarios, in

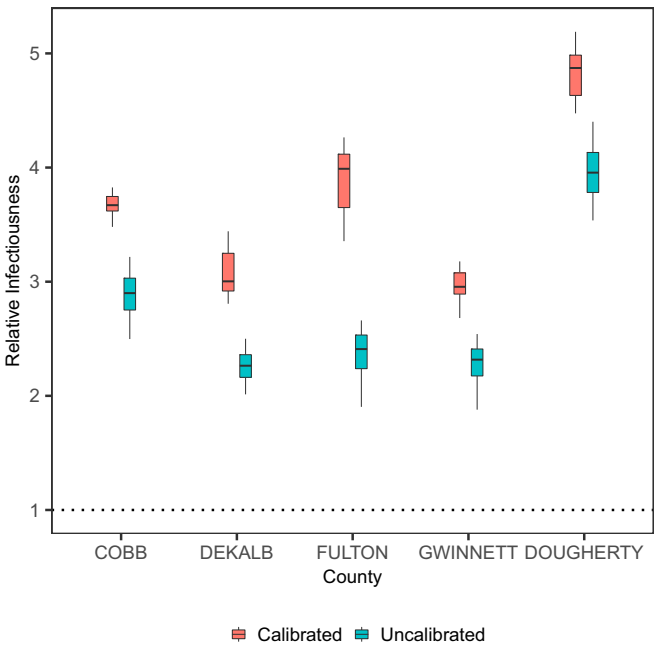

Fig. 4. Infectiousness of younger patients (<60 y) relative to the older patients (≥60 y), calibrated for underreported cases.

Sensitivity Analysis. Reassuringly, our main conclusions appear to be largely robust. Second, with the lack of detailed individual movement data, our model implicitly assumes transmission occurred mostly near people's homes. Nevertheless, most activities (hence transmissions) were likely to have clustered around homes due to the increasing public concern over the pandemic and announcements of shelter-in-place orders back in March and April. For example, Fulton County, one of the counties with the highest number of cases, closed its school as early as March 10, about 1 mo before the shelter-in-place order on April 3. As activities outside of homes (e.g., gatherings at pubs) may tend to facilitate clusters of transmissions and superspreading, our model may have thus underestimated both the infectiousness of younger adults who are more socially active and the degree of superspreading. Third, we only consider modeling age-specific infectiousness in two age groups ($<$60 y and $\geq$60 y). The model would tend to be overparameterized if we further break down the age groups that we have considered (e.g., by having a group for younger than 15 y), mainly due to an imbalance of the reported number of cases between age groups (particularly, markedly low reported numbers in the very young population). And, given the very few cases among younger children, our results are mostly relevant for younger adults and the elderly. 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 superspreading, future work in linking infectiousness directly with biological factors (e.g., age-specific viral loads) may shed further light.

Materials and Methods

Spatiotemporal Transmission Process Model. We formulate an age-specific spatiotemporal transmission modeling framework that allows us to infer the unobserved infection times and transmission tree among cases, integrating detailed individual-level surveillance data, geospatial location data, and highly resolved population density (grid) data, and aggregate mobility data. This approach also allows us to infer explicitly the distribution of the offspring (i.e., number of secondary cases generated) of each case. Our framework represents an extension of the models we developed (8, 9), which were validated generally and applied successfully to dissect the transmission dynamics of the Ebola outbreak in western Africa between 2014 and 2016. Fig. 5 gives a schematic overview of the model.

More specifically, we model the occurrence of a new infection as a first event in a nonhomogeneous Poisson process with a time-varying rate $r(t) = n(t) \times \beta(t)$, where $n(t)$ is the number of infectious individuals at time t and $\beta(t)$ is the time-dependent infectiousness of a case. We consider that $\beta(t)$ remained constant (i.e., the baseline infectiousness) before the statewide shelter-in-place order was announced, and declined exponentially after the order according to a rate ω. We also allow a primary infection rate α which may account for infections that are not explicitly modeled (e.g., noise or imported cases). We further assume that the two age groups ($<$60 y and $\geq$60 y) have their own (free) baseline infectiousness parameters to allow for age-specific infectiousness.

Spatially, it is assumed that the probability of the new infection being at a certain position (polar coordinates measured by distance r and direction θ) away from the source of infection is determined by the movement patterns of infectious individuals and the population density. Specifically, r and θ are drawn from an appropriate density function $g(r, \theta; \eta, \hat{\boldsymbol{s}})$ (9). Details of $g(r, \theta; \eta, \hat{\boldsymbol{s}})$ are also given in *SI Appendix, SI Text*. Recall that $\hat{\boldsymbol{s}}$ is the population density (within many 100 m $\times$ 100 m grids) across a county, and η is the mean of the spatial kernel f. The mean of movement distance η, after the issuing of the shelter-in-place order, is assumed to change according to the county-wise percentage change of movement distance (which is computed from the Facebook mobility data; see below). We calculated the average distance of all trips per day occurring in a county from March 23 to April 2, 2020, to establish baseline mobility in distance prior to implementation of statewide social distancing measures. We compared baseline mobility with mobility during the period from April 3 to May 5, after implementation of social distancing measures. Mobility data from Fulton, DeKalb, Gwinnett, and Cobb counties was readily available; data from Dougherty County was not. As a proxy for mobility in Dougherty

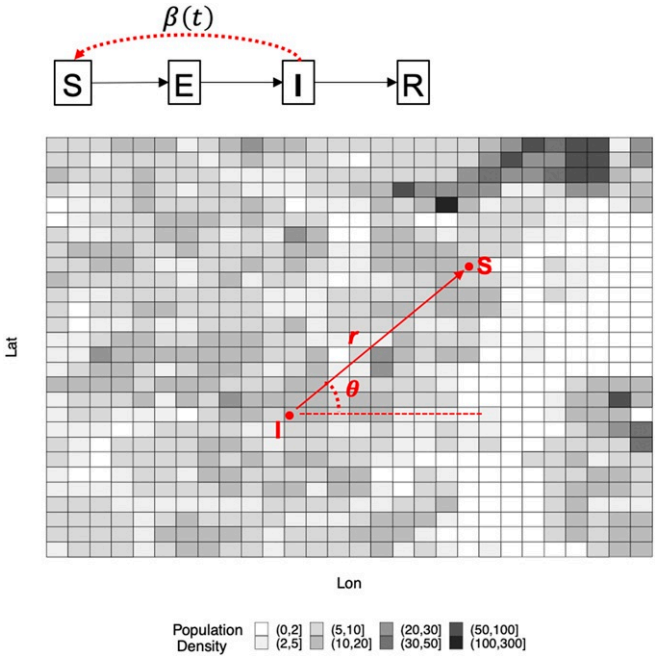

Fig. 5. Schematic description of our model. We model individual-level transmission of SARS-CoV-2, in continuous time and space and over a heterogeneous landscape with varying population density over 100 m $\times$ 100 m grids. Disease status of an individual is assumed to follow the susceptible–exposed–infectious–recovered framework. Infectiousness of an infectious individual $\beta(t)$ is time dependent and decreases due to social distancing. Likelihood of transmission from the infectious individual to a susceptible individual, at distance r and angle θ measured from the infectious source, is determined by 1) a spatial movement kernel (density) function f with mean η, 2) change of the mean movement distance due to a shelter-in-place order (informed by the aggregate mobility data), and 3) spatial distribution of population denoted by $\hat{\boldsymbol{s}}$ (i.e., detailed grid-level population density shown in the figure), which, all together, could more realistically account for heterogeneous mixing of individuals in the population (9).

County, we used data from Liberty and Glynn counties, which are also classified as "Small Metro" according to the 2010 US Census Bureau urbanicity classifications (21).

A new infection would go through an exponentially distributed incubation period with a mean parameter μ, before showing symptoms and becoming infectious. It is assumed that patients $<$60 y and $\geq$60 y old have, respectively, probabilities 0.06 and 0.17 of being hospitalized (22). The older age group is also assumed to be twice as susceptible as the younger group ($<$60 y) (3). The sojourn time between symptom onset and hospitalization follows $Exp(c)$. A nonhospitalized case is assumed to follow an exponential distribution with a mean of 14 d before recovery.

We estimate Θ (i.e., the parameter vector) in the Bayesian framework by sampling it from the posterior distribution $P(\Theta|\mathbf{z})$, where $\mathbf{z}$ are the data (8, 9, 23, 24). Denoting the likelihood by $L(\theta; \mathbf{z})$, the posterior distribution of Θ is $P(\Theta|\mathbf{z}) \propto L(\Theta; \mathbf{z})\pi(\Theta)$, where $\pi(\Theta)$ is prior distribution for Θ. Noninformative uniform priors for parameters in Θ are used (*SI Appendix*). 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. Details of the inferential algorithm are given in *SI Appendix, SI Text*. Posterior distributions of parameters are given in *SI Appendix*, Table S1.

Methods for Sensitivity Analysis. The number of total underreported 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 is the probability of being reported. We consider these probable scenarios: In March, probabilities of being reported for younger cases ($<$60 y) and older cases are, respectively, 0.1 and 0.2; to account for increased testing capacity, these probabilities in April increase to 0.3 and 0.6, 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 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 infectiousness.

Data Availability. The authors are not given permission to share the individual-level COVID-19 data. Data requests should be formally submitted to Public Health Information Portal (PHIP) of GDPH at https://dph.georgia.gov/phip-data-request. Computer code is available at https://github.com/msylau.

ACKNOWLEDGMENTS. We thank Dr. Laura Edison, Dr. Hannah Copper, and Dr. Allison Chamberlain for their efforts in leading and coordinating the research partnership between GDPH and Emory University. We also acknowledge Nishant Kishore and the COVID-19 Mobility Network for their support in 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. B.L. and K.N. acknowledge support from NIH/ National Institute of General Medical Sciences (Grants R01 GM124280 and 3R01GM124280-03S1), NIH/National Institute of Allergy and Infectious Diseases (Grant 3R01AI143875-02S1), and NSF (Grant 2032084).

1. Q. Bi *et al.*, Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: A retrospective cohort study. *Lancet Infect. Dis.* **20**, 911–919 (2020).
2. K. Prem *et al.*, The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study. *Lancet Public Health* **5**, e261–e270 (2020).
3. N. G. Davies *et al.*, Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Med.* **26**, 1205–1211 (2020).
4. S. Y. Park *et al.*, Early release—Coronavirus disease outbreak in call center, South Korea.*Emerg. Infect. Dis.* **26**, 1666–1670 (2020).
5. A. Endo, S. Abbott, A. J. Kucharski, S. Funk; Group CftMMoIDCW, Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res.* **5**, 67 (2020).
6. D. Adam *et al.*, Clustering and superspreading potential of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections in Hong Kong. https://doi.org/10.21203/rs.3.rs-29548/v1 (21 May 2020).
7. C. O. Buckee *et al.*, Aggregated mobility data could help fight COVID-19. *Science* **368**, 145–146 (2020).
8. M. S. Y. Lau *et al.*, Spatial and temporal dynamics of superspreading events in the 2014–2015 West Africa Ebola epidemic. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 2337–2342 (2017).
9. M. S. Y. Lau *et al.*, A mechanistic spatio-temporal framework for modelling individual-to-individual transmission—With an application to the 2014-2015 West Africa Ebola outbreak. *PLoS Comput. Biol.* **13**, e1005798 (2017).
10. M. S. Y. Lau *et al.*, A competing-risks model explains hierarchical spatial coupling of measles epidemics en route to national elimination. *Nat. Ecol. Evol.* **4**, 934–939 (2020).
11. M. U. G. Kraemer *et al.*, The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).
12. R. Li *et al.*, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* **368**, 489–493 (2020).
13. P. Maas, "Facebook disaster maps: Aggregate insights for crisis response & recovery" in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Association for Computing Machinery, Anchorage, AK, 2019), p. 3173.
14. H. S. Badr *et al.*, Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. *Lancet Infect. Dis.*, 10.106/S1473-3099(20)30553-3.
15. S. A. Lauer *et al.*, The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Ann. Intern. Med.* **172**, 577–582 (2020).
16. T. V. Inglesby, Public health measures and the reproduction number of SARS-CoV-2. *J. Am. Med. Assoc.* **323**, 2186–2187 (2020).
17. J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, W. M. Getz, Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005).
18. E. Barry, Days after a funeral in a Georgia town, coronavirus 'hit like a bomb.' *NY Times* , March 30, 2020. https://www.nytimes.com/2020/03/30/us/coronavirus-funeral-albany-georgia.html. Accessed 1 July 2020.
19. K. Sun, J. Chen, C. Viboud, Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: A population-level observational study. *Lancet Dig. Health* **2**, e201–e208 (2020).
20. E. Shim, A. Tariq, W. Choi, Y. Lee, G. Chowell, Transmission potential and severity of COVID-19 in South Korea. *Int. J. Infect. Dis.* **93**, 339–344 (2020).
21. US Census Bureau, *2010 Census urban and rural classification and urban area criteria*. https://census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/2010-urban-rural.html. Accessed 1 July 2020.
22. R. Verity *et al.*, Estimates of the severity of coronavirus disease 2019: A model-based analysis. *Lancet Infect. Dis.* **20**, 669–677 (2020).
23. M. S. Y. Lau, G. Marion, G. Streftaris, G. Gibson, A systematic Bayesian integration of epidemiological and genetic data. *PLoS Comput. Biol.* **11**, e1004633 (2015).
24. G. J. Gibson, E. Renshaw, Estimating parameters in stochastic compartmental models using Markov chain methods. *Math. Med. Biol.: A J. IMA* **15**, 19–40 (1998).