Check for updates

Real-time, interactive website for US-county-level COVID-19 event risk assessment

Aroon Chande [1,2], Seolha Lee [3], Mallory Harris [4], Quan Nguyen[1], Stephen J. Beckett [1], Troy Hilley[1], Clio Andris [3,5] ✉ and Joshua S. Weitz [1,6] ✉

Large events and gatherings, particularly those taking place indoors, have been linked to multitransmission events that have accelerated the pandemic spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). To provide real-time, geo-localized risk information, we developed an interactive online dashboard that estimates the risk that at least one individual with SARS-CoV-2 is present in gatherings of different sizes in the United States. The website combines documented case reports at the county level with ascertainment bias information obtained via population-wide serological surveys to estimate real-time circulating, per-capita infection rates. These rates are updated daily as a means to visualize the risk associated with gatherings, including county maps and state-level plots. The website provides data-driven information to help individuals and policy makers make prudent decisions (for example, increasing mask-wearing compliance and avoiding larger gatherings) that could help control the spread of SARS-CoV-2, particularly in hard-hit regions.

As of 5 October 2020, there have been over seven million documented cases of SARS-CoV-2 in the United States, leading to more than 200,000 documented fatalities[1]. Despite large-scale efforts to suppress disease spread via lockdown orders and other non-pharmaceutical interventions including mask wearing[2], there has been a resurgence of SARS-CoV-2 cases in the United States since late summer 2020; particularly in the south and west followed by a resurgence of cases in the midwest in early autumn 2020 (refs. [3,4]). The rise of cases threatens public health, economic recovery and the re-opening of K-12 schools as well as colleges and universities[5,6].

The basic reproduction number and large fraction of asymptomatic cases represent challenges for controlling SARS-CoV-2 (ref. [7]). Early estimates of the basic reproduction number of SARS-CoV-2 range from 2.1 to 4.5 (ref. [8]), with current best estimates from the Centers of Disease Control and Prevention (CDC) indicating a basic reproduction number of 2.5 (ref. [9]). Early studies found that approximately half of cases may be via presymptomatic, mild or asymptomatic transmission[10,11]. The absence of commonly associated symptoms (fever, cough, shortness of breath) may be more pronounced in younger individuals. In addition, effective isolation of symptomatic cases may increase the fraction of circulating cases that are mild or asymptomatic.

The strong and often undocumented spread of SARS-CoV-2 is exacerbated by large transmission incidents[12], referred to as 'super-spreading' events[13]. Super-spreading of SARS-CoV-2 has been documented in multiple, indoor events or large gatherings in which a single infector is putatively associated with the infection of dozens (or more) of individuals[14–16]. Large gatherings pose particular challenges for preventing the spread of SARS-CoV-2. First, the risk that one (or more) individuals is infected increases rapidly with group size; increasing the inherent risk of a potential exposure as groups increase in number. Second, the number of potential interactions increases with gathering size (up to the square of the

number of individuals in small groups where all individuals might be in contact). Third, follow-up contact tracing is problematic given the potential unknown nature of identifying close interactions. Although the last two challenges can be hard to quantify due to logistical and privacy reasons, this first category of risk is quantifiable, presents a gateway to action taking and should be communicated to the public at large.

In March 2020, one of us (J.S.W.) developed a scenario-driven approach to assess the risk that one (or more) individuals in a group was infected in groups of size 10, 100, 1,000, 10,000 and even 100,000 (ref. [17]). The risk chart highlighted combinations of event size n and circulating cases in the United States that had equal risk. The visualized risk contours can be defined via a binomial statistical model as a set of values (p,n) such that risk is a constant $r = 1 - (1-p)^n$ (the implications of these risk contours at the early stages of the SARS-CoV-2 outbreak in the United States are discussed in refs. [18–20]). Given a risk level r defined between 0 and 1, the per-capita probability along an equirisk contour scales as $1/n$ (converging rapidly to 0 when n is large). Hence, large events can potentially seed transmission even when the per-capita probability that an individual is infected remains low.

With shelter-in-place orders now suspended in most of the United States, many businesses (from retail to sports), recreational facilities, daycare centres, schools (both K-12 and colleges/universities) are evaluating re-opening plans. These plans must also gauge the probable risk of transmission. The COVID-19 Real-Time Event Risk Assessment website (https://covid19risk.biosci.gatech.edu/) uses a data-driven approach to connect circulating case reports with risk assessment by adapting a binomial model of risk to real-time estimates at the county level. The central purpose is to quantify and visualize the expected risk associated with gatherings of different sizes and to help guide action taking by policy makers and public health departments, as well as event planners and visitors. The interactive website has drawn over two million visitors since the launch

[1]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA. [2]Applied Bioinformatics Laboratory, Atlanta, GA, USA. [3]School of City and Regional Planning, Georgia Institute of Technology, Atlanta, GA, USA. [4]Department of Biology, Stanford University, Palo Alto, CA, USA. [5]School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, USA. [6]School of Physics, Georgia Institute of Technology, Atlanta, GA, USA. ✉e-mail: clio@gatech.edu; jsweitz@gatech.edu
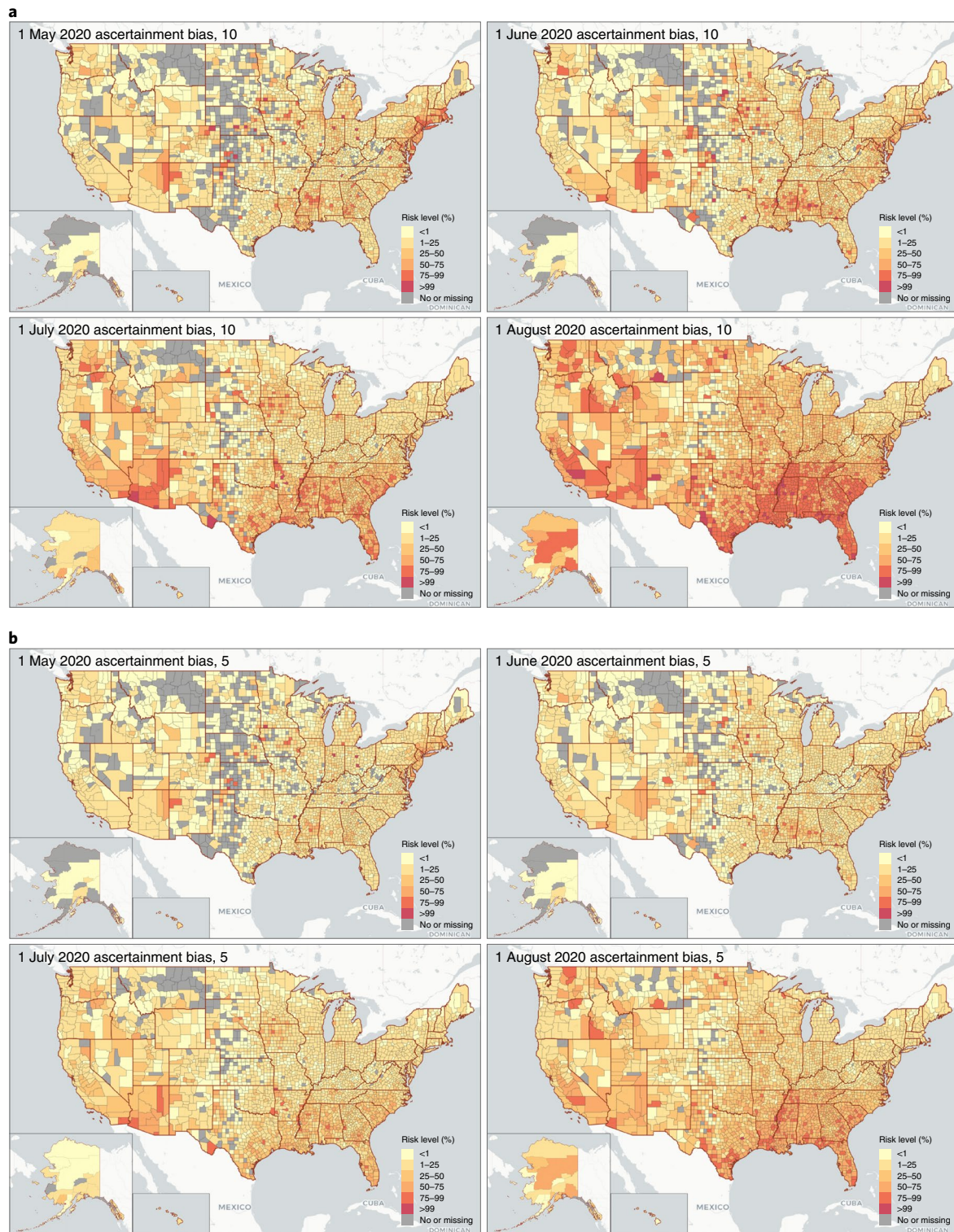
Fig. 1 | Heterogeneous risk map. a,b, The map depicts risk given events of size 50 using ascertainment biases of ten times (**a**) and five times (**b**) on 1 May, 1 June, 1 July and 1 August 2020. Alaska and Hawaii were resized to be smaller than they actually are on the web.

of the county-level map on 7 July 2020, is updated daily and continues to provide real-time estimates of event-associated risk with extensions to risk assessment in select countries globally (released on 5 October 2020). As we describe below, the visualized risk maps are intended to inform individuals on the need to take preventative steps to reduce new transmission, for example, by avoiding large gatherings and wearing masks when in close contact with others.

Results

Real-time risk is heterogeneous, reflecting recent increases in cases. We used a binomial probability model to assess the risk that one or more individuals is infected with SARS-CoV-2 (see probability model, Methods). The risk that one or more individuals is infected at an event is equivalent to one minus the probability that no individuals are infected given a per-capita infection risk
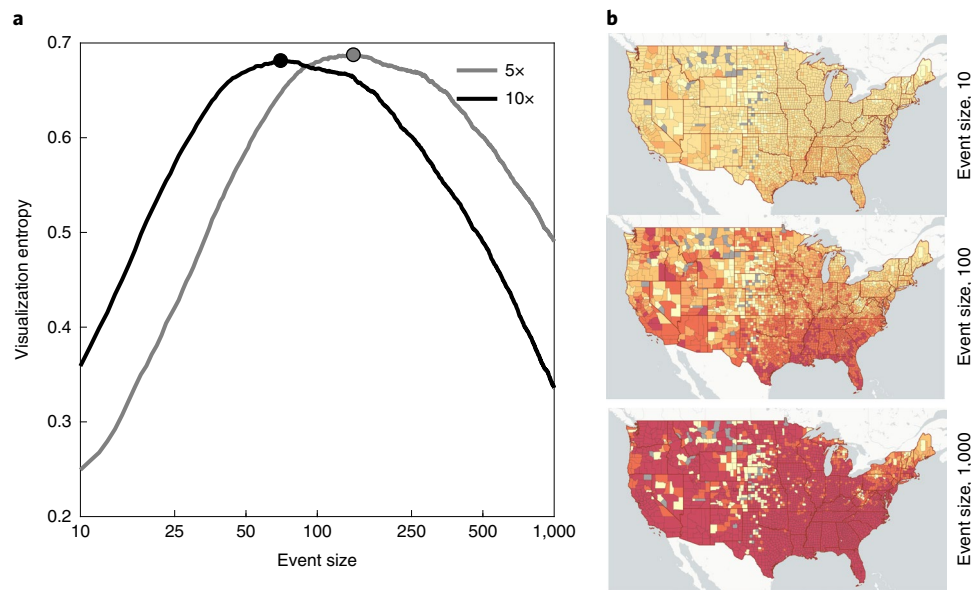
Fig. 2 | Visualizations of event-associated risk. An information-based index of heterogeneity in risk reveals that intermediate event sizes differentiate spatially heterogeneous risk as of 1 August 2020. **a**, Visual map information as a function of event size using five and ten times ascertainment biases for event sizes between 10 and 1,000 people. **b**, Maps illustrating that most counties appear to have similarly low risk when events are small (fewer than ten individuals) or similarly high risk when events large (>1,000 individuals). In contrast, the highest level of heterogeneity in risk is revealed given intermediate event sizes (50–150 individuals). Map visualizations use an assumption of five-times ascertainment bias.

estimated using recent case reports multiplied by an ascertainment bias informed by serological surveys. To assess risk variation, we measured county-level heterogeneity by combining time-varying estimates given reported case counts from May to August at the county level. Figure 1 shows four snapshots, spaced monthly, corresponding to estimated risk associated with gatherings of 50 individuals on 1 May, 1 June, 1 July and 1 August 2020. These snapshots reveal that gathering-associated risk was heterogeneous and concentrated in the northeast (and to some extent in the southwest) in early May with higher risk associated with the south and southeast beginning in early June. Critically, the regional shift in current risk means that use of cumulative case or death counts does not necessarily provide near-term actionable information on ongoing risk. We note that estimates are affected by uncertainty in the ascertainment bias; the default option is ten times corresponding to the median of serologically positives to PCR positives in locale-centred population surveys conducted in April to May 2020[21] (Methods). In light of increased testing, we also include a five-times ascertainment bias (see Discussion for more details).

Information can be conveyed by focusing on risk associated with intermediate-scale events. One key choice in visually displaying risk is selecting event sizes that are meaningful in a public health context, can be precomputed and effectively communicate differential risk. Precomputation is key to accommodate a large number of users simultaneously. The choice of gathering size strongly influences the information content of county-level maps. The map includes six different coloured bins representing the probability of an infected individual being present at an event: <1, 1–25, 25–50, 50–75, 75–99 and >99%. We note estimation of risk as calculated via the binomial model saturates at 1 when the size of the gathering n is much larger than $1/p$. For example, if $p = 0.005$ or 1 in 200, then events much larger than 200 will saturate near 1; in contrast, if $p = 0.0001$ or 1 in 10,000, then events much larger than 10,000 will saturate near 1. As a result, the map will be uniformly 'light' (associated with low risk) when events are sufficiently small and uniformly 'dark' (associated with high risk) when events are sufficiently large.

This also suggests that displaying risk associated with intermediate size events will more effectively communicate differences between counties and states.

We used an information-based metric to assess the overall spatial heterogeneity of the county-level risk map. We denote the visual map information as the sum of $-q_i \log(q_i)$ where q_i denotes the fraction of counties in the ith risk category where $i = 1$ to 6 (per the number of data bins on the map). Note that small counties are weighted equally to large counties, and future work could use a population-weighted cartogram to allow users to visualize county areas in proportion to their respective populations. Figure 2 quantifies the information conveyed associated with visualizations across sizes from 10 to 1,000 on 1 August 2020. The peak information is found at sizes of $n = 70$ and 142 for ascertainment biases ten and five times, respectively, consistent with the maximum colour divergence at intermediate risk sizes. This peak indicates that in early August, whereas most small events (of about ten) had relatively low risk everywhere and most large events (greater than 1,000 people) had a relatively high risk everywhere, the risk associated with intermediate-sized events was strongly variable with region. Such variability is critical to informing opening decisions.

State-level variation in critical event sizes. The spatiotemporal variation in risk can be viewed a different way: by evaluating the location-dependent risk associated with a given event size. To do so, we fixed the event size at 50 people and then estimated the state-level risk $(1 - (1-p)^{50})$. Figure 3 arranges states as well as Washington DC and Puerto Rico in order of their relative risk effective 15 August 2020 from no. 1 (lowest state-level risk) to no. 52 (highest state-level risk). In many cases, states with high risk levels in May and June experienced declines throughout July and August, particularly in the northeast. In contrast, states with lower risk levels in May and June experienced upsurges of cases (and risk) in July and August, especially in the south. This analysis further reinforces the spatiotemporal variation of event risk, as many states continue to have elevated risk associated with gatherings of 50 (corresponding to a social gathering, bar, restaurant, business event or
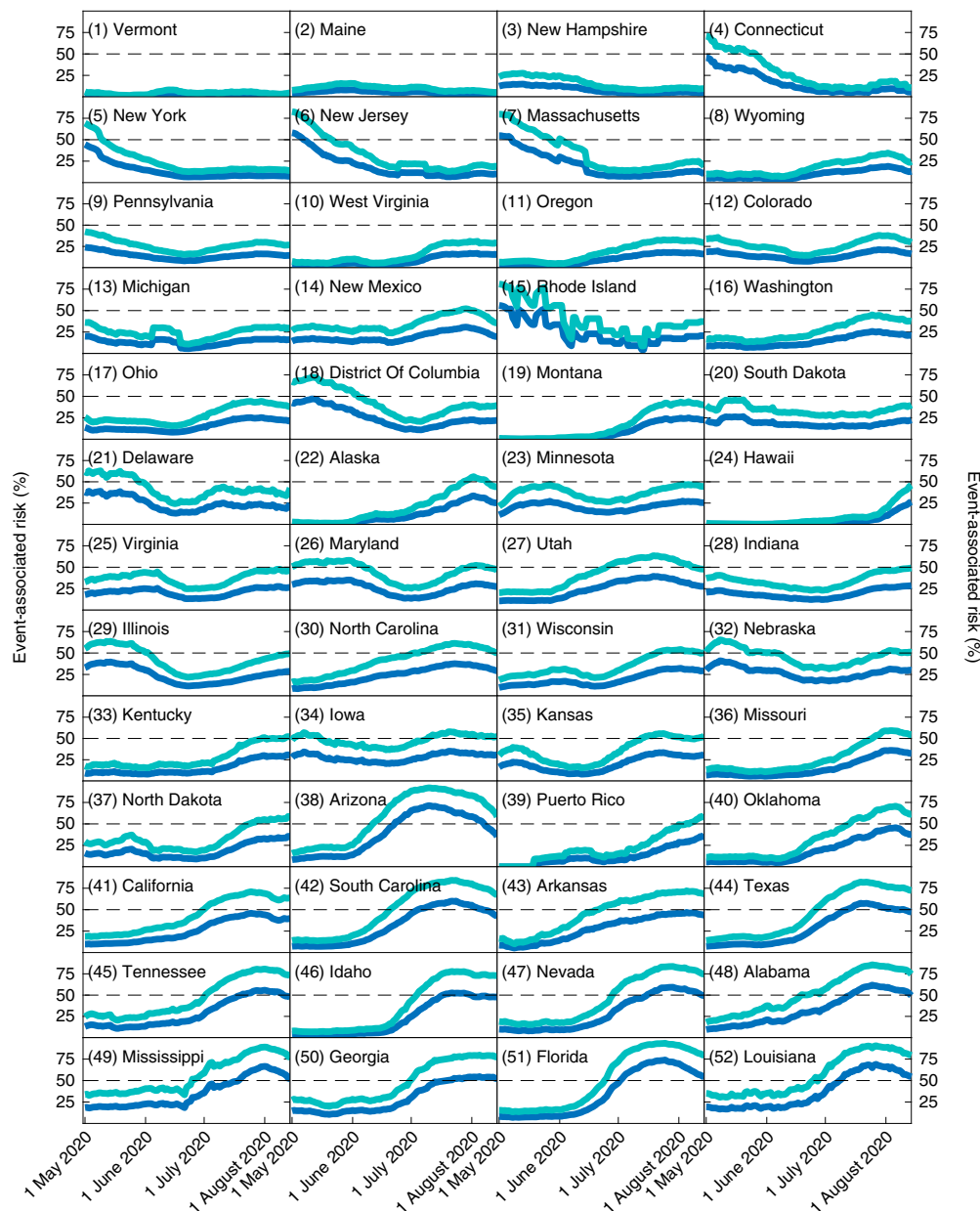
Fig. 3 | State-level risk associated with events of size 50 over time. The curves denote risk estimates assuming 5:1 (dark blue) and 10:1 (light blue) ascertainment biases. States are ordered as a function of ascending risk level as of 14 August 2020 (last point shown).

approximately two K-12 classes). Specifically, of the 52 locales, we identify 51, 49, 44 and 22 that have more than 5, 10, 25 and 50% risk, respectively, that one or more individuals with SARS-CoV-2 are present in events of size 50 effective 15 August 2020 assuming ten times ascertainment bias (and 49, 46, 31 and four locales assuming five-times ascertainment bias). This finding indicates that plans to reopen schools, colleges and businesses should operate knowing that there is an elevated risk of within-event transmission if precautions are not taken; this elevated risk is robust to the choice of either ten or five-times ascertainment bias.

Discussion
The COVID-19 Event Risk Assessment Tool provides real-time, localized information on risk associated with gatherings. The risk highlights the probability that one (or more) individuals may be infected with SARS-CoV-2 in events of different sizes. By integrating real-time information aggregated via state health

departments nationwide along with a simple statistical model, the website is able to capture, calculate and disseminate information relevant to decision-making by the public that could help reduce risk and new transmission. The risk model addresses the probability that an infected person is present at events of different sizes rather than estimating the likelihood that someone will become infected at that event. Addressing the latter would involve analyses beyond the scope of this paper including environmental models and behaviour assessment (for example, the need for mask wearing and physical distancing[22,23]).

Static and interactive maps and as well as interactive data dashboards, i.e., sets of linked visualizations for data exploration[24], have proliferated since the start of SARS-CoV-2. Most dashboards allow visitors to choose epidemic-associated variables to display: number of cases, cases per capita (for example, per 100,000 people), number of deaths and deaths per capita, for divisions within a single country or for countries on a global map. Other behavioural maps have

illustrated the reduction in mobility[25] or polling results such as attitudes towards masks[26]. Like these maps, the COVID-19 Event Risk Assessment Tool describes the relationship between disease spread and behaviour; albeit in an effort to change rather than track behaviour. This map is designed as a spatial decision support system[27] that allows individuals to measure the risk of their own actions and plan accordingly. It removes the burden of interpreting what case rates mean in a quantitative context by directly communicating a probability of encountering an infected individual via interactions. As a result, individuals can visualize themselves in a group and decide whether this risk is worth taking. Risk assessment and tolerance varies considerably between individuals. The same risk value from the tool (for example, 50% risk) will differentially affect an individual's decision whether to attend an event or hold an event, and/or shape their perceptions of events. The intention of the tool is to promote informed behaviour by providing a quantity analogous to other likelihoods that may be familiar to users (for example, weather forecasts). Follow-up work is necessary to characterize how behaviour changed on the basis of engagement.

The interpretable risk levels provided by the COVID-19 Event Risk Assessment website encourages visitors to take steps to reduce their risk of infection, such as physical distancing, washing hands and wearing a mask. By illustrating how risk increases nonlinearly with event size, the tool may be particularly useful in encouraging large event planners to reschedule or cancel events, move to a safer format (for example, outdoors where transmission risk is reduced or online when possible), thereby averting potential exposures. As such, the website is of particular relevance given the relaxation of shelter-in-place orders across the United States, including restrictions on gatherings. These relaxations of non-pharmaceutical interventions indicate that individuals must remain informed of the personal risk involved with everyday activities so as to modify their behaviour accordingly. By providing a quantitative tool to convey the ongoing risk of the pandemic, we hope to supplement and bolster local public health advisories. The model's risk estimate is designed to display information that is tailored to an individual's immediate locale in a unit of measurement that is relevant and interpretable. We note that regardless of the risk values calculated, individuals should continue to follow their administrative unit's local public health policies.

There are multiple ways to extend these findings to improve local estimates. First, the model uses a binomial probability of risk that assumes that risk is homogeneous at county levels. We anticipate there will be variation within counties (for example, see studies on heterogeneous risk within New York City boroughs[28]). However, because data on cases are reported at the county level, further refinement to tract or zip code levels is not yet feasible. In addition, the website does not break down risk in terms of other socioeconomic correlates, or by race, gender or other personalized factors including the effect of mobility from nearby counties. Second, the risk model assumes that individuals are equally likely to attend an event, whereas increases in symptomatic case isolation indicates that a fraction of infectious individuals are unlikely to attend events (the same applies to those hospitalized, albeit that is a much smaller fraction of the total). We note that, despite their inclusion in our calculations, symptomatic individuals would probably already limit their attendance at large gatherings. However, the fraction of symptomatic individuals may change with the age structure of infections and prevalence of certain comorbidities (for example, asthma, diabetes, heart disease). Furthermore, an individual's ability to isolate may depend on additional socioeconomic factors and vary over the course of infection. While we assume uniform infectiousness across a 10-day period, infectiousness varies between individuals and across time for a single individual. Improved estimates of the duration of infectiousness could improve these calculations.

Yet, perhaps the largest driver of uncertainty remains ascertainment bias. Ascertainment bias denotes the number of actual cases for each documented case. A recent population-wide CDC serosurvey found that ascertainment bias ranged from 6–24-fold above PCR documented cases in March and April[21]. Phase 2 serology surveys of populations revealed a range in ascertainment bias from between two and 14, with a median of nine to ten times[29]. Rapid, population-wide serosurveys are needed to connect case reports to localized estimates of ascertainment bias. Integrating such serosurveys at state levels or improvements in estimates of ascertainment bias using statistical or mechanistic models[30,31] could further refine variation in event risk estimates.

In closing, by connecting real-time case reports in the context of risk associated with events, the website attracted more than 2 million visitors in the first 2 months after release of a county-level risk tool. This interest showcases the importance of translating epidemiological statistics into real-world context. In doing so, we hope that health departments in the United States and globally consider integrating event-associated risk models in current and future pandemic responses as part of public awareness campaigns. Indeed, we have already extended the same approach to subnational risk estimation in three European countries, Italy, Switzerland and the UK, and note that an Italian language risk assessment map for Italy based on the current approach is also available at http://covid19eventi.datainterfaces.org/ and that a Spanish-language risk assessment map for Spain based on the current approach is also available at https://eventosycovid19.es/. Spatial risk models can help to convey heterogeneous risk at local levels, and provide accessible information that can help to justify the choice of restrictions on gatherings as part of integrative campaigns to control spread. For SARS-CoV-2, the open-source and publicly available dashboard highlights the fact that there is a >99% risk that one (or more) individuals may be infected in groups of 500–1,000, in the vast majority of locations as of October 2020; these sizes are consistent with typical enrollment at K-12 schools. Hence, it is critical that re-openings of businesses and schools devise policies for testing, mask wearing and other non-pharmaceutical interventions to ensure that one case does not soon become many.

Methods

Probability model. We estimated the probability that one or more individuals may have SARS-CoV-2 in events of different sizes via a binomial assumption of homogeneous risk. Let p denote the probability that a randomly selected individual in a focal population is infected. Hence, the probability that each of n individuals is not infected must be $(1-p)^n$ and by extension the probability that one (or more) individuals is infected must be $1-(1-p)^n$; we define this as the event gathering risk. This formalism was used as the basis for early estimates to communicate risk of large gatherings in March 2020 using a scenario-based approach to estimating p within the United States[17–20].

Circulating case estimate. At a county level, the circulating per-capita probability of infection is defined as the estimated number of circulating cases divided by the census population. The circulating case counts are defined, operationally, in two stages. First, the number of newly documented cases over the past 10 days (d) are obtained via data via state departments of public health. Data were aggregated and accessed from the *New York Times*' repository of COVID-19 data[32] using a standard application programming interface. The choice of 10 d is consistent with CDC guidelines on durations of infectiousness[33]. Second, the number of newly documented cases is multiplied by an ascertainment bias to yield the estimated number of circulating cases. The default ascertainment bias is ten times, consistent with a median of 9–10 in population-wide serological surveys conducted by the Centers for Disease Control and Prevention[21,29]; with a secondary option of five times.

Visualization code. The code to visualize county- and state-level risk was written in R and used the R-Shiny Package for map deployment. The input data was a county shapefile from the US Census that included all 50 states, the District of Columbia and Puerto Rico whose boundaries were generalized using the 'rmapshaper' package. This file was converted to a geojson file for faster drawing. The projection was relegated to a web Mercator standard instead of a traditional conic projection due to the constraints of the R package. New York City was

agglomerated as a set of five counties to accommodate the *New York Times*' county level case data[32], which reported New York City as a single region. The risk value shown on the county-level map takes into account the county's new cases for the past 10 d, the user's chosen ascertainment bias (five or ten) from a radio button and the user's chosen case size from a slider with eight discrete increments (10, 25, 50, 100, 500, 1,000, 5,000 and 10,000). The map symbology was chosen as a univariate colour ramp showing intensity in red, and allows for interactive zooming and panning. On hovering over it, a pop up shows the county name and the likelihood (in terms of a percentage) that an individual at that event is infected with SARS-CoV-2.

Web application. The web application is built using the R-Shiny web development framework and deployed as a self-contained Docker container using the open-source Shiny-server. Application containers are deployed to a fleet of servers hosted at Georgia Institute of Technology, with multiple application instances running on each. Users are load-balanced across instances using Nginx. All static data used in the application (for example, map HTML files, data used for interactive plots) are automatically updated and distributed to each application instance.

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

Data availability

Population demographics for US states and counties were obtained from the publicly available United States Census Bureau American Community Survey for 2018 (ref. [34]). State-level cases were obtained from CovidTracking.com, a project developed by Alexis Madrigal and colleagues at The Atlantic[4]. County-level cases were obtained from the *New York Times* Github site[32].

Code availability

The website code is open source and available on Github at https://github.com/appliedbinf/covid19-event-risk-planner.

References

1. *Coronavirus Disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update* (World Health Organization, 2020); https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports
2. Ferguson, N. et al. *Report 9: Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID19 Mortality and Healthcare Demand* (Imperial College London, 2020). https://doi.org/10.25561/77482
3. Coronavirus in the U.S.: latest map and case count. *New York Times* (continuously updated); https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html
4. *The COVID Tracking Project* (The Atlantic Monthly Group, 2020); https://covidtracking.com
5. Gottlieb, S. & Strain, M. R. Want schools to open? Get serious about outbreaks. *Wall Street Journal* https://www.wsj.com/articles/want-schools-to-open-get-serious-about-outbreaks-11596396822 (2 August 2020).
6. Levinson, M., Cevik, M. & Lipsitch, M. Reopening primary schools during the pandemic. *N. Engl. J. Med.* **383**, 981–985 (2020).
7. Fauci, A. S., Lane, H. C. & Redfield, R. R. COVID-19—navigating the uncharted. *N. Engl. J. Med.* **382**, 1268–1269 (2020).
8. Park, S. W. et al. Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: framework and applications to the novel coronavirus (SARS-CoV-2) outbreak. *J. R. Soc. Interface* **17**, 20200144 (2020).
9. *COVID-19 Pandemic Planning Scenarios* (Centers for Disease Control and Prevention, 2020); https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html
10. Lavezzo, E. et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature* **584**, 425–429 (2020).
11. Nishiura, H., Linton, N. M. & Akhmetzhanov, A. R. Serial interval of novel coronavirus (COVID-19) infections. *Int. J. Infect. Dis.* **93**, 284–286 (2020).
12. Li, R. et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020).
13. Kain, M. P., Childs, M. L., Becker, A. D. & Mordecai, E. A. Chopping the tail: how preventing superspreading can help to maintain COVID-19 control. Preprint at *medRxiv* https://doi.org/10.1101/2020.06.30.20143115 (2020).
14. Liu, Y., Eggo, R. M. & Kucharski, A. J. Secondary attack rate and superspreading events for SARS-CoV-2. *Lancet* **395**, e47 (2020).
15. Xu, X.-K. et al. Reconstruction of transmission pairs for novel Coronavirus disease 2019 (COVID-19) in mainland China: estimation of superspreading events, serial interval, and hazard of infection. *Clin. Infect. Dis.* https://doi.org/10.1093/cid/ciaa790 (2020).
16. Frieden, T. R. & Lee, C. T. Early release-identifying and interrupting superspreading events—implications for control of severe acute respiratory syndrome Coronavirus 2. *Emerg. Infect. Dis.* **26**, 1059–1066 (2020).
17. Weitz, J. COVID-19 Event Risk Assessment Planner. *figshare* https://doi.org/10.6084/m9.figshare.11965533.v1 (2020).
18. Arnold, C. COVID-19: Biomedical research in a world under social-distancing measures. *Nat. Med.* https://doi.org/10.1038/d41591-020-00005-1 (2020).
19. Maggiacomo, T. & Greshko, M. See why keeping groups small can save lives in the era of COVID-19. *National Geographic* https://www.nationalgeographic.com/science/2020/03/graphic-see-why-small-groups-are-safer-during-covid-19-coronavirus-pandemic/ (19 March 2020).
20. Downey, M. Scientists do the math to show how large events like March Madness could spread coronavirus. *Atlanta Journal Constitution* https://www.ajc.com/blog/get-schooled/scientists-the-math-show-how-large-events-like-march-madness-could-spread-coronavirus/g1pVdzQgJS5aoPnadBqyXO (13 March 2020).
21. Havers, F. P. et al. Seroprevalence of antibodies to SARS-CoV-2 in 10 sites in the United States, March 23-May 12, 2020. *JAMA Intern. Med.* https://doi.org/10.1001/jamainternmed.2020.4130 (2020).
22. Bourouiba, L. Turbulent gas clouds and respiratory pathogen emissions: potential implications for reducing transmission of COVID-19. *JAMA* **323**, 1837–1838 (2020).
23. Mittal, R., Ni, R. & Seo, J.-H. The flow physics of COVID-19. *J. Fluid Mech.* https://doi.org/10.1017/jfm.2020.330 (2020).
24. Yi, J. S., Kang, Y. A., Stasko, J. T. & Jacko, J. A. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Trans. Vis. Comput. Graph.* **13**, 1224–1231 (2007).
25. *Mapping Mobility Changes in Response to COVID-19* (GeoDS Laboratory, University of Wisconsin-Madison, 2020); https://geods.geography.wisc.edu/covid19/physical-distancing/
26. Katz, J., Sanger-Katz, M. & Quealy, K. A detailed map of who is wearing masks in the U.S. *New York Times* https://www.nytimes.com/interactive/2020/07/17/upshot/coronavirus-face-mask-map.html (17 July 2020).
27. Armstrong, M. P. & Densham, P. J. Database organization strategies for spatial decision support systems. *Int. J. Geogr. Inf. Syst.* **4**, 3–20 (1990).
28. Sy, K. T. L., Martinez, M. E., Rader, B. & White, L. F. Socioeconomic disparities in subway use and COVID-19 outcomes in New York City. Preprint at *medRxiv* https://doi.org/10.1101/2020.05.28.20115949 (2020).
29. *Commercial Laboratory Seroprevalence Survey Data* (Centers for Disease Control and Prevention, accessed 15 August 2020); https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/commercial-lab-surveys.html
30. Perkins, T. A. et al. Estimating unobserved SARS-CoV-2 infections in the United States. *Proc. Natl Acad. Sci. USA* **117**, 22597–22602 (2020).
31. Beckett, S. J., Dominguez-Mirazo, M., Lee, S., Andris, C. & Weitz, J. S. Spread of COVID-19 through Georgia, USA. Near-term projections and impacts of social distancing via a metapopulation model. Preprint at *medRxiv* https://doi.org/10.1101/2020.05.28.20115642 (2020).
32. *Coronavirus (COVID-19) Data in the United States* (New York Times, accessed 10 August 2020); https://github.com/nytimes/covid-19-data
33. *Duration of Isolation and Precautions for Adults with COVID-19* (Centers for Disease Control and Prevention, 2020); https://www.cdc.gov/coronavirus/2019-ncov/hcp/duration-isolation.html
34. *American Community Survey Dataset, 2018* (United States Census Bureau, accessed 15 April 2020); https://www.census.gov/programs-surveys/acs

Author contributions

A.C. led the development of the web server, contributed analysis and analysed data. S.L. led the development of the map-based visualizations, contributed analysis and analysed data. M.H. contributed to website design, study design and edited the manuscript. S.J.B. and Q.N. contributed to the development of the international risk assessments. T.H. contributed to the development of the web server. C.A. codesigned the study, provided oversight to all aspects of map development, analysed data and cowrote the manuscript. J.S.W. designed the study, provided oversight for all aspects of the study, developed the core modelling framework, analysed data and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

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

nature research

Corresponding author(s): Joshua S. Weitz; Clio Andris

Last updated by author(s): 10/10/2020

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 | Visualization code: The code to visualize county- and state-level risk was written in R and used the R Shiny Package for map deployment. The input data was a county shapefile from the U.S. Census that included all 50 states, the District of Columbia, and Puerto Rico whose boundaries were generalized using the 'rmapshaper' package. This file was converted to a geojson file for faster drawing. The projection was relegated to a web Mercator standard instead of a traditional conic projection due to the constraints of the R package. New York City was agglomerated as a set of five counties in order to accommodate the New York Times' county level case data (New York Times, 2020a), which reported New York City as a single region. The risk value shown on the county-level map takes into account the county's new cases for the past 10 days, the user's chosen ascertainment bias (5 or 10) from a radio button, and the user's chosen case size from a slider with eight discrete increments (10, 25, 50, 100, 500, 1,000, 5,000, and 10,000). The map symbology was chosen as a univariate color ramp showing intensity in red, and allows for interactive zooming and panning. Upon hover, a pop up shows the county name and the likelihood (in terms of a percentage) that an individual at that event is infected with SARS-CoV-2.

Web application: The web application is built using the R-Shiny web development framework and deployed as a self-contained Docker container using the open-source shiny-server. Application containers are deployed to a fleet of servers hosted at Georgia Institute of Technology, with multiple application instances running on each. Users are load-balanced across instances using Nginx. All static data used in the application (e.g. map HTML files, data used for interactive plots) are automatically updated and distributed to each application instance.

Data Availability: Population demographics for US states and counties were obtained from the publicly available United States Census Bureau American Community Survey for 2018 (U.S. Census Bureau, 2019). State-level cases were obtained from CovidTracking.com – a project developed by Alexis Madrigal and colleagues at The Atlantic (The Atlantic Monthly Group, 2020). County-level cases were obtained from the New York Times github site (New York Times, 2020a).

Code availability: The website code is open source and available on Github: https://github.com/appliedbinf/covid19-event-risk-planner.

Data analysis	The website code is open source and available on Github: https://github.com/appliedbinf/covid19-event-risk-planner.

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 Availability: Population demographics for US states and counties were obtained from the publicly available United States Census Bureau American Community Survey for 2018 (U.S. Census Bureau, 2019). State-level cases were obtained from CovidTracking.com – a project developed by Alexis Madrigal and colleagues at The Atlantic (The Atlantic Monthly Group, 2020). County-level cases were obtained from the New York Times github site (New York Times, 2020a).

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

Life sciences study design

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

Sample size	Data Availability: Population demographics for US states and counties were obtained from the publicly available United States Census Bureau American Community Survey for 2018 (U.S. Census Bureau, 2019). State-level cases were obtained from CovidTracking.com – a project developed by Alexis Madrigal and colleagues at The Atlantic (The Atlantic Monthly Group, 2020). County-level cases were obtained from the New York Times github site (New York Times, 2020a).
Data exclusions	N/A
Replication	N/A
Randomization	N/A
Blinding	N/A

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

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☒ ☐	Animals and other organisms
☒ ☐	Human research participants
☒ ☐	Clinical data
☒ ☐	Dual use research of concern

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging