



Disease and healthcare burden of COVID-19 in the United States

Ian F. Miller¹✉, Alexander D. Becker¹, Bryan T. Grenfell^{1,2} and C. Jessica E. Metcalf^{1,2}

As of 24 April 2020, the SARS-CoV-2 epidemic has resulted in over 830,000 confirmed infections in the United States¹. The incidence of COVID-19, the disease associated with this new coronavirus, continues to rise. The epidemic threatens to overwhelm healthcare systems, and identifying those regions where the disease burden is likely to be high relative to the rest of the country is critical for enabling prudent and effective distribution of emergency medical care and public health resources. Globally, the risk of severe outcomes associated with COVID-19 has consistently been observed to increase with age^{2,3}. We used age-specific mortality patterns in tandem with demographic data to map projections of the cumulative case burden of COVID-19 and the subsequent burden on healthcare resources. The analysis was performed at the county level across the United States, assuming a scenario in which 20% of the population of each county acquires infection. We identified counties that will probably be consistently, heavily affected relative to the rest of the country across a range of assumptions about transmission patterns, such as the basic reproductive rate, contact patterns and the efficacy of quarantine. We observed a general pattern that per capita disease burden and relative healthcare system demand may be highest away from major population centers. These findings highlight the importance of ensuring equitable and adequate allocation of medical care and public health resources to communities outside of major urban areas.

SARS-CoV-2 was first identified in December 2019 in Wuhan, China⁴, and the first infection was detected in the United States on 7 January 2020 (ref. ⁵). The virus is both highly transmissible and virulent. Estimates of the basic reproductive number, R_0 , range from 1.4 to 6.49 (ref. ⁶), with an estimated overall case fatality rate of 1.4% (ref. ⁷), which is highly varied across age classes. As of 24 April 2020, >2,626,000 confirmed cases of SARS-CoV-2 and >181,000 deaths had been recorded globally. Currently, the cumulative reported incidence of COVID-19 in the United States is the highest in the world¹.

As the COVID-19 epidemic expands within the United States, a central focus of public health efforts will be limiting fatalities. A key driver of this outcome will be keeping the case burden of patients with COVID-19 within the treatment capacity of the healthcare system. If the medical system is overwhelmed, the standard of care for all individuals seeking medical care could be reduced, thereby exacerbating negative health outcomes⁸. Patients critically ill with COVID-19 might fare particularly poorly. High mortality rates within this group will probably be further compounded by shortages of intensive care facilities and/or access to mechanical ventilation equipment⁹. Patients without COVID-19 and who require care for other conditions will also be affected by the health system's inability to meet their needs.

Effective allocation of limited medical resources, such as healthcare workers, protective equipment and ventilators, is required to reduce the likelihood of the healthcare system being overwhelmed. However, to achieve this, information on the distribution of the burden of disease and how that burden aligns with healthcare system capacity is required.

Several factors probably contribute to the heterogeneous distribution of COVID-19 burden across the United States. The first of these is demography. The incidence of COVID-19 consistently rises with increasing age^{2,3} (noting that incidence observed from diagnostic testing does not necessarily reflect total infections). This age-dependent pattern of infection seems to be largely driven by differences in susceptibility and symptomatic infection rates between age classes rather than by differences in transmission potential^{3,10}. Rates of hospitalization and intensive care unit admission are also higher in individuals aged >60 years than in younger age classes¹¹. Thus, variation in age structure between counties could lead to differences in the per capita burden of disease between regions. Access to healthcare could also affect the distribution of COVID-19 burden. Many rural areas of the United States might have insufficient or no resources to provide acute or critical care. Residents of such areas could therefore be at increased risk for insufficient treatment. Finally, limited healthcare system capacity in rural areas could lead to an unexpected influx of cases to hospitals in more densely populated regions.

The temporal distribution of COVID-19 spread could also contribute to heterogeneity in disease burden across the United States. The magnitude and timing of the epidemic peak, for example, determine the minimum healthcare system capacity needed to provide adequate care. However, obtaining accurate predictions of the epidemic peak is often challenging in emerging outbreaks due to limited and often unreliable data on incidence, as well as to the challenges associated with modeling the effects of rapidly deployed and changing mitigation efforts. County-level variability in testing standards and efforts^{12,13}, nonpharmaceutical interventions (NPIs) such as social distancing¹⁴ and outbreak onset¹⁵, and a lack of serological data also limit efforts to accurately model epidemic trajectories beyond several weeks. By contrast, projections of cumulative disease burden are less hindered by these challenges as these are not aimed at describing an epidemic time course. Although such projections miss the nuance of the intensity and timing of outbreaks, their estimates of the spatial footprint of disease burden contain core information relevant to informing resource distribution. Comparing the expected cumulative number of critical and severe infections against healthcare resources in each county in the United States allows for the identification of regions that may experience particularly high disease burden. Furthermore, analysis of simulations of multiple transmission scenarios (for example, different

¹Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. ²Woodrow Wilson School of Public Affairs, Princeton University, Princeton, NJ, USA. ✉e-mail: ifmiller@princeton.edu

Table 1 | Policy summary

Background	The now widely prevalent new coronavirus SARS-CoV-2 poses a grave public health threat in the United States, particularly in areas where cases could overwhelm healthcare systems. Allocation of emergency medical and public health resources can help to mitigate this problem, but this requires efficient identification of regions that are severely affected.
Main findings and limitations	Leveraging the relationship between age and COVID-19 disease severity, we projected the cumulative disease burden for each of 3,142 counties and county equivalents in the United States using a modified SEIR model. Comparing these burdens with hospital capacity, we identified counties that will probably be consistently, heavily affected relative to the rest of the country, and found that the per capita burden of disease and the relative strain on the healthcare system could be higher in rural areas. To circumvent uncertainty surrounding the time course of the epidemic, we projected the long-term cumulative burden of COVID-19 rather than peak burden. These projections ignore the temporal component of disease spread and assume that cumulative incidence is identical among counties. Due to these limitations, our results should be interpreted as an approximation of the spatial distribution of COVID-19 burden rather than as a set of epidemiological forecasts.
Policy implications	Our findings highlight the importance of ensuring equitable and adequate distribution of medical care and public health resources to urban and rural areas to reducing the total mortality associated with COVID-19.

contact patterns) allows for possible identification of those regions with consistently high disease burdens without needing to forecast an exact epidemic trajectory.

Here, we project the cumulative case burden (case numbers) and cumulative per capita burden (cases per person) of severe and critical COVID-19 cases in each county within the United States by combining demographic data and age-specific risk factors under the assumption that 20% of the population becomes infected. We calculate the cumulative healthcare system burden, using case/bed ratio that each county could experience, as its own residents (and those from nearby counties with limited or nonexistent medical resources) seek care. We repeat this analysis for a range of transmission scenarios, map the expected burden of COVID-19 for each scenario and identify those regions consistently expected to experience the highest cumulative burden of disease. A summary of the main findings, limitations and policy implications of this study is given in Table 1.

We developed a modified, age-stratified susceptible–exposed–infected–recovered (SEIR) epidemic model (based on the model of Davies et al.³) to project the number of COVID-19 cases for all counties (and county equivalents such as independent cities) in the United States. In this model, susceptible individuals (*S*) become infected in a density-dependent fashion and enter the exposed (*E*) class, before eventually becoming either asymptotically infected (*I_A*) or mildly symptomatic (but not yet clinically presenting) (*I_P*). Following published estimates³, we assume that relative susceptibility to infection and the fraction of individuals who become mildly symptomatic rather than asymptotically infected are higher in older age classes than in younger age classes. Individuals in the *I_P*

class eventually become fully symptomatic (*I_C*). Asymptomatic and symptomatic individuals recover with immunity to classes *R_A* and *R_S*, respectively. All individuals in the infected classes (*I_A*, *I_P*, *I_C*) are infectious; however, our model assumes that the relative infectiousness of asymptomatic individuals is scaled by factor *b_A*, and the relative infectiousness of fully symptomatic individuals is scaled by factor *b_C* to account for the effects of case isolation and quarantine. Mixing between individuals of different age classes is determined by a parameter θ . For $\theta=1$, mixing patterns reflect empirically measured rates for the United Kingdom¹⁶. For $\theta=0$, mixing patterns are homogeneous; for $0 < \theta < 1$, mixing patterns are intermediate. This model aims to specifically project the age distribution of cases over a wide variety of transmission scenarios, and is not intended to produce epidemiological forecasts. As such, we include epidemiological details that could result in differences in disease burden between age classes, such as age-specific mixing patterns and rates of symptom presentation. However, we do not vary the components of our model linked to interventions (for example, transmission rate, mixing patterns) over time or by location.

We investigated a scenario in which 20% of the population in each county becomes infected. A 20% cumulative infection rate represents a pessimistic scenario over the next few months, but perhaps this will be an optimistic scenario beyond that time frame¹⁷. A 20% cumulative infection rate is independent of *R₀* and is not equivalent to 20% of the herd immunity threshold. We intentionally ignored spatial variation in the progression of the epidemic, to simplify comparisons of disease burden between regions.

As we aim to provide general estimates of relative distribution of disease burden rather than make precise predictions of case load over time, we sought to identify patterns of disease burden that are robust to different assumptions about the dynamics of epidemiological spread. Accordingly, we varied our assumptions about the overall transmissibility of COVID-19, age structure of contact patterns and the contributions of fully symptomatic individuals to transmission. For each set of assumptions, we simulated our model for each county in the United States using demographic data from the 2018 American Community Survey¹⁸. We then extracted the number of individuals in each age class who had become symptomatically infected by the time the cumulative population infection rate had reached 20%. We present detailed results for the most optimistic scenario and most pessimistic scenario. In the optimistic scenario (transmission, *R₀*=2, relative infectivity of fully symptomatic individuals, *b_C*=0.1 and mixing structure, $\theta=1$; see Methods) transmission is slow, fully symptomatic individuals are effectively quarantined and mixing patterns exhibit a strong age structure, potentially decreasing transmission from asymptotically infected (and thus nonquarantined) individuals in less vulnerable age classes (such as children) to individuals in more vulnerable age classes (such as the elderly). By contrast, the pessimistic scenario (*R₀*=5, $\theta=0$, *b_C*=1) is characterized by high transmission, well-mixed contact patterns and ineffective quarantine. Results for 25 alternate combinations of *R₀*, θ and *b_C* are summarized in Extended Data Figs. 1–4.

To evaluate the sensitivity of our results to the effects of crowding on transmission and epidemic size, we also investigated alternative scenarios in which *R₀* increases as a linear function of urban population, from *R₀*=2 in counties with 0% of residents living in urban areas, to *R_{0,max}* in counties with 100% of residents living in urban areas. In an optimistic scenario we set *R_{0,max}* to 3 (other parameters: $\theta=1$, *b_C*=0.1) and, in a pessimistic scenario, we set *R_{0,max}* to 5 (other parameters: $\theta=0$, *b_C*=1). Disease burden in each county was calculated when the cumulative number of infections reached 20% of the herd immunity threshold multiplied by the population size, rather than 20% of the population size (see Methods). The relationship between crowding and *R₀* has not been definitively established and, as such, these results should be interpreted cautiously.

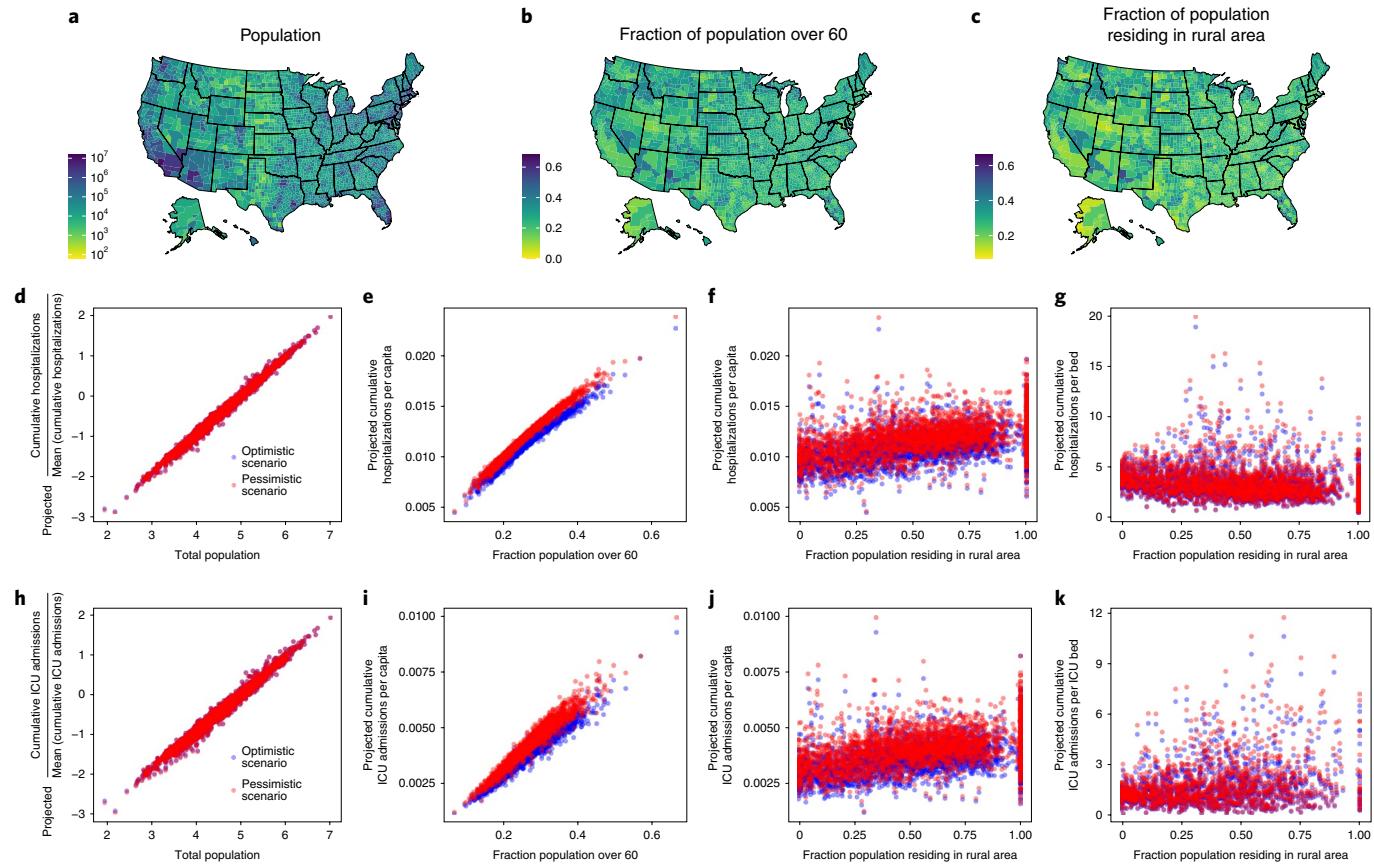


Fig. 1 | Population characteristics of the United States and their relationships with disease burden. **a**, Population of each county. **b**, Fraction of individuals within each county >60 years of age. **c**, Fraction of the population of each county classified as living in a rural area according to the 2010 US Census²³. **d–k**, The relationship between population characteristics (x axes) and metrics of disease burden (y axes) for the optimistic transmission scenario (blue) and pessimistic scenario (red). **d**, Total population vs. projected cumulative hospitalizations. **e**, Fraction of population over 60 vs. projected cumulative hospitalizations per capita. **f**, Fraction of population residing in rural area vs. projected cumulative hospitalizations per hospital bed. **g**, Total population vs. projected cumulative ICU admissions. **h**, Fraction of population over 60 vs. projected cumulative ICU admissions per capita. **i**, Fraction of population residing in rural area vs. projected ICU admissions per capita. **j**, Fraction of population residing in rural area vs. projected ICU admissions per ICU bed.

Using our projections of cumulative symptomatic infections, we further estimated the number of severe cases (that is, requiring hospitalization) and critical cases (that is, requiring intensive care) using published rates of these outcomes for various age classes¹¹. In all transmission scenarios, the areas with high relative burdens of hospitalizations and intensive care unit (ICU) admissions generally had large populations (Figs. 1a,d,h, 2a,d and 3a,d). However, we observed the opposite pattern for the per capita burden of hospitalizations and ICU admissions, which were distributed heterogeneously and were higher in rural areas than in major population centers (Fig. 1c,f,j). Due to the positive correlation between age and disease severity, areas with the highest per capita burden were those with the highest percentages of individuals >60 years of age (Figs. 1b,e,i, 2b,e and 3b,e). Although more elderly age classes were disproportionately affected in the pessimistic transmission scenario (Figs. 2g and 3g), the sets of counties with very high projected burdens of per capita hospitalizations and ICU admissions remained similar across different transmission scenarios. Indeed, of the 315 counties at or above the 90% quantile of per capita hospitalization in the optimistic transmission scenario (Fig. 1, legend), 308 counties were also at or above this quantile in the pessimistic scenario. The median percentage of people residing in rural areas among these 308 counties was 100%, which is significantly greater than the median of all counties (57.54%, Mann–Whitney $U=696,348$, $n_1=308$, $n_2=3,142$, two-sided $P<2.2\times 10^{-16}$).

Of the 315 counties at or above the 90% quantile of per capita ICU admissions in the optimistic transmission scenario, 313 were also at or above this quantile in the pessimistic scenario. Again, the median percentage of people residing in rural areas among these 313 counties was 100%, significantly greater than the median of all counties (Mann–Whitney $U=725,670$, $n_1=313$, $n_2=3,142$, two-sided $P<2.2\times 10^{-16}$).

Next, we evaluated how projected case burdens aligned with healthcare system capacity. We obtained data on the number of hospital beds and ICU beds in each county from the American Hospital Association 2018 annual survey¹⁹. We distributed cases to healthcare systems within and outside of their county of origin based on an allocation algorithm (see Methods). This algorithm distributes severe and critical cases based on relative distance and the relative capacity of healthcare systems to provide care (quantified as the number of hospital beds and ICU beds, respectively). The majority of cases originating from within a county with substantial medical resources stay within that county. Most severe and critical cases originating from within a county with few hospitals or ICU beds are allocated to nearby counties with greater care capacity. All severe or critical cases originating in a county that lacks the capacity to provide appropriate care entirely are distributed to nearby counties.

The maps of relative hospitalizations per bed (Fig. 2c,f) and relative ICU admissions per bed (Fig. 3c,f) indicate those counties

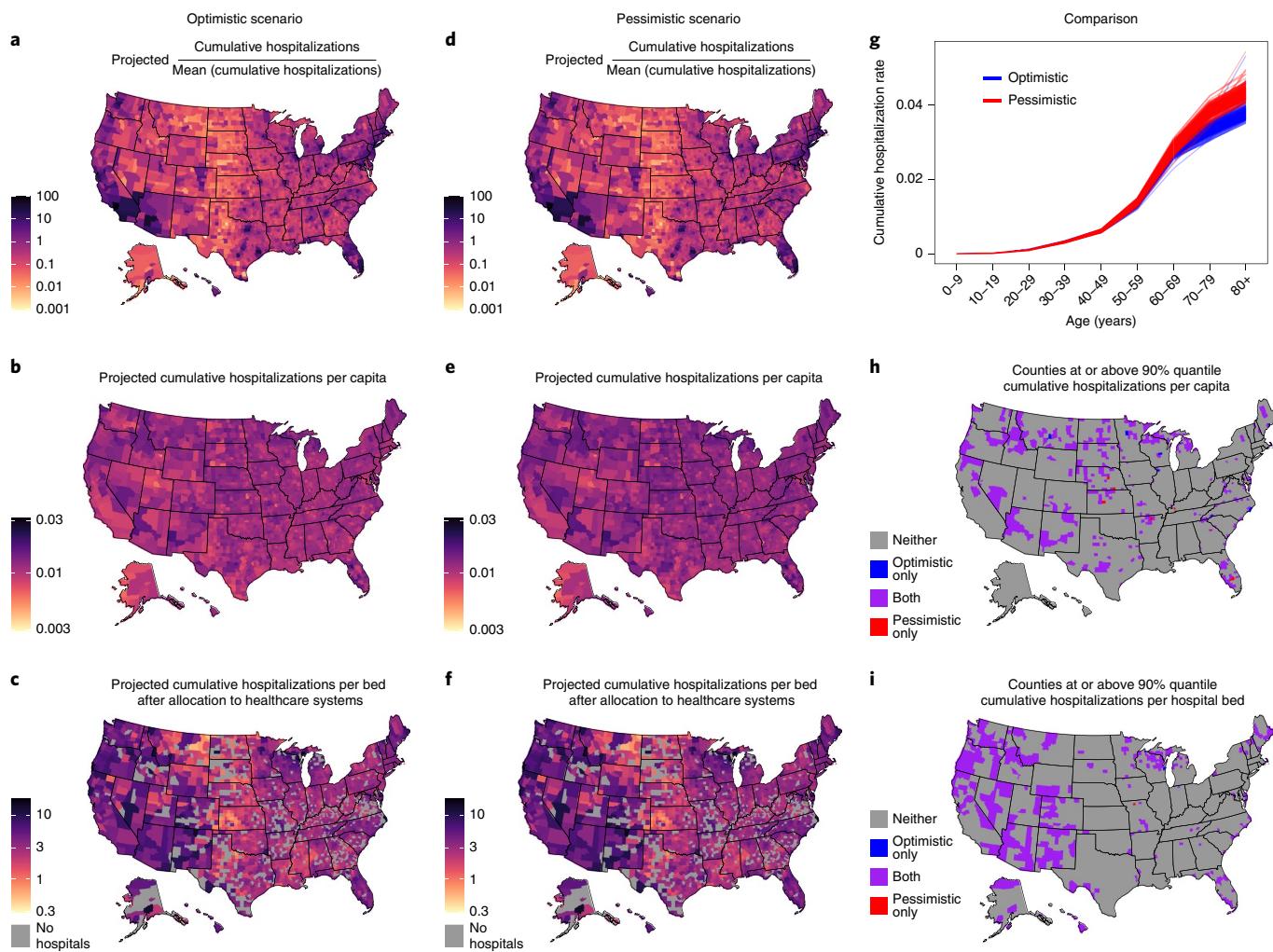


Fig. 2 | Projected cumulative burden of hospitalizations in the United States. **a–c**, Optimistic scenario; **d–f**, pessimistic scenario. **a, d**, Relative number of hospitalizations in each county. **b, e**, Number of projected hospitalizations per capita in each county. **a, b, d, e**, Cases not yet allocated to healthcare systems. **c, f**, Cumulative number of hospitalizations per hospital bed after allocation of cases to healthcare systems. **g**, Cumulative fraction of each age class hospitalized in each transmission scenario. Each of the 315 lines for each transmission scenario represents a different county. **h, i**, Counties estimated to be in the 90% quantile of hospitalizations per capita (**h**) and hospitalizations per hospital bed (**i**) (after case allocation). Colors in **h, i** indicate whether these counties were estimated to be in the 90% quantile in the optimistic scenario, the pessimistic scenario, both or neither. A high-resolution version of this figure is provided in Supplementary Information.

expected to experience a higher burden of disease relative to medical resources. The burden of cases relative to hospital and ICU beds was generally highest away from urban centers in counties with substantial rural populations (Extended Data Fig. 5d,h). Several regions have a high concentration of counties with a high burden, including much of the western United States, the northern Midwest, Florida and northern New England. These patterns are robust to assumptions about transmission rates and age-specific mixing patterns. The optimistic and pessimistic transmission scenarios each identified 248 counties as being at or above the 90% quantile of cumulative hospitalizations per hospital bed; 247 counties were identified in both transmission scenarios. The median percentage of people residing in rural areas among these 247 counties (38.97%) is lower than the median for all counties with hospital beds (51.82%, Mann–Whitney $U=246,652$, $n_1=247$, $n_2=2,478$, two-sided $P=4.64 \times 10^{-7}$). Nevertheless these data indicate that the healthcare system burden is not concentrated in urban centers. In the case of ICU admissions per bed, all of the 136 counties identified as being at or above the 90% quantile were the same for both transmission scenarios. These 136 counties (median percentage of residents living

in rural areas = 31.11%) were not identified as being more rural than all counties with ICU beds (median percentage of residents living in rural areas = 36.21%, Mann–Whitney $U=85,746$, $n_1=136$, $n_2=1,353$, two-sided $P=0.19$) but, again, a pattern emerges of healthcare system burden not being concentrated in urban areas.

For analyses where R_0 varied as a function of the percentage of population residing in urban areas, the per capita and per hospital and ICU bed burdens of disease were not generally higher in rural areas, but rather were distributed heterogeneously across urban and rural areas (Extended Data Fig. 6). Counties at or above the 90% quantile for various metrics of disease burden were less rural than comparable counties, but were not heavily concentrated in the urban end of the urban–rural distribution (Extended Data Figs. 9–10). Otherwise, results from these analyses (Extended Data Figs. 6–10) largely agree with those presented above, indicating that our finding that disease burden is not expected to be concentrated only in urban areas is robust to assumptions about the effects of crowding on transmission patterns and epidemic size.

Even with unprecedented efforts to rapidly develop a vaccine²⁰, a pharmaceutical intervention against COVID-19 is unlikely to be

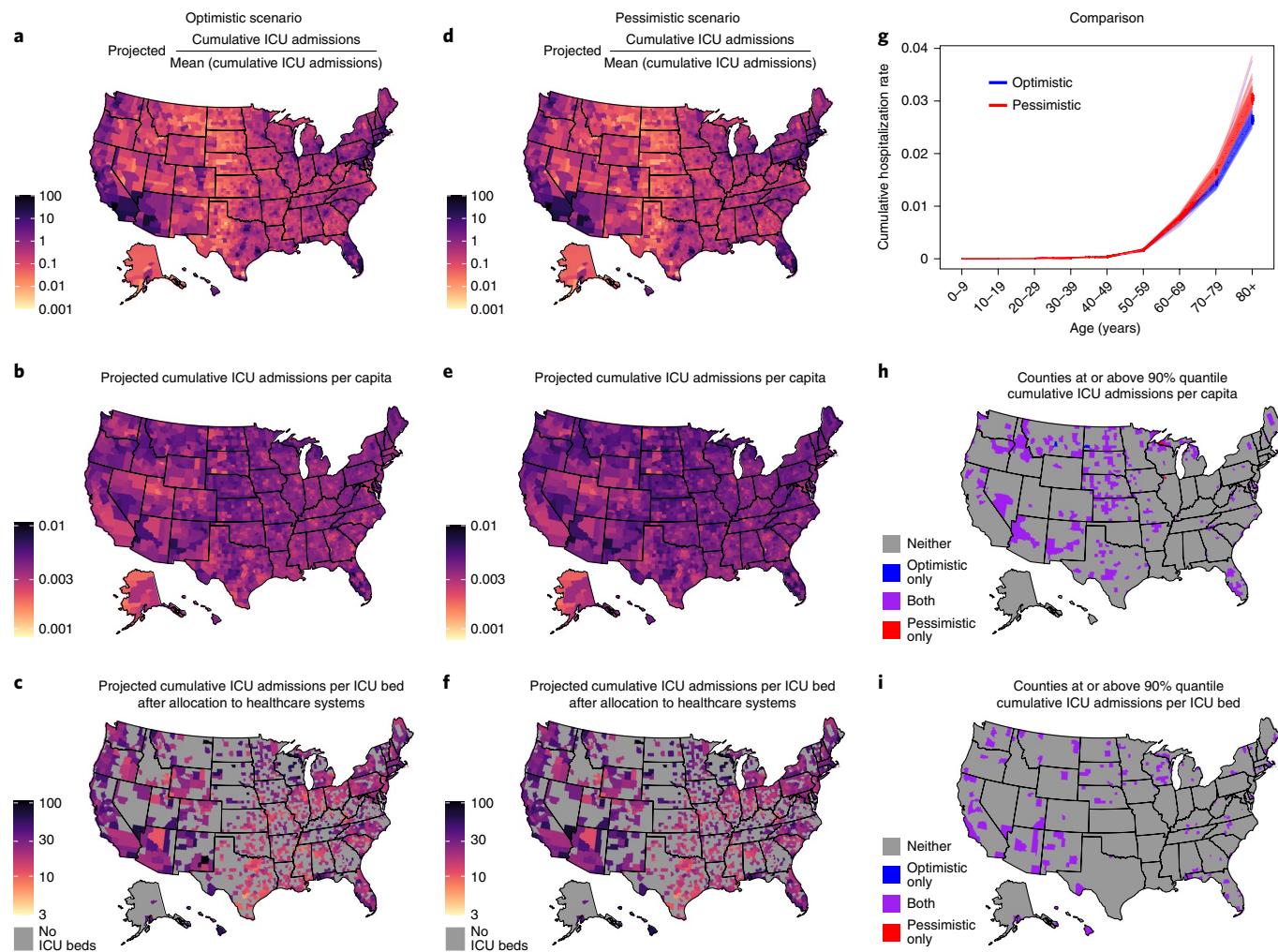


Fig. 3 | Projected cumulative burden of ICU admissions in the United States. **a–c**, Optimistic scenario; **d–f**, pessimistic scenario. **a,d**, Relative number of ICU admissions in each county. **b,e**, Number of projected ICU admissions per capita in each county. **a,b,d,e**, Cases not yet allocated to healthcare systems. **c,f**, Cumulative number of ICU admissions per ICU bed after cases have been allocated to healthcare systems. **g**, Cumulative fraction of each age class requiring ICU admission in each transmission scenario. Each of the 315 lines for each transmission scenario represents a different county. **h,i**, Counties estimated to be in the 90% quantile of ICU admissions per capita (**h**) and ICU admissions per ICU bed (**i**) (after case allocation). Colors in **h,i** indicate whether these counties were estimated to be in the 90% quantile in the optimistic scenario, the pessimistic scenario, both or neither. A high-resolution version of this figure is provided in Supplementary Information.

available in the near future. SARS-CoV-2 transmission is expected to continue over the coming months and will probably affect every locality in the United States. We aimed to identify counties that consistently emerge as being likely to experience a large burden of disease on their population and healthcare systems (across a range of assumptions about transmission patterns). We identified several regions in need of additional support, including much of the western portion of the country, the northern Midwest, Florida and northern New England. At a fine geographical scale, our results suggest that considerable rural–urban inequities exist, with the per capita burden of disease being higher away from major population centers.

Before even considering the increased case burden that these more rural places are projected to experience relative to the rest of the country, it is evident that hospitals—and, to a greater extent, hospitals with the capacity to provide intensive care—are unevenly distributed. Many regions have limited, or no, facilities equipped to provide the type of acute or critical care required to treat COVID-19 (ref. ¹⁹). Case fatality rates in these regions could rise above the national average if people are unable to access care.

Bolstering the capacity of rural health systems, ensuring equitable access to care and implementing public health measures such as testing and contact tracing in both urban and rural areas should be central goals of COVID-19 management strategies in the United States. While the healthcare systems of major population centers were not identified as weak spots in our analysis, they do service a much larger number of people. Given the consequences of their potential failure, they should remain a priority for response efforts.

Our findings are robust to different assumptions about transmission patterns. However, it is imperative that they be interpreted in the context of our methodology. We were deliberately conservative in not considering the impact of potential therapeutics and vaccines. Our results only underscore the urgency of developing these interventions. Likewise, we did not consider the impact of other NPIs such as social distancing. Our findings point to the importance of implementing these measures in urban and rural regions. We specifically did not attempt to predict the epidemic peak timing or magnitude. Given the time-invariant scenario we model (that is, 20% of the population acquires infection), it is likely that our

projections will not precisely match future observed patterns of disease burden in the short term, as many regions are still in the early phases of their epidemics, or in the long term, as the timing, extent and efficacy of interventions will vary among regions. However, our results provide an approximation of the expected patterns of burden rooted in basic features of demography and health system capacity. Notably, we did not consider how other factors linked with an increased risk of severe disease, such as comorbidities²¹ (for example, hypertension, pulmonary disease), or decreased access to medical care, such as noninsurance rate and socioeconomic status²², might exacerbate disease burden in certain regions. Incorporating such factors into mathematical models and their forecasts is an essential area of future research, and could reveal additional 'hotspots' of disease burden that were not identified in our analyses, which considered the role of demography alone. Future work should also seek to identify if and where disease burden is disproportionately high in certain racial or minority groups. Finally, we urge public health officials using our results to carefully consider location-specific details and nuances not explicitly included in our analyses when planning their response, and to focus on patterns of relative burdens rather than projections for individual counties.

In conclusion, we have identified areas in the United States expected to be particularly heavily affected by COVID-19. Our findings suggest that ensuring equitable allocation of medical care and public health resources to communities away from major population centers will be crucial as the country attempts to mitigate the consequences of the ongoing COVID-19 epidemic.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-020-0952-y>.

Received: 6 April 2020; Accepted: 22 May 2020;
Published online: 16 June 2020

References

1. *Coronavirus Disease 2019 (COVID-19) Situation Report 95* (World Health Organization, 2020).
2. COVIDView Week 13 (Centers for Disease Control and Prevention, 2020).
3. Davies, N. G. et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. Preprint at *medRxiv* <https://doi.org/10.1101/2020.03.24.20043018> (2020).
4. Wang, C., Horby, P. W., Hayden, F. G. & Gao, G. F. A novel coronavirus outbreak of global health concern. *Lancet* **395**, 470–473 (2020).
5. Holshue, M. L. et al. First case of 2019 novel coronavirus in the United States. *N. Engl. J. Med.* **382**, 929–936 (2020).
6. Liu, Y., Gayle, A. A., Wilder-Smith, A. & Rocklöv, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.* **27**, 1–4 (2020).
7. Wu, J. T. et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat. Med.* **26**, 506–510 (2020).
8. Emanuel, E. J. et al. Fair allocation of scarce medical resources in the time of Covid-19. *N. Engl. J. Med.* **382**, 2049–2055 (2020).
9. White, D. B. & Lo, B. A framework for rationing ventilators and critical care beds during the COVID-19 Pandemic. *JAMA* **323**, 1773–1774 (2020).
10. Hay, J. A., Haw, D. J., Hanage, W. P., Metcalf, C. J. E. & Mina, M. J. Implications of the age profile of the novel coronavirus. Preprint at <https://dash.harvard.edu/handle/1/42639493> (2020).
11. Ferguson, N. M. et al. *Impact of Non-pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand* (MRC Centre for Global Infectious Disease Analysis, 2020).
12. Gamio, L., Cai, W. & Hassan, A. Where the U.S. stands now on coronavirus testing. *The New York Times* (26 March 2020).
13. The COVID Tracking Project. *The Atlantic* <https://covidtracking.com/data> (2020).
14. Adolph, C., Amano, K., Bang-Jensen, B., Fullman, N. & Wilkerson, J. Pandemic politics: timing state-level social distancing responses to COVID-19. Preprint at *medRxiv* <https://doi.org/10.1101/2020.03.30.20046326> (2020).
15. Covid-19-data. *The New York Times* <https://github.com/nytimes/covid-19-data> (2020).
16. Mossong, J. et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, 381–391 (2008).
17. Kissler, S. M. et al. Social distancing strategies for curbing the COVID-19 epidemic. Preprint at *medRxiv* <https://doi.org/10.1101/2020.03.22.20041079> (2020).
18. *2018 American Community Survey 1-Year Estimates, Table DP05* (U.S. Census Bureau, 2018).
19. *2018 AHA Annual Survey Database* (American Hospital Association, 2018).
20. Lurie, N., Saville, M., Hatchett, R. & Halton, J. Developing Covid-19 vaccines at pandemic speed. *N. Engl. J. Med.* **382**, 1969–1973 (2020).
21. Yang, J. et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *Int. J. Infect. Dis.* **94**, 91–95 (2020).
22. Deng, X. et al. Case fatality risk of novel coronavirus diseases 2019 in China. Preprint at *medRxiv* <https://doi.org/10.1101/2020.03.04.20031005> (2020).
23. *2010 U.S. Census, Table P2* (U.S. Census Bureau, 2010).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

Data. We obtained counts of the number of individuals in 10-year age bins for all counties in the United States (we include non-county federally incorporated places in the set of all counties for the purposes of our analyses) from the 2018 American Community Survey, available from the United States Census Bureau¹⁸. We define the set of age categories as $G = \{0-9, 10-19, \dots, 70-79, 80+\}$. We obtained data on hospital location and bed number from the American Hospital Association 2018 annual survey¹⁹. We used the calculated total of all beds for each hospital to represent the number of hospital beds, and the number of adult medical/surgical intensive care beds to represent the number of ICU beds. We aggregated hospital and ICU bed data by county in accordance with American Hospital Association data use policy. We obtained the numbers of individuals in each county living in rural and urban areas from the 2010 US census²³.

Mechanistic models. We developed an age-stratified mechanistic epidemiological model based on that of Davies et al.³ that follows a SEIR framework. This model assumes no births or deaths. The subscript i denotes the index of the age strata. The parameter r_i denotes the rate of symptomatic infection for age class G_i . The parameter u_i denotes the relative susceptibility of age class G_i . We set values for r_i and u_i according to the means of the consensus estimates from Davies et al.³:

$$r = \{0.40, 0.25, 0.37, 0.42, 0.51, 0.59, 0.2, 0.76, 0.76\}$$

$$u = \{0.33, 0.37, 0.69, 0.81, 0.74, 0.8, 0.89, 0.77, 0.77\}$$

The infected class is decomposed into asymptomatic (I_A), symptomatic, pre-clinical (I_P) and symptomatic, clinical (I_C) classes to reflect relevant aspects of SARS-CoV-2 epidemiology, namely that not all infected individuals show symptoms and that individuals are frequently quarantined upon presenting symptoms. We also decomposed the recovered class into separate compartments for those recovered from symptomatic infection, R_S and those recovered from asymptomatic infection, R_A , to simplify calculations of total symptomatic and asymptomatic cases. This model framework allows us to impose assumptions about the infectivity of asymptomatic and fully symptomatic individuals (b_A and b_C , respectively) relative to the infected class probably responsible for the bulk of transmission (I_P).

$$\frac{dS_i}{dt} = -S_i u_i \beta \sum_{j=1}^9 C_{ij} \frac{I_{P_j} + b_C I_{C_j} + b_A I_{A_j}}{N_j}$$

$$\frac{dE_i}{dt} = S_i u_i \beta \sum_{j=1}^9 C_{ij} \frac{I_{P_j} + b_C I_{C_j} + b_A I_{A_j}}{N_j} - \delta_E E_i$$

$$\frac{dI_{P_i}}{dt} = r_i \delta_E E_i - \delta_P I_{P_i}$$

$$\frac{dI_{C_i}}{dt} = \delta_P I_{P_i} - \delta_C I_{C_i}$$

$$\frac{dI_{A_i}}{dt} = (1 - r_i) \delta_E E_i - \delta_A I_{A_i}$$

$$\frac{dR_S}{dt} = \delta_C I_{C_i}$$

$$\frac{dR_A}{dt} = \delta_A I_{A_i}$$

Here, C is the contact matrix whose entries C_{ij} correspond to the mean number of contacts between individuals in the i th and j th age classes of G , δ parameters determine the mean amount of time (t) that individuals spend in each class and β is the transmission parameter.

We used this model to simulate a wide range of plausible epidemiological scenarios. Specifically, we considered values for b_C in $\{0.1, 0.5, 1\}$, values for R_0 in $\{2, 4, 6\}$ and values for the degree of homogeneous mixing in $\{0, 0.5, 1\}$. In the sections below, we describe how we constructed the contact matrix C . We set the values of the following model parameters according to published estimates³: $b_A = 0.5$, $\delta_E = \frac{1}{3}$, $\delta_P = \frac{1}{2.1}$, $\delta_C = \frac{1}{2.9}$, $\delta_A = \frac{1}{5}$. After constructing C and fixing these variables, we used numerical methods combined with the next-generation matrix approach²⁴ to calculate the value for β that corresponds to the value R_0 we wished to assume for each scenario.

Rescaling the contact matrix. We used the ‘socialmixr’²⁵ R package to retrieve the UK contact matrix from the POLYMOD study¹⁶, with contacts binned according to the following age categories: $\{0-9, 10-19, \dots, 60-69, 70+\}$. We term this matrix A . No finer resolution was available for contacts involving individuals over the age

of 70. However, to account for differences between individuals in the age classes 70–79 and 80+ in terms of relevant COVID-19 parameters, we synthesized a new matrix, B , that includes contacts for individuals in the age classes 70–79 and 80+:

$$B_{i,70-79} = A_{i,70+} \frac{N_{70-79}}{N_{70+}}$$

$$B_{i,80+} = A_{i,70+} \frac{N_{80+}}{N_{70+}}$$

$$B_{70-79,j} = A_{70+j}$$

$$B_{80+j} = A_{70+j}$$

where N_x is the number of individuals in the entire United States in age class x .

Next, we constructed the contact matrix used in our model C by rescaling B to reflect our assumptions about mixing patterns:

$$C_{i,j} = \frac{(1 - \theta) \sum_{j=1}^9 B_{i,j}}{9} + \theta B_{i,j}$$

The quantity θ represents the degree of homogeneous mixing. When $\theta = 1$, contact patterns are identical to the POLYMOD contact patterns. When $\theta = 0$, contact rates are homogenous across age classes. Values of θ between 0 and 1 correspond to mixing patterns intermediate between the POLYMOD and homogenous scenarios. This rescaling procedure preserves the total number of contacts experienced by each age class while changing the identity of those contacts.

Model simulation. For each scenario in each county, we used the following conditions to initiate the model:

$$S_i = N_i - 4$$

$$E_i = 1$$

$$I_{P_i} = 1$$

$$I_{C_i} = 1$$

$$I_{A_i} = 1$$

$$R_{S_i} = 0$$

$$R_{A_i} = 0$$

The number of individuals within each age class for the county of interest is N_i .

We then simulated the model in R using the ‘ode’ function in the ‘deSolve’ package²⁶ with the ‘lsoda’ integrator and a step size of 0.25. We truncated the simulation when $\sum_{i=1}^9 \frac{I_{P_i} + I_{C_i} + I_{A_i} + R_{S_i} + R_{A_i}}{S_i + E_i + I_{P_i} + I_{C_i} + I_{A_i} + R_{S_i} + R_{A_i}} = 0.2$

and then extracted the number of individuals in each age-stratified compartment.

Case estimation. We calculated the total number of symptomatic infections in each age class by the time that the cumulative infection rate reached 20% as $I_{P_i} + I_{C_i} + R_{S_i}$ at the end of the simulation. We then calculated the number of hospitalizations in each age class by multiplying the number of symptomatic infections in each age class by age-stratified estimates¹¹ of hospitalization rates for symptomatic cases: $\{0.001, 0.003, 0.012, 0.032, 0.049, 0.102, 0.166, 0.243, 0.273\}$.

We then calculated the number of ICU admissions in each age class by multiplying the number of hospitalizations by age-stratified estimates¹¹ of the rate of ICU admissions for patients given hospitalization: $\{0.05, 0.05, 0.05, 0.063, 0.122, 0.274, 0.432, 0.709\}$.

Case distribution. We distributed cases originating in a given county to the healthcare systems of that county and other counties using the following algorithm.

- Let the county of origin be denoted as c_0 and the potential destination counties as c_1, \dots, c_N
- Let the distances between the center of population of the county of c_0 and each potential destination county c_i be $d_{0,i}$. We obtained the latitude and longitude of the center of population for each county from publicly available data from the 2010 US census, and calculated pairwise distances between counties using the R package ‘geosphere’²⁷.
- We next removed all destination counties with $d_{0,i} > 400$ km.

- We calculated a distance weight, y_i , for each remaining potential destination county as $y_i = \frac{1}{20} e^{\frac{d_{i,1}}{20}}$
- We calculated a bed weight, z_j , for each county as the number of total hospital beds in c_j . For projections involving ICU admissions, we used the number of ICU beds rather than the number of hospital beds.
- We then calculated a composite weight, w_i , for each county as
$$w_i = \frac{y_i}{\sum_{j=0}^9 y_j} \frac{z_j}{\sum_{j=0}^9 z_j}$$
- Lastly, cases originating in c_0 were then distributed to counties c_0, \dots, c_N proportional to $\frac{w_0}{\sum_{i=0}^9 w_i}, \dots, \frac{w_N}{\sum_{i=0}^9 w_i}$

'Alternate optimistic' and 'alternate pessimistic' scenarios. For the two scenarios in which we varied R_0 between counties according to the percentage of the population residing in rural areas, the value of R_0 for each county was calculated as:

$$R_0 = 2 + (R_{0\max} - 2) \times \text{percentage population residing in urban area}$$

Instead of truncating our simulations at a 20% cumulative infection rate, we truncated our simulations when the following condition was met, indicating that the cumulative infections rate was equal to 20% of the herd immunity threshold:

$$\sum_{i=1}^9 \frac{I_{P_i} + I_{C_i} + I_{A_i} + R_{S_i} + R_{A_i}}{S_i + E_i + I_{P_i} + I_{C_i} + I_{A_i} + R_{S_i} + R_{A_i}} = 0.2 \times \left(1 - \frac{1}{R_0}\right)$$

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data underlying the maps of disease burden are available from the corresponding author upon request. The data used in our analyses, with the exception of hospital capacity data from the American Hospital Association Annual Survey, which are not publicly available, are included as Supplementary Information.

Code availability

The code used to conduct analyses and generate figures is included as Supplementary Information, and is available at <https://github.com/ianfmiller/covid19-burden-mapping>.

References

24. 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).
25. Funk, S. *socialmixr: Social Mixing Matrices for Infectious Disease Modeling* <https://cran.r-project.org/web/packages/socialmixr/vignettes/introduction.html> (2020).
26. Soetaert, K., Petzoldt, T. & Setzer, R. W. Package deSolve: solving initial value differential equations in R. *J. Stat. Softw.* **33**, 1–25 (2010).
27. Hijmans, R. J. *geosphere: Spherical Trigonometry* <https://cran.r-project.org/web/packages/geosphere/index.html> (2019).

Acknowledgements

We thank N. Davies and R. Eggo for providing data. I.F.M. and A.D.B. acknowledge funding from National Science Foundation Graduate Research Fellowships.

Author contributions

I.F.M., A.D.B., C.J.E.M. and B.T.G. conceived the study. I.F.M. and A.D.B. conducted the analyses. All authors contributed to writing and revision of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41591-020-0952-y>.

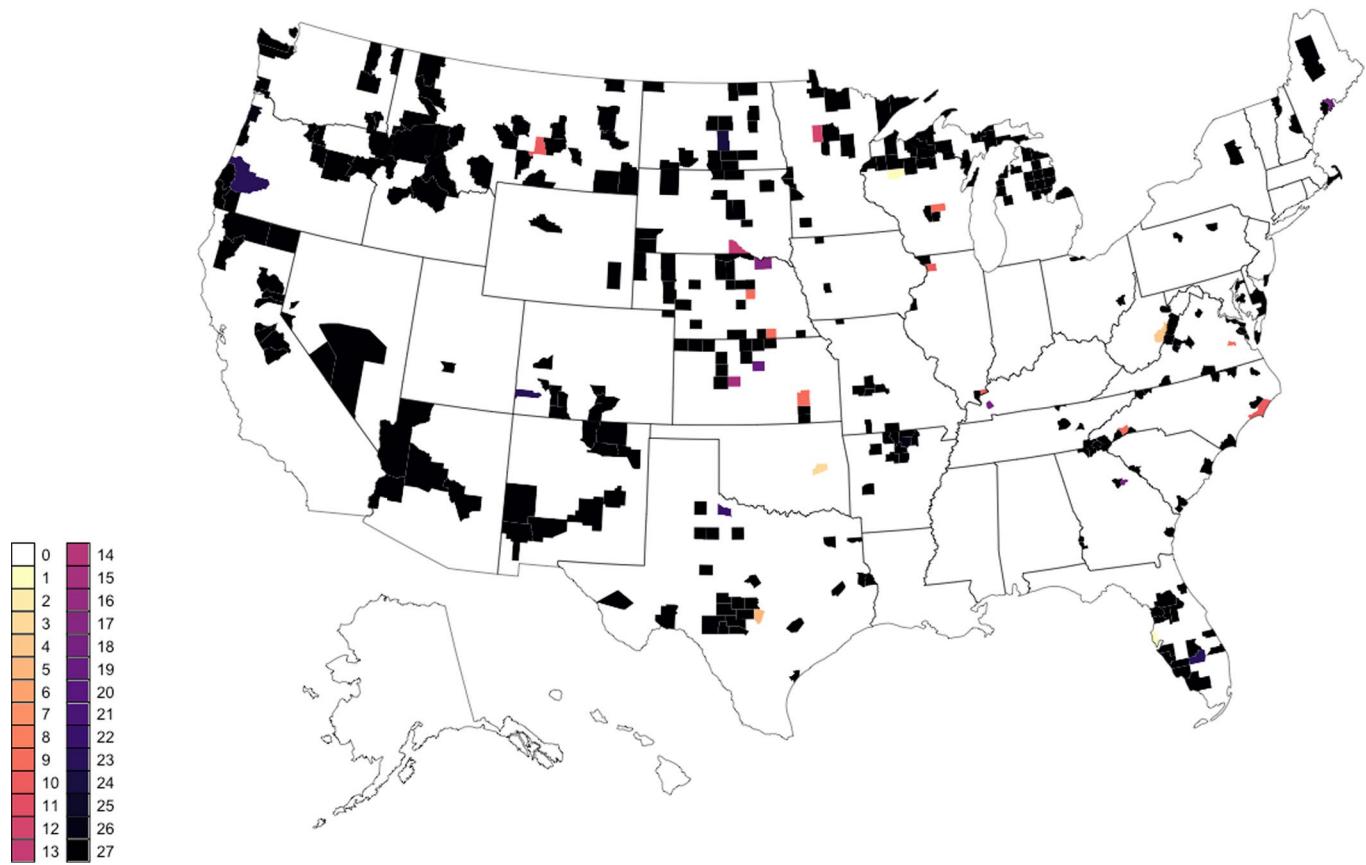
Supplementary information is available for this paper at <https://doi.org/10.1038/s41591-020-0952-y>.

Correspondence and requests for materials should be addressed to I.F.M.

Peer review information Jennifer Sargent was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

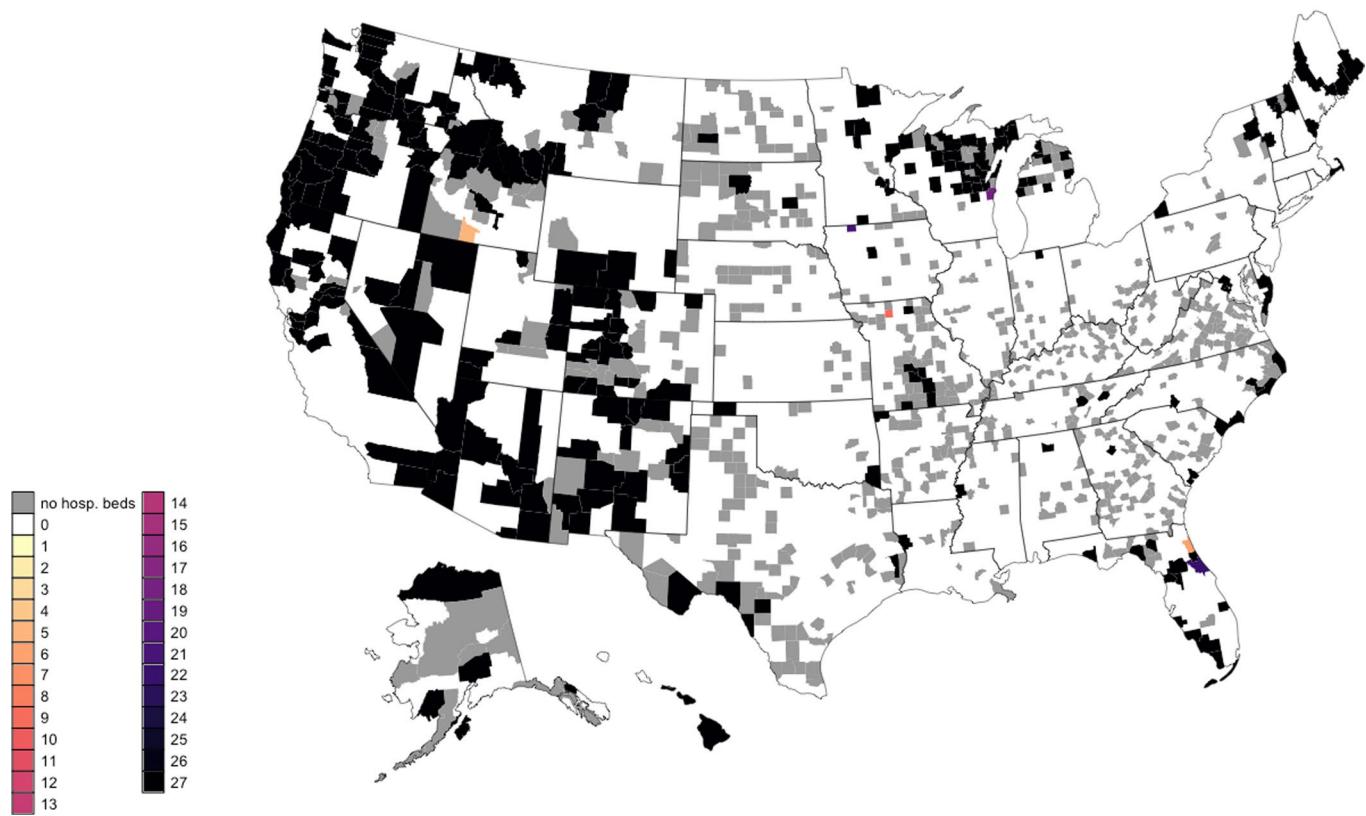
Reprints and permissions information is available at www.nature.com/reprints.

N transmission scenarios identifying county as being at or above 90% quantile of per capita hospitalizations



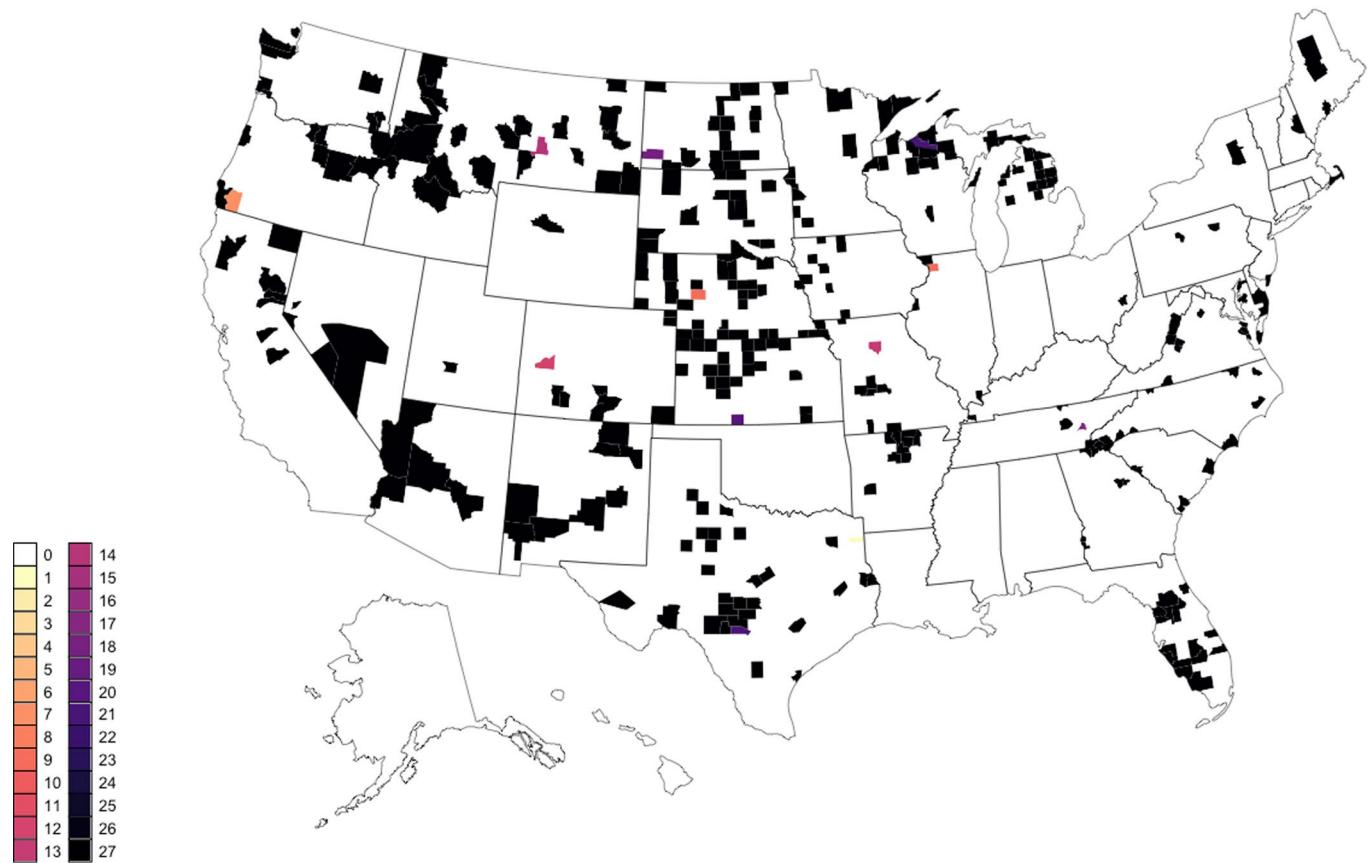
Extended Data Fig. 1 | The rural-urban distribution of U.S. counties with heavy COVID-19 burden. a, Per capita hospitalizations. **b**, Cumulative hospitalizations per hospital bed. **c**, per capita ICU admissions. **d**, Cumulative hospitalizations per ICU bed. Counties identified as being in the 90% quantile for various metrics of disease burden in the optimistic and pessimistic transmission scenarios (purple bars) and comparable counties across the entire US (grey bars).

N transmission scenarios identifying county as being at or above 90% quantile of hospitalizations per hospital bed



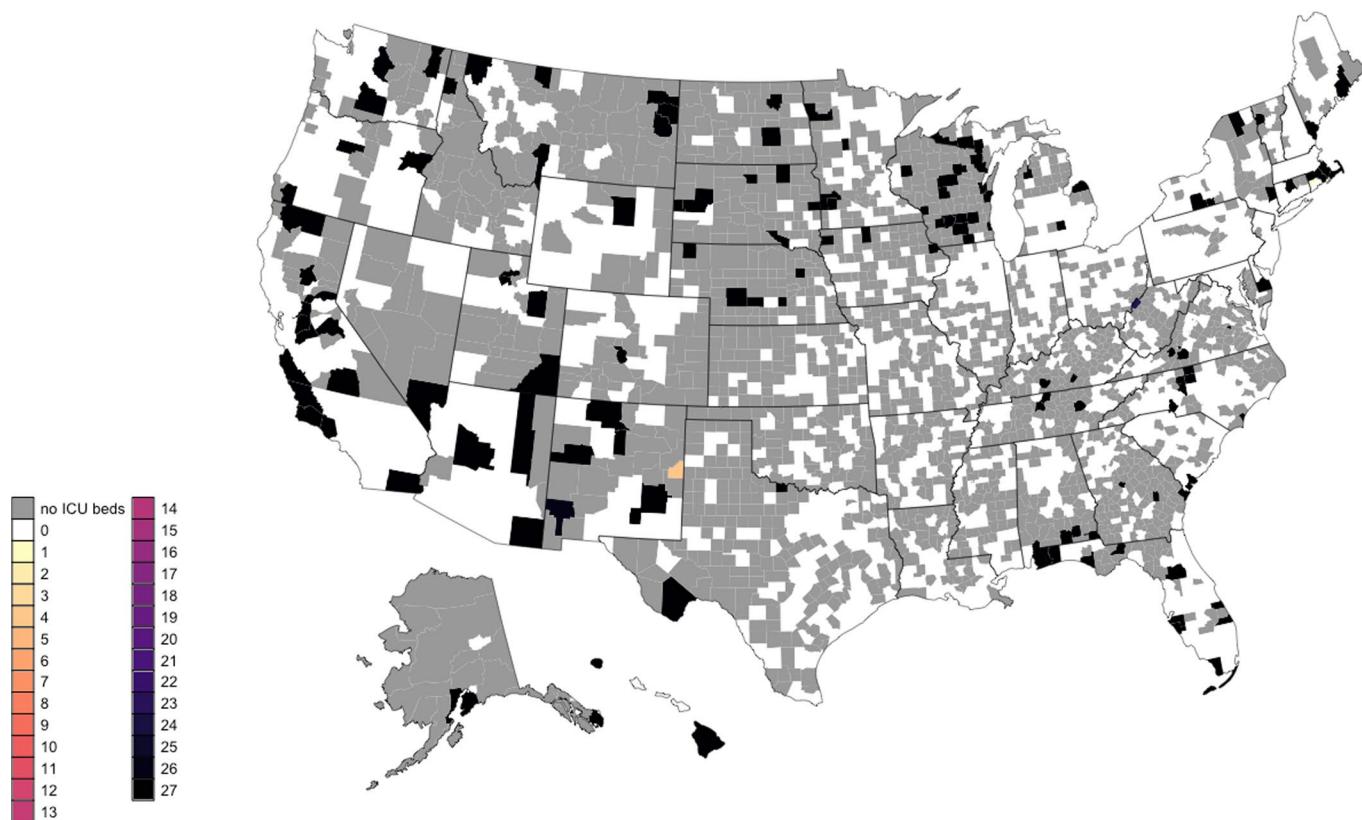
Extended Data Fig. 2 | Counties at or above 90% quantile of per capita hospitalizations. Colors indicate the number of transmission scenarios in which each county was identified as being at or above the 90% quantile of per capita hospitalizations.

N transmission scenarios identifying county as being at or above 90% quantile of per capita ICU admissions

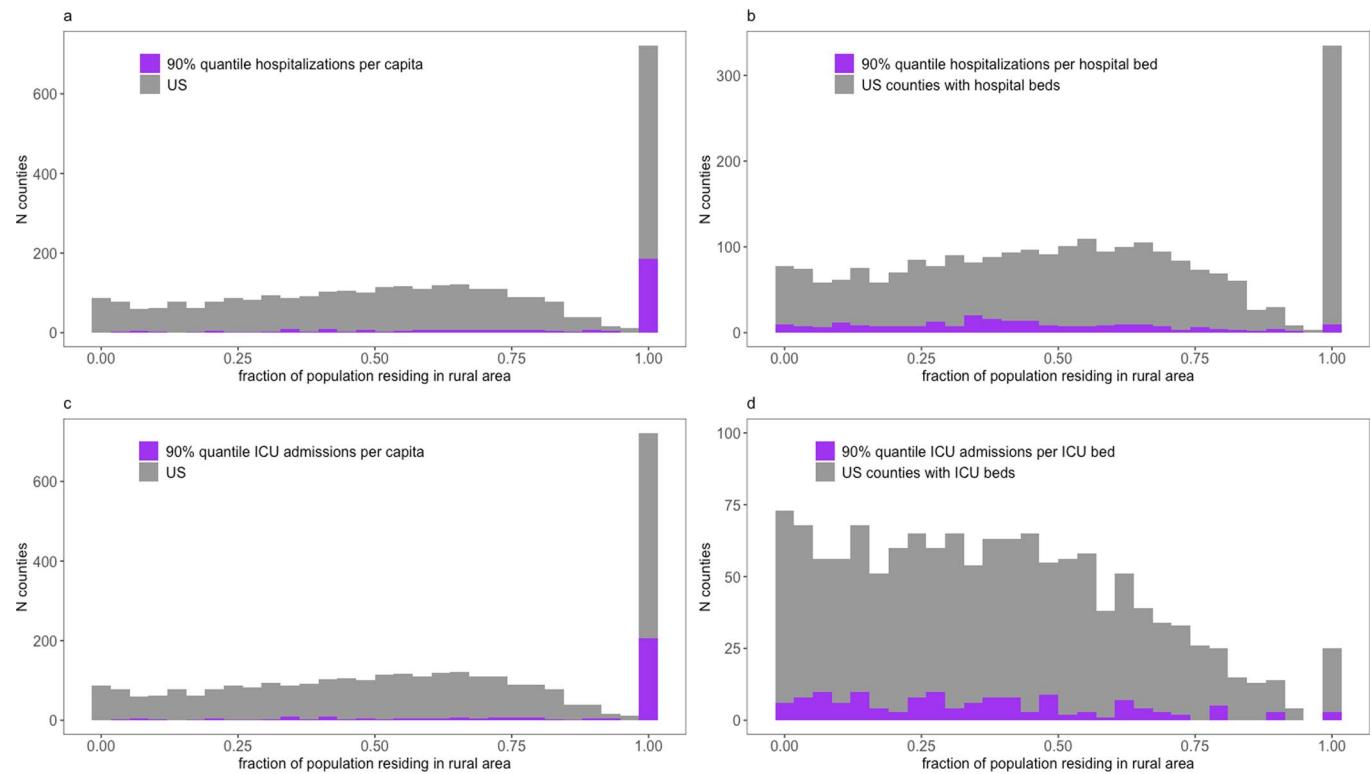


Extended Data Fig. 3 | Counties at or above 90% quantile of hospitalizations per hospital bed. Colors indicate the number of transmission scenarios in which each county was identified as being at or above the 90% quantile of hospitalization per hospital bed. Counties without hospital beds are colored grey.

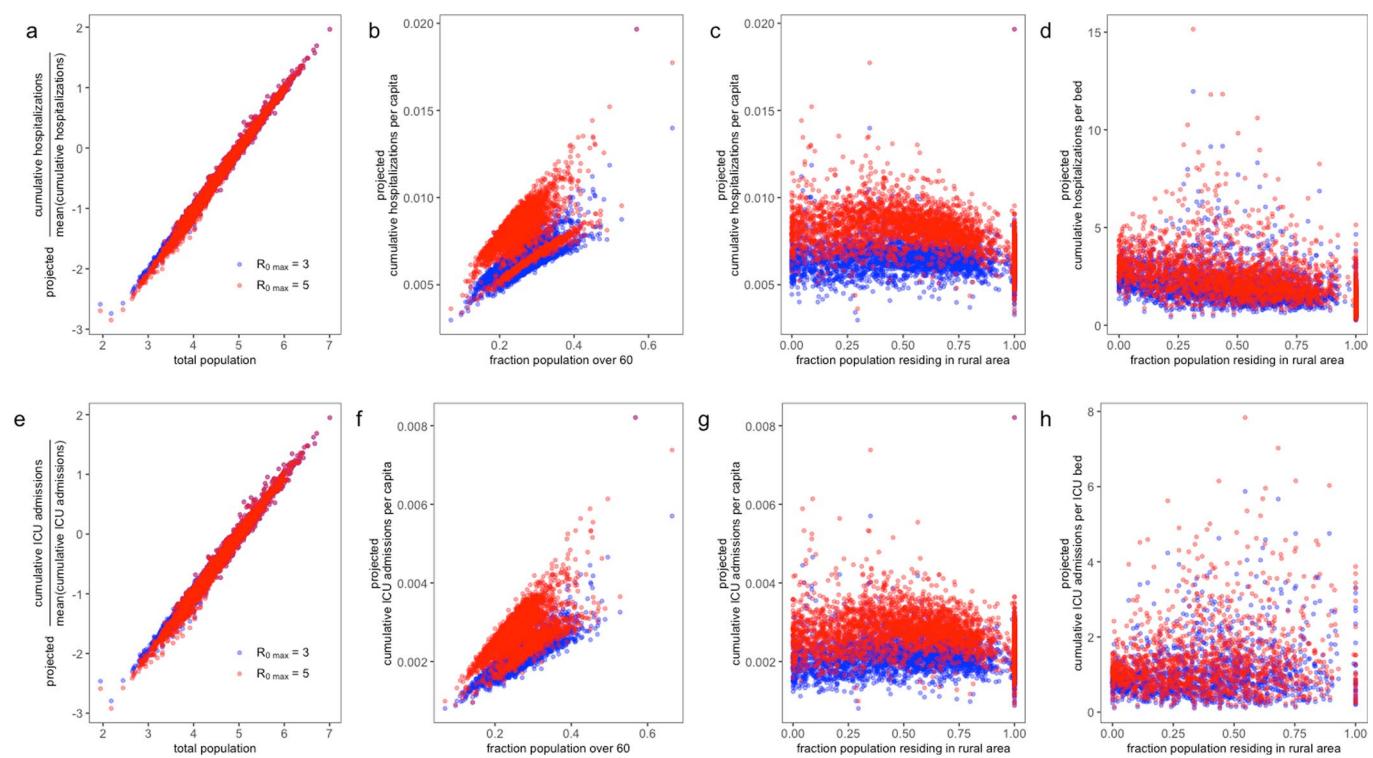
N transmission scenarios identifying county as being at or above 90% quantile of ICU admissions per ICU bed



Extended Data Fig. 4 | Counties at or above 90% quantile of per capita ICU admissions. Colors indicate the number of transmission scenarios in which each county was identified as being at or above the 90% quantile of ICU admissions per capita.

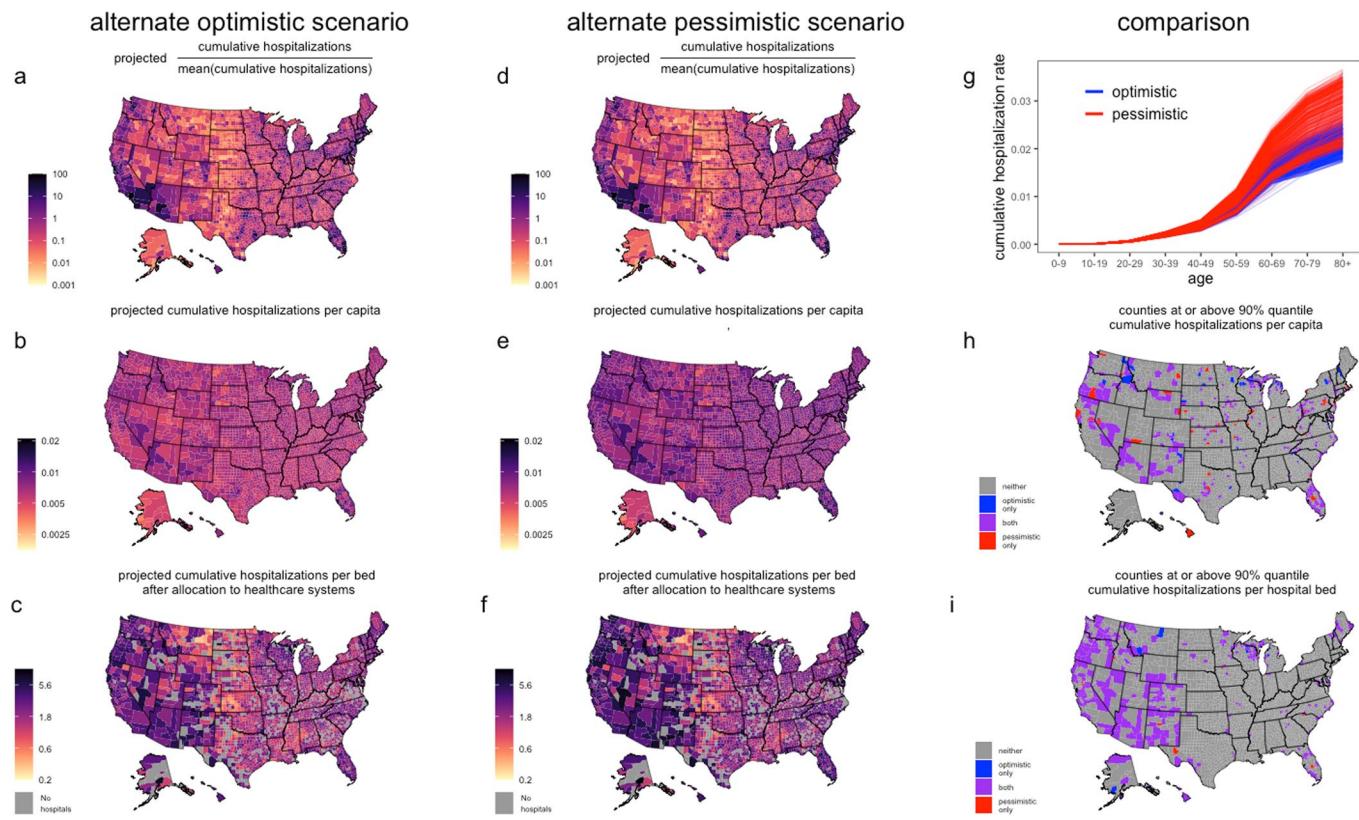


Extended Data Fig. 5 | Counties at or above 90% quantile of per capita ICU admissions per ICU bed. Colors indicate the number of transmission scenarios in which each county was identified as being at or above the 90% quantile of ICU admissions per ICU bed. Counties without ICU beds are colored grey.

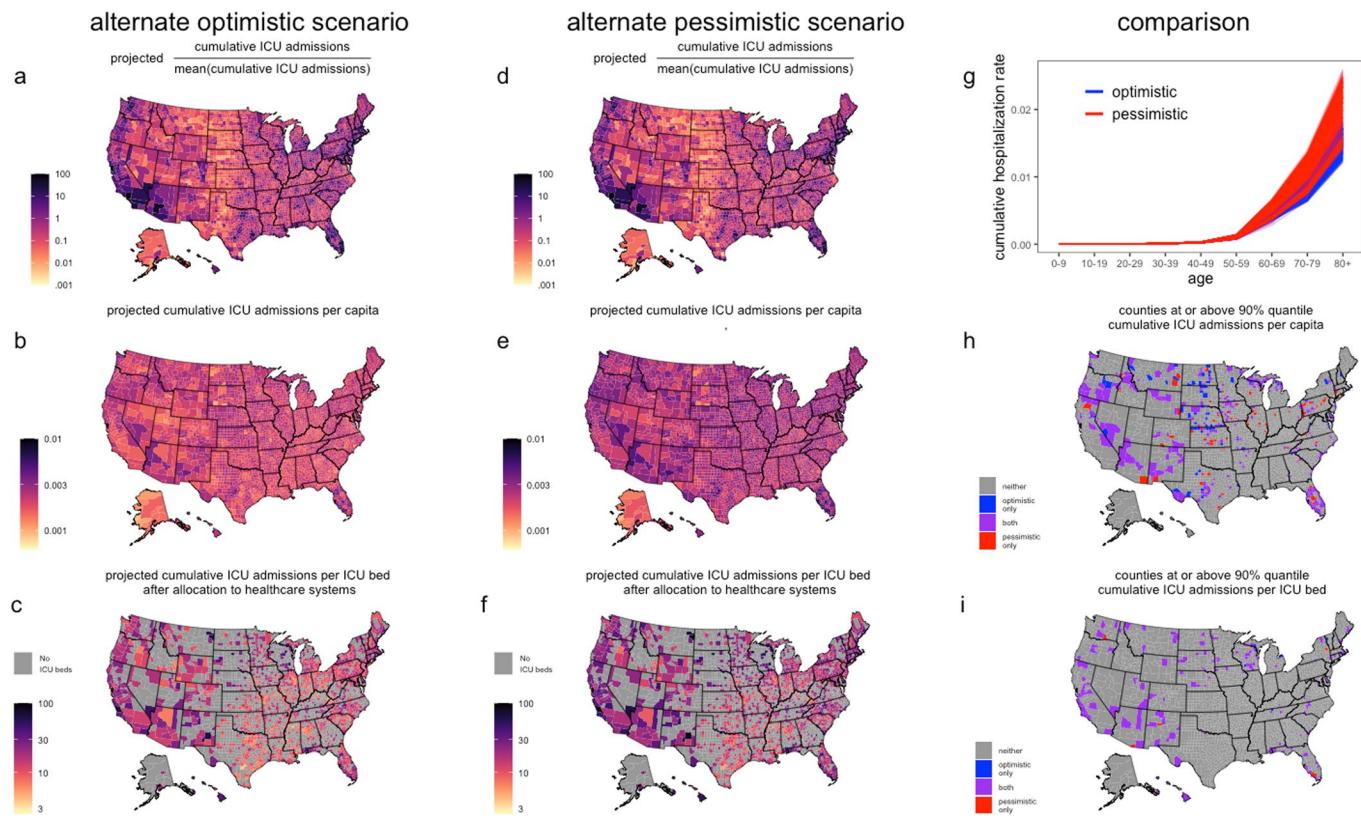


Extended Data Fig. 6 | Relationships between population characteristics and disease burden for alternative optimistic and pessimistic scenarios.

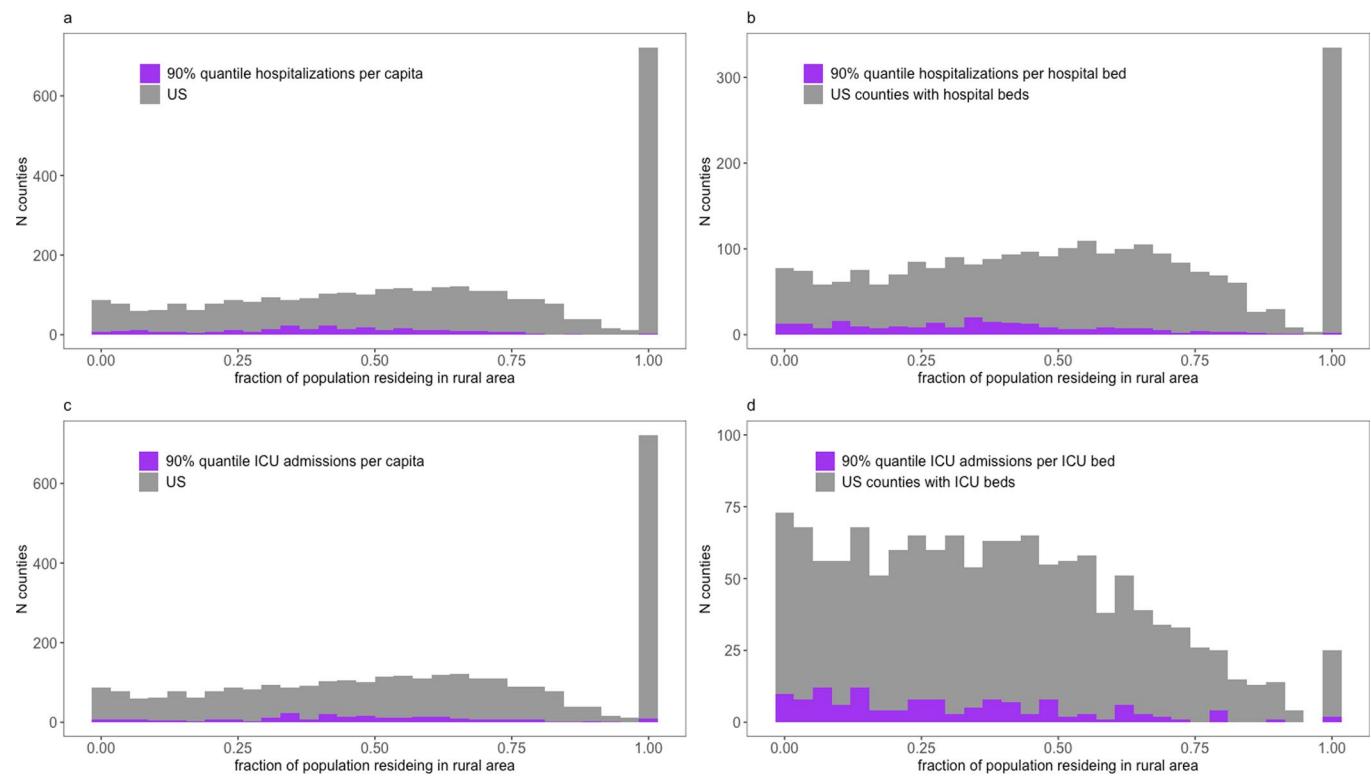
A-H Relationship between population characteristics (x-axes) and metrics of disease burden (y-axes) for the alternative optimistic (blue points) and pessimistic (red points) scenarios in which RO increased as a linear function of urban population.



Extended Data Fig. 7 | Projected cumulative burden of hospitalizations in the U.S. for alternative optimistic and pessimistic scenarios. In the alternative optimistic and pessimistic scenarios, R_0 increased as a linear function of urban population. **(a-c)** optimistic scenario. **(d-f)** pessimistic scenario. **a, d**, Relative number of hospitalizations in each county. **b, e**, Number of projected hospitalizations per capita in each county. In **a, b, d**, and **e** cases have not yet been allocated to healthcare systems. **c, f**, Cumulative number of hospitalizations per hospital bed after cases have been allocated to healthcare systems. **g**, Cumulative fraction of each age class hospitalized in each transmission scenario. Each of the 315 lines for each transmission scenario represents a different county. **h, i**, Counties estimated to be in the 90% quantile of hospitalizations per capita and hospitalizations per hospital bed (after case allocation). Colors in **h, i** indicate whether the counties were estimated to be in the 90% quantile in the optimistic scenario, the pessimistic scenario, both, or neither. A high-resolution version of this figure is provided in the Supplementary Information.



Extended Data Fig. 8 | Projected cumulative burden of ICU admissions in the U.S. for alternative optimistic and pessimistic scenarios. In the alternative optimistic and pessimistic scenarios, R_0 increased as a linear function of urban population. **(a-c)** Optimistic scenario. **(d-f)** Pessimistic scenario. **a, d**, Relative number of ICU admissions in each county. **b, e**, Number of projected ICU admissions per capita in each county. In **a, b, d**, and **e** cases have not yet been allocated to healthcare systems. **c, f**, Cumulative number of ICU admissions per ICU bed after cases have been allocated to healthcare systems. **g**, Cumulative fraction of each age class requiring ICU admission in each transmission scenario. Each of the 315 lines for each transmission scenario represents a different county. **h** and **i**, Counties estimated to be in the 90% quantile of ICU admissions per capita and ICU admissions per ICU bed (after case allocation). Colors in **h** and **i** indicate whether the counties were estimated to be in the 90% quantile in the optimistic scenario, the pessimistic scenario, both, or neither. A high-resolution version of this figure is provided in the Supplementary Information.



Extended Data Fig. 9 | The rural-urban distribution of U.S. counties with heavy COVID-19 burden for alternative optimistic and pessimistic scenarios. In the alternative optimistic and pessimistic scenarios, RO increased as a linear function of urban population. **a**, Per capita hospitalizations. **b**, Cumulative hospitalizations per hospital bed. **c**, per capita ICU admissions. **d**, Cumulative hospitalizations per ICU bed. Counties identified as being in the 90% quantile for various metrics of disease burden in the optimistic and pessimistic transmission scenarios (purple bars) and comparable counties across the entire US (grey bars).

Metric	Median percentage of population residing in rural area		Mann-Whitney U	Two-sided p value
	Counties identified at or above 90% quantile	Comparable counties		
Per capita hospitalizations	41.01 (N = 269)	59.54 (N total counties = 3142)	277982	<2.2*10 ⁻¹⁶
Per capita ICU admissions	45.18 (N = 245)	59.54 (N total counties = 3142)	291914	2.35*10 ⁻¹⁰
Hospitalizations per hospital bed	34.52 (N = 241)	51.82 (N counties with hospital beds = 2478)	201652	<2.2*10 ⁻¹⁶
ICU admissions per ICU bed	27.02 (N = 131)	36.21 (N counties with ICU beds = 1353)	73457	1.2*10 ⁻³

Extended Data Fig. 10 | The comparative rurality of U.S. counties with heavy COVID-19 burden for alternative optimistic and pessimistic scenarios. In the alternative optimistic and pessimistic scenarios, R0 increased as a linear function of urban population. Mann-Whitney U tests indicated that the sets of counties identified as being heavily burdened in both scenarios were less rural than the set of comparable counties.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.

A description of all covariates tested

A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.

For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection No code was used to collect data.

Data analysis Data were analyzed using R version 3.6.1.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data underlying the maps of disease burden are available from the corresponding author upon request. The data used in our analyses, with the exception of hospital capacity data from the American Hospital Association Annual Survey which is not publicly available, are included as Supplementary Information.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We analyzed publicly available data using a mathematical model.
Research sample	We used publicly available data.
Sampling strategy	No data was collected.
Data collection	No data was collected.
Timing and spatial scale	No data was collected.
Data exclusions	No data was collected.
Reproducibility	No data was collected.
Randomization	No data was collected.
Blinding	No data was collected.

Did the study involve field work? Yes No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	Antibodies	<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Eukaryotic cell lines	<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	Palaeontology and archaeology	<input checked="" type="checkbox"/>	MRI-based neuroimaging
<input checked="" type="checkbox"/>	Animals and other organisms		
<input checked="" type="checkbox"/>	Human research participants		
<input checked="" type="checkbox"/>	Clinical data		
<input checked="" type="checkbox"/>	Dual use research of concern		