nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01654-1

Mass gatherings for political expression had no discernible association with the local course of the COVID-19 pandemic in the USA in 2020 and 2021

Received: 2 August 2022

Accepted: 14 June 2023

Published online: 31 July 2023

Check for updates

Eric Feltham [1,2] ✉, Laura Forastiere [1,3], Marcus Alexander[1,4] & Nicholas A. Christakis [1,2,5,6]

Epidemic disease can spread during mass gatherings. We assessed the impact of a type of mass gathering about which comprehensive data were available on the local-area trajectory of the COVID-19 epidemic. Here we examined five types of political event in 2020 and 2021: the US primary elections, the US Senate special election in Georgia, the gubernatorial elections in New Jersey and Virginia, Donald Trump's political rallies and the Black Lives Matter protests. Our study period encompassed over 700 such mass gatherings during multiple phases of the pandemic. We used data from the 48 contiguous states, representing 3,108 counties, and we implemented a novel extension of a recently developed non-parametric, generalized difference-in-difference estimator with a (high-quality) matching procedure for panel data to estimate the average effect of the gatherings on local mortality and other outcomes. There were no statistically significant increases in cases, deaths or a measure of epidemic transmissibility (R_t) in a 40-day period following large-scale political activities. We estimated small and statistically non-significant effects, corresponding to an average difference of −0.0567 deaths (95% CI = −0.319, 0.162) and 8.275 cases (95% CI = −1.383, 20.7) on each day for counties that held mass gatherings for political expression compared to matched control counties. In sum, there is no statistical evidence of a material increase in local COVID-19 deaths, cases or transmissibility after mass gatherings for political expression during the first 2 years of the pandemic in the USA. This may relate to the specific manner in which such activities are typically conducted.

Coronavirus disease 2019 (COVID-19), like any serious outbreak of a contagious disease, can place the virtues of public health and civic engagement—in the form of mass gatherings—into direct conflict[1]. Indeed, epidemics pose an especially difficult problem for the political process, which relies on an active and engaged citizenry so that the public may hold their leaders accountable for policies enacted to manage the epidemic in the first place. Even if the will of the majority is aligned with the public interest of most effectively containing, mitigating and recovering from an epidemic, political participation is often necessary to enact the majority's preferences. Yet, when citizens

A full list of affiliations appears at the end of the paper. ✉e-mail: eric.feltham@aya.yale.edu

gather in person, participation itself can contribute to a worsening of the epidemic. Both from a public health perspective and an individual point-of-view, the question arises whether in-person mass gatherings, in particular for political purposes, are safe.

In the spring of 2020, the USA was in the midst of its presidential primary elections when the COVID-19 pandemic began in earnest. Concerns immediately arose about the potential of in-person voting to lead to a rise in community transmission of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus, making elections into potential super-spreader events with the adverse consequence of otherwise-preventable excess deaths in communities across the USA. Furthermore, election procedures themselves became a partisan issue under COVID-19[2,3], with some Democrats attempting to signal a commitment to public health by avoiding the polls, and some Republicans signalling defiance of COVID-19 by showing up in person. COVID-19 sparked a partisan uproar over postponing primaries and mail-in voting, reaching its zenith with the Wisconsin Supreme Court battle[4], where a Republican effort blocked an attempt by Governor Tony Evers, a Democrat, to reschedule the primary election, given his public health concerns. By November 2020, Dr Anthony Fauci, the then-director of the National Institute of Allergy and Infectious Diseases, had stated that in-person voting was likely to be safe, so long as social distancing measures are in place[5].

In addition to the primary elections, which were held in the spring of 2020 in all 50 states, former President Donald Trump held 67 political rallies in 15 states across the country during the 2020 election season. The Trump rallies were anticipated to cause a surge in COVID-19 and were among the largest events held in the USA at the time[6], drawing between hundreds and thousands of people[7]. The USA also saw a surge in large protests for racial justice in the summer of 2020 and sporadically thereafter. The 'Black Lives Matter' (BLM) protests (of which there were hundreds) have been characterized as the largest concerted protest movement in US history[8], and cities and towns across the country saw demonstrations ranging in size from tens to hundreds of thousands of people. Early on, some experts worried that the BLM protests were a risk for increasing COVID-19 transmission while the USA was in the middle of its second wave of the epidemic[9]. At the same time, over 1,000 public health workers controversially signed a letter that minimized the health risks of participating in the protests, arguing that the political issue of racial justice outweighed any public health risks from COVID-19[10,11].

Political activities as a type of mass gathering during the COVID-19 pandemic were a contentious issue and pressing policy problem around the world. For instance, in late 2021, a wave of teachers were reported to have tested positive and died subsequent to working at the polls in India[12], and Argentina decided to push back its 2021 midterm elections, while nine countries in Latin America held other elections in 2021[13]. Leaders around the globe decided whether to carry out or postpone their upcoming elections and wrestle with rules for voting[14–17]. In addition, large protests continue around the world in places where the pandemic was ongoing, including in India, Colombia, Cuba and others (Global Protest Tracker, https://carnegieendowment.org/publications/interactive/protest-tracker). Cuba had a record-breaking number of daily COVID-19 deaths around the time of its protests[18]. In India, the government has urged the dissolution of large-scale protests, citing the dangers of COVID-19 transmission[19]. In Russia in February and March 2022, anti-war protests were banned and COVID-19 was used as the justification[20]. In most of these places, powerful vaccines remain scarce and traditional non-pharmaceutical interventions (NPIs) (including a ban on mass gatherings) were the primary strategy to mitigate the spread of COVID-19.

Political activities are thus an important subset of mass gatherings that have occurred during the COVID-19 pandemic, with some properties that distinguish them from concerts, sporting events or other gatherings marked by more sustained or intimate or indoor contact. Still, the political gatherings studied here are crucial to the functioning of democracy and are important in their own right from a public health perspective. Moreover, it is possible to comprehensively enumerate certain types of political mass gathering and to assemble data about them. We thus examine political gatherings both because it is possible to collect complete data on such gatherings and because of their special importance.

Results

Methodology summary

We assembled a complete dataset of essentially all mass gatherings of particular types for a 2-year period (see Methods for data availability). We examined the impact of five different sorts of major political events on the spread of COVID-19 in the USA, in the period from 17 March 2020 to 2 November 2021: the US primary elections (17 March 2020 to 11 August 2020); the Georgia (GA) special election (5 January 2021); the gubernatorial elections in New Jersey (NJ) and Virginia (VA) (2 November 2021); Trump's 2020 campaign rallies (6 June 2020 to 2 November 2020); and the BLM protests (Summer 2020).

We covered every state-wide election over our whole study period in which there was a substantial amount of in-person voting (and excluded the national general election on 3 November 2020). We included 100% of the official Trump rallies (67 events). For the BLM protests, we specifically examined 100% of the large-scale protests in the time period considered—a total of 658 county-level events with a crowd of at least 800 persons, of which 94% are directly linked to the BLM movement. The events we studied thus span the types of political activity engaged in by the public, the geographic area of the USA and the various phases of the epidemic, ranging from its early onset in March 2020, the second wave over the summer of 2020, the peak of the third wave in January 2021 and the fourth wave in 2021, in which the Delta variant became the dominant strain in the USA.

The preponderance of studies that examine the effectiveness of various NPIs and the impact of in-person gatherings on the spread of the SARS-CoV-2 virus so far only consider the Alpha variant[21,22]. However, the Delta variant was found to be between 40–80% more infectious than the initial strain[23,24] and caused a marked increase in hospitalizations and deaths, despite the increasing administration of vaccines by the time it emerged[25]. The increased transmissibility of new variants may fundamentally alter the relative effectiveness of various NPIs, including the banning of mass gatherings. Therefore, we included in our analysis events that took place in late 2021 to examine the impact of the same type of event under different predominant strains of the virus. Furthermore, our analyses are naturally stratified by vaccine period. In the USA, the first vaccines were administered to the public on 14 December 2020[26]. The 2020 primary elections, the BLM protests and the Trump campaign rallies all predate the vaccine rollout in the USA (the last rally in our data was held on 2 November 2020). Similarly, the GA special election took place on 5 January 2021, at which time under 1% of the population in the state of GA had their first vaccine dose. By contrast, the NJ and VA gubernatorial elections were held when 64 and 62% of those states' populations had complete vaccine series (two doses of a two-dose vaccine, or one dose of a single-dose vaccine), respectively[27].

Modelling the spread of COVID-19 is particularly challenging for two major reasons. The first relates to the quality of the underlying data, which can be incomplete or noisy. The USA has struggled to test an adequate number of individuals and, especially early on, there were not enough tests to adequately track the number of cases. Many of the mass gatherings in our data occurred during a rapid increase in the number of tests administered (see Supplementary Fig. 1), which also occurred unevenly across US counties. Moreover, testing relies on the interest and ability of individuals to get tested and of institutions to test them. Consequently, there was substantial variation in testing capacity and rates at the county level, as well as over time[28]. Furthermore, given that 66% of registered voters reported being at least somewhat concerned about the safety of voting in a March 2020 survey[29] and

evidence that individuals may have altered their behaviour in response to the BLM protests[30], it is plausible that testing may be endogenous to treatment; that is, individuals in counties that held political events may have been more likely to seek testing, thus potentially inducing bias in the outcomes measured across counties. Consequently, to track the course of the epidemic, we primarily focused on (more accurate) 'mortality' data, using both death counts and rates. However, we also explored other outcomes, including case rates, virus transmissibility (R_t) (from an epidemiological model tailored to COVID-19)[31–33] and mobility (based on mobile phone data).

Second, many of the available statistical tools make assumptions that are not appropriate to modelling the spread of a virus[34]. Consequently, we developed and applied a non-parametric generalized difference-in-differences estimator for panel data[35] to account for the nonlinear epidemic curve, with a matching procedure at the county level (see Methods for methodological details and code availability). We made use of county-level data, as they are the most granular COVID-19 data widely available in the USA, affording the best possible means to track changes in the spread of the virus. For each timepoint (day), we matched 'treated' counties (that had a political gathering) with multiple 'control' counties (that did not have any political gathering during a sufficiently long time window) that had similar socio-demographic characteristics and similar dynamics of the epidemic before and up to the time of the 'treatment'.

Crucially, to capture the epidemic characteristics, we matched on the county-level cumulative death rate and also the date of the first reported case in a county. We selected, at most, the five best matches to a treated county as the basis for comparison, which is a widely used strategy for methods with multiple matches[36]. Calipers were applied individually to specific matching covariates as needed to ensure that covariate (im)balance remains below 0.1, on average, over the matching period. Consequently, fewer than five matches were used when a match was of poor quality. Treated observations were dropped from the analysis if they had no good-quality matches. Supplementary Fig. 2 displays the distribution of the number of matches for the main analyses; see Supplementary Information (Section 3) for details on the numbers of matches and treated units for each model presented. In addition, we present the results for each analysis before refining the matches to, at most, the best five, presented in Supplementary Information Section 3. Indeed, we conducted a broad variety of robustness checks regarding the quality of this matching (see Methods and Supplementary Information Section 2 for further details); we present robustness results with varying caliper widths, pre-outcome-window results and different matching covariate specifications.

Our approach extends a statistical approach that has been recently introduced and validated[35]. In particular, we define different 'causal' parameters that are more meaningful for this setting and implement a sliding time window to find matches for units that can be considered 'treated' at each timepoint (see Methods). This approach avoids certain limitations in other methods previously applied to assess the effect of mass gatherings on the spread of COVID-19 whose assumptions may not be well met in this setting[37,38]. The present approach may be applied in any setting governed by a nonlinear contagion process, whether the 'treatment' occurs in a limited time frame, is repeated over time (time-varying treatment) or varies across the population.

Furthermore, each of the event types we analysed presents different challenges that benefit from alteration of our estimation procedure; in the case of the Trump rallies, for instance, we assumed the presence of geographic spillover effects[39], and in the case of the BLM rallies, we had many units with multiple treatment-like protests within a narrow window of time.

We measured the average effect of an event on the observations where the event actually occurred (the so-called 'average treatment effect on the treated' or ATT) by comparing the change in outcome over time across matched units in different treatment and control groups. The difference-in-differences strategy requires that, in the absence of the event, the outcome would have followed the same trend in the treated and control arms. Nonetheless, our approach makes less stringent assumptions than many traditional approaches for the estimation of causal effects with panel data[35]. As opposed to the standard difference-in-differences estimator, this method relies on a parallel trend assumption only after conditioning on both baseline and time-varying covariates before the intervention, including the pre-treatment outcome history.

Covariate adjustment was conducted using a matching procedure whereby the five best matches were selected and used for each treated county (see Methods for details). For each model, the quality of the five best matches was examined with reference to a balance score, calculated on the basis of the standardized distance between a matched unit and its best matches in terms of covariate similarity. In each case, acceptable balance scores were obtained on average for each matching covariate (see Methods and Supplementary Information Section 2 for detailed reporting on match quality for each analysis). This approach also allows for time-varying treatments that can occur multiple times over the observed window.

We used a range of variables describing the counties, including demographic, political and mobility variables. Our mobility data tracked daily visits to points of interest at individual locations across the USA in census block groups, typically covering between 600 and 3,000 individuals, aggregated to our 3,108 counties (see Methods). Mobility data have previously been linked to the spread of COVID-19[39–41]. Many existing studies use general movement patterns[30,42,43], such as overall time spent away from home, which does not necessarily indicate whether individuals engaged in the sort of close interpersonal contact that is heavily responsible for the spread of COVID-19[40]. However, more recent studies have used finer-grained mobility metrics[28,40,41,44] in line with our approach. We tracked movement specifically at full-service restaurants, grocers, and fitness and recreation facilities, some of which are known to be high-risk locations[28,41], and directly included these measures in the matching procedure.

In our analysis, a unit is the entire time series of a US county, X_i. A unit is treated on a specific day, $X_{i,t}$, if it holds an event (mass gathering) of interest (for example, an election, rally or protest) on that day. For such a given treated observation, we matched on that county's characteristics for 30 d up to 1 d before treatment. Our primary outcome of interest is the county-level death rate (measured in deaths per 10,000 people, per million people; or in absolute counts of deaths). We contrasted the treated and matched unit's outcome for each day in an outcome window defined as 10–40 d after an event, which is the epidemiologically informed period in which we would expect an effect on the mortality rate from an event where contagion with SARS-CoV-2 might occur[45,46]. This window captures the first wave of deaths that would be expected to occur from infection on the day of a mass gathering.

For consistency, we also used the same time window for the primary analyses of the other outcomes (see Methods), that is, we used the same window for cases. While the incubation period is shorter than 10 d, we believe that this window effectively accounts for reporting delays in testing, and the period from 10 to 40 d can capture subsequent waves of infections that could be traced to infections on the day of an event.

For each day that we estimated in the outcome window, we defined the 'crossover window' as the period in which we examined the treatment history of potential matched control counties. A control county may not be treated during the post-treatment portion of the crossover window, defined as 30 d before the given day in the outcome window (a day F between 10 and 40) up to the day before treatment. Eligible control units must have treatment histories that are similar to the treated unit in the pre-treatment crossover window. Thus, we handled the treatment histories in a nuanced manner, separately from the fixed covariate matching window (Fig. 1). In addition, for each of the main analyses presented, we also calculated the ATTs in the crossover

https://doi.org/10.1038/s41562-023-01654-1

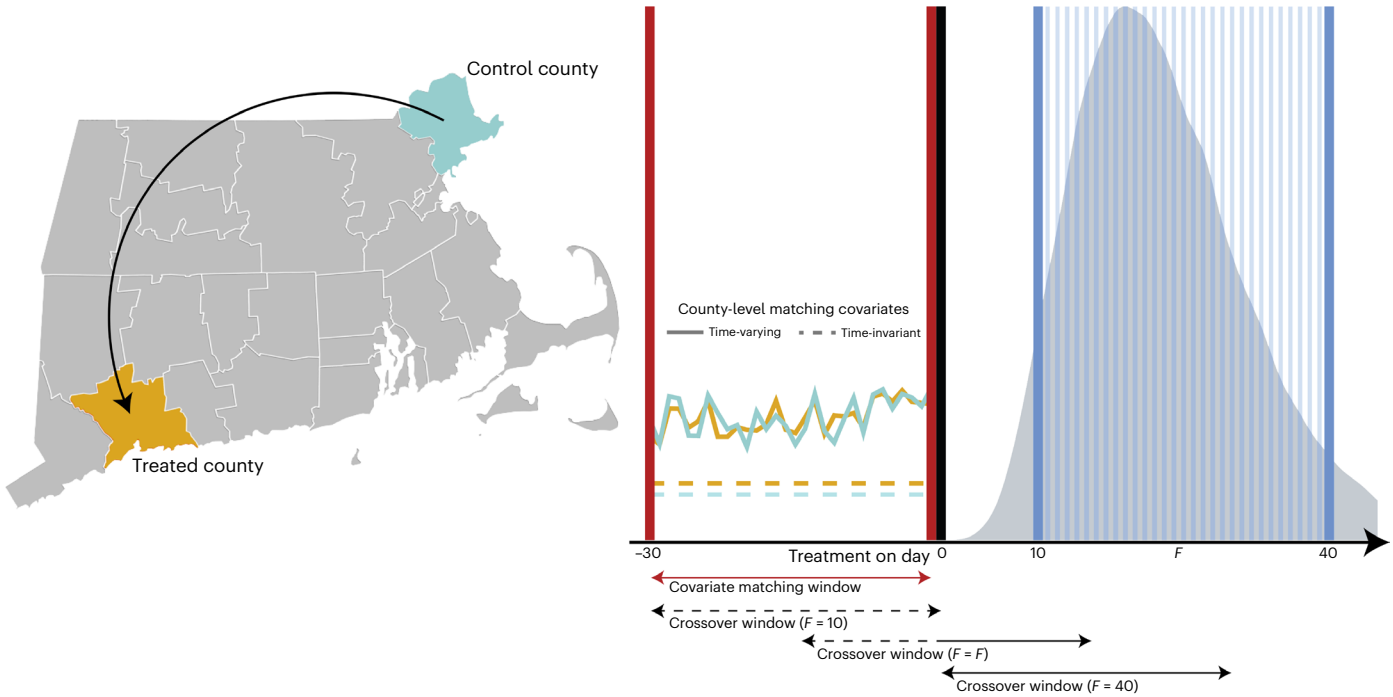


Fig. 1 | Overview of matching and estimation. We estimated the ATT for each day in the outcome window, from 10 to 40 d after a treatment event (an election, rally or protest), which corresponds to the 5th and 95th percentiles of the empirically observed distribution of times from infection to death from SARS-CoV-2 (vertical blue lines). Given this, we used the same time window for the primary analyses of the other outcomes. Treated counties were matched to, at most, the five best control counties on the basis of similarity of their covariate values from 30 d to 1 d before a treatment event, that is, the 'covariate matching window' (red solid line arrow), although here we show only one matched county for illustration (respectively, represented as yellow and green lines). Treated counties may have between one and five matches, contingent on the quality of available matches. Covariates are either time-varying (represented as solid yellow and green lines; for example, the cumulative death rate, mobility based on phone data) or static (represented as dashed yellow and green lines; for example, population density) over our study horizon. The 'crossover window' is the period in which we examined the treatment history of potential matched control counties. A control county may not be treated during the post-treatment portion of the crossover window (solid black arrows), defined as 30 d before *F*, up to the day before treatment. Eligible control units must have treatment histories that are similar to the treated unit in the pre-treatment crossover window, 30 d before *F* to 10 d before *F* (when defined for the post-treatment period, shown by the dashed black arrows). That is, we did not only match on a fixed period 30 d before treatment (which is what we did for the matching covariates), but also dealt with the treatment histories in a nuanced manner. Finally, calipers were applied individually to specific matching covariates as needed to ensure that covariate (im)balance remains below 0.1 on average over the matching period. Treated observations were dropped from the analysis if they had no good-quality matches.

window and up to the start of the outcome window (in these cases, we observed results that are consistent with those in the outcome window itself; see Extended Data Fig. 1).

Valid estimation in this context requires that there is sufficient variation of the treatment onsets to find a suitable set of matches for each treated unit. For example, starting with an analysis of the primaries, the rolling schedule of these primaries in the USA from February to August 2020 helps to satisfy this condition. Furthermore, while states are not randomly assigned to primary election days and states with later elections may differ in their characteristics from those with earlier elections, we observed sufficient county-level variation in the features of interest to obtain acceptable matching to the treated units (see Methods and Supplementary Information Section 2). While several states had primaries that were either cancelled or rescheduled due to COVID-19, it seems justified to treat cancelled primaries as control units over those dates. Furthermore, for those primaries that were rescheduled, the overall difference in the case and death rates between the original and rescheduled dates is close to zero (Supplementary Fig. 3). The rescheduled primaries were moved to periods in which community spread still occurred, allowing us to use the counties therein as valid treated units.

Overall assessment of local mortality and cases

We first report a summary analysis that examined the overall effect for 'all' event types and time periods combined, using county-level death rates and case rates (Fig. 2a), which correspond to an average ATT of −0.257 (95% confidence interval (CI) = −3.482–2.777, *P* value (*P*) = 0.853, Bayes factor (BF) = 0.030) deaths per million persons for the death rates, and an average ATT of −20.949 (95% CI = −80.987–27.604, *P* = 0.414, BF = 0.038) cases per million persons. The point estimates on these effects are, in fact, negative. The high quality of the covariate balance between cases and controls for this omnibus analysis is shown in Fig. 2b,c.

We emphasize that in most cases, the changes in the wake of the events are small in magnitude in absolute terms as well. In the context of the death counts, the foregoing rate translates to an ATT for the death 'count' of −0.0567 (95% CI = −0.319–0.162), such that 0.0567 more deaths would have occurred on average for each day over the outcome window in the counties that held events, compared with a situation in which the political events had not been held.

We also stress that mortality is the more reliable and less biased metric (since people might be motivated to modify their testing behaviour before or after events), as we found a meaningful increase in the total number of tests performed over each day of the outcome window and many of the earlier events occurred during a rapid ramp-up in testing at the national level.

Specific types of event and local mortality and cases

We first analysed the primary elections of the spring and summer of 2020. Our estimates for the effect of the primary elections for each day

https://doi.org/10.1038/s41562-023-01654-1

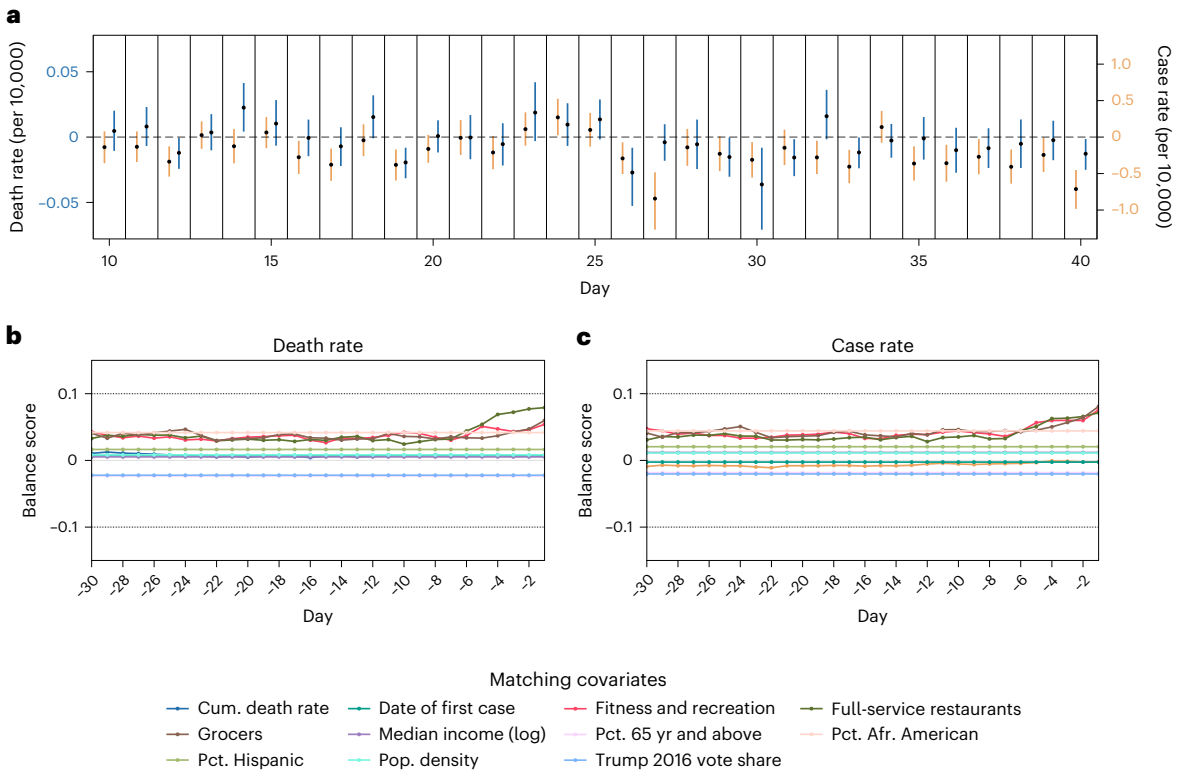


Fig. 2 | Impact of all mass gatherings for political expression on COVID-19 mortality and case rates in the USA. Generally, we observed no statistically significant increase on average for treated counties from a period of 10–40 d after a primary election was held for either outcome. **a**, Overall ATT estimates for the political events, representing the average difference in the change in death (blue) and case (orange) rates, from the day before treatment to 10–40 d after an election. Error bars indicate 95% CIs. For the death rate outcome, for each day on days 10–40, N = 1,493, 1,486, 1,477, 1,474, 1,472, 1,467, 1,461, 1,458, 1,456, 1,456, 1,455, 1,451, 1,450, 1,449, 1,445, 1,445, 1,444, 1,443, 1,442, 1,442, 1,442, 1,434, 1,433, 1,433, 1,433, 1,433, 1,433, 1,433, 1,431, 1,430 and 1,429. N is the number of treated counties on a day in the outcome window. There are, on average, 6,085 matched units across the period. For the case rate outcome, for each day on days 10–40, N = 1,680, 1,672, 1,663, 1,661, 1,656, 1,652, 1,649, 1,646, 1,645, 1,644, 1,642, 1,640, 1,638, 1,637, 1,636, 1,636, 1,634, 1,633, 1,632, 1,631, 1,631, 1,621, 1,621, 1,620, 1,620, 1,619, 1,619, 1,619, 1,618, 1,618 and 1,618. N is the number of treated counties on a day in the outcome window. There are, on average, 7,277 matched units across the period. **b**, Covariate balance for the death rate model. The balance scores reflect the similarity between the treated units and their matched counties for 30 d before, up to 1 d before a political event. They reflect the covariate balance after matching refinement and the application of a caliper to ensure match quality. The balance scores are, on average over the matching window, within the threshold of 0.1, indicating sufficient similarity between the treated and matched counties for the estimated ATTs. **c**, Covariate balance for the case rate model. The balance scores are, on average over the matching window, again within the threshold of 0.1. See Extended Data Figs. 2 and 3 and Methods for more details.

in the outcome window throughout the USA are generally not significant at the 5% level (Fig. 3a). Furthermore, the averages across the entire outcome window for the difference in the change in deaths per million persons (ATT = 0.180; 95% CI = −1.256–1.927, P = 0.827, BF = 0.037) and cases per million persons (ATT = −8.221, 95% CI = −47.524–32.47, P = 0.667, BF = 0.038) were not significant (Supplementary Table 2), in keeping with the day-by-day results.

Because the number of voters present at the polls (and, in particular, the number of voters in a tightly congested area) might affect the risk of transmission, we did additional analyses stratified on the turnout rate. We observed that no level of turnout yielded a positive and significant increase for either outcome compared with the matched control units (Fig. 3b). Separately, we conducted additional analyses that stratified on the population density, primary date, census region, 2016 Trump vote share, the number of days between the first reported case and the primary date, the cumulative case rate on the primary day and the cumulative death rate on the primary day for the treated counties (Supplementary Figs. 4–17), and consistently found no clear evidence for an effect of the primary elections on COVID-19 in the subsequent period.

Second, we estimated the effect of the 5 January 2021 US Senate run-off elections in Georgia. This election is notable both because it (1) occurred during the third peak of COVID-19 in the USA, when the rate of new cases was at the highest point during the pandemic; and (2) was extremely competitive and high-profile, attracting attention at the national level[47]. The latter meant that this primary election set a record for voter turnout[48] and the mean in-person turnout rate was higher than that of the primary elections (Supplementary Fig. 18a,b). Yet, despite these differences and consistent with the primary election results, we observed a mixture of non-significant and positive significant effects on a day-by-day basis (Fig. 4a). However, the outcome-window-wide average for the death rate represents an average growth in the death rate on the order of 10 deaths per million persons and is not significant at the 5% level (ATT = 9.197, 95% CI = −8.972–29.541, P = 0.385, BF = 0.141). The case rate outcome average was also not significant (ATT = −69.949 per million persons; 95% CI = −325.136–175.960, P = 0.552, BF = 0.105) (Supplementary Table 2).

Furthermore, models that stratified by in-person voter turnout did not reveal significant differences in the ATTs as the turnout rate increased (Supplementary Figs. 19 and 20). Moreover, we conducted an additional analysis that stratified by the county-level propensity to wear masks; this analysis did not yield significant effects on average across the outcome window, although the effect size diminished by half in counties that had a tendency to report mask-wearing above the median for counties in Georgia (Supplementary Figs. 21 and 22, and Supplementary Table 2). Finally, a similar effect was observed in

https://doi.org/10.1038/s41562-023-01654-1

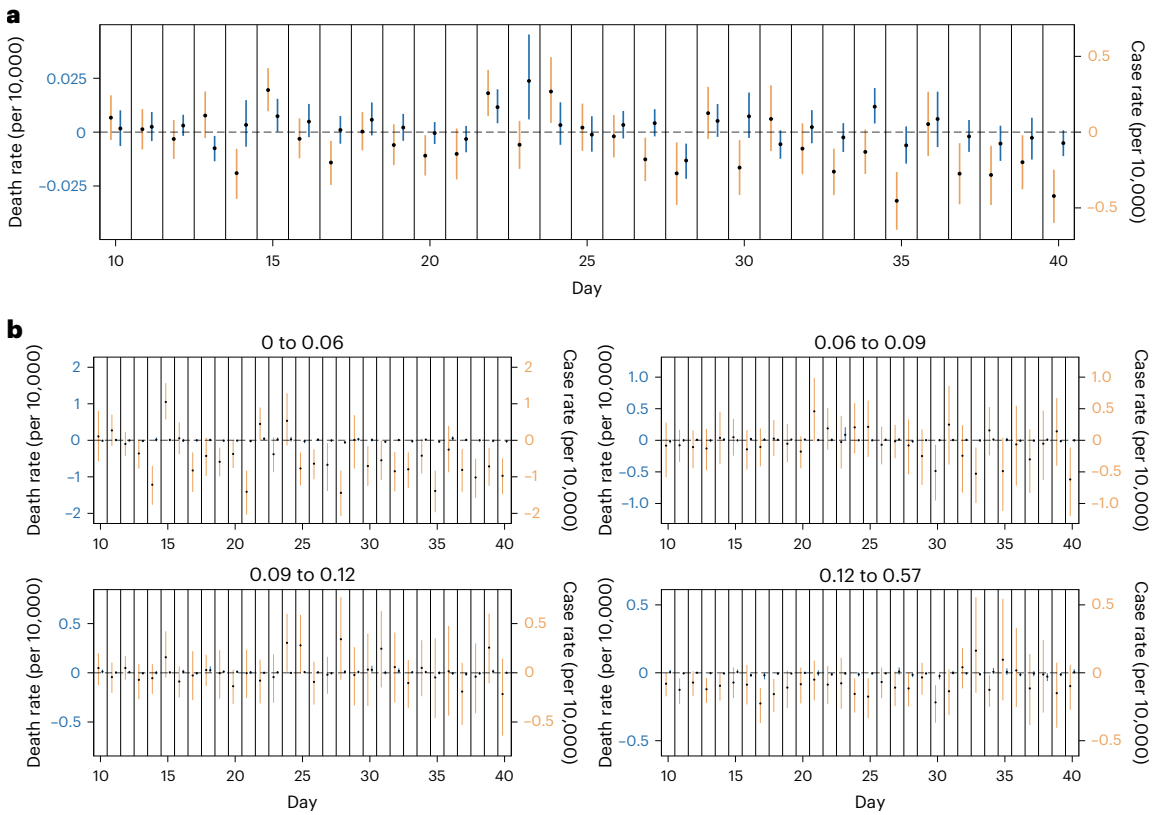


Fig. 3 | Impact of the primary elections on COVID-19 mortality and case rates. Generally, we observed no statistically significant increase on average for treated counties from a period of 10–40 d after a primary election was held. Error bars indicate 95% CIs. **a**, Overall ATT estimates for the primary elections, representing the average difference in the change in death (blue) and case (orange) rates from the day before treatment to 10–40 d after an election. For the death rate outcome, $N = 965$ for days 10–23, 961 for days 24–30 and 956 for days 31–40. For the case rate outcome, $N = 1,057$, 1,056 and 1,051 for the same ranges in the outcome window. N represents the number of treated units. The average numbers of matched units are 4,284 and 4,878 for the death and case rates, respectively, over the whole period. **b**, ATT estimates stratified by the in-person voter turnout rate (defined as the number of in-person voters out of the total county population). The strata represent quantiles of the distribution of turnout rates over the counties with elections. Since turnout data were not available for CO, IA, KY, MA, MI, MN, MS, MO, MT, NV, NY, OK, PA, RI, SD, UT and WV, these states were estimated as a separate category (see Methods). For the death rate, for each turnout bin, $N = 132$, 141, 121 and 88 on days 10–23; 131, 141, 121 and 88 on days 24–30; and 130, 140, 121 and 88 on days 31–40. N represents the number of treated counties on a day. For the case rate, $N = 567$, 664, 646 and 426 for days 10–16; 566, 664, 646 and 426 for days 17–23; 566, 662, 646 and 426 for days 24–30; and 557, 656, 640 and 424 for days 31–40. The average numbers of matched units over the period are 470, 563, 548 and 416 for the death rate and 154, 156, 142 and 90 for the case rate in respective strata. Covariate balance scores for the results in **a** and **b** are, on average over the matching window, within the threshold of 0.1, indicating sufficient similarity between the treated and matched counties for the estimated ATTs (see Extended Data Fig. 4 and Supplementary Figs. 38–42).

an analysis that stratified by above or below 50% vote share for Donald Trump in the 2016 election (Supplementary Figs. 23 and 24).

Third, we assessed the effect of the 2 November 2021 gubernatorial elections in New Jersey and Virginia. These elections had in-person turnout rates that were over twice that of the primary elections (Supplementary Fig. 18c,d), and similar to the election in Georgia, were prominent on the national political stage and set records for levels of turnout[49,50]. In addition, these elections were held during the surge of the Delta variant in the USA and occurred one month before a new set of policy responses to mitigate the spread of the Omicron variant. The elections were also held during a period of concern about 'pandemic fatigue' (drops in adherence to social distancing and masking protocols as frustration with pandemic restrictions in the public was rising)[51–53]. While we could not directly match on the county-level vaccination status in these late elections, our included demographic characteristics are strongly associated with the vaccination levels in the USA[54,55]. Here again, we observed no significant evidence for an effect of the elections on the mortality rate or case rate after the elections in both states combined, with an outcome-window-wide average ATT for the death rate of −15.927 per million persons (95% CI = −29.19–4.298, $P = 0.013$, BF = 2.472) and non-significant average for the case rate (ATT = −220.749 per million persons, 95% CI = −482.639–135.296, $P = 0.139$, BF = 0.296) (Fig. 4b and Supplementary Table 2).

Fourth, we estimated the effect of Donald Trump's political rallies held over the course of 2020. These rallies varied in attendance and were held outdoors in all but three cases (Tulsa, 20 June; Phoenix, 23 June; Henderson, 13 September). The rallies themselves were most commonly held at airports. Estimates for the rally sizes range from hundreds to thousands[7], although no precise or consistent data are available. We assumed that individuals were willing to attend a rally from up to three counties away from the rally location and consequently applied a model that accounts for spillover effects onto neighbouring counties. Specifically, we defined four levels of exposure to the treatment: (1) direct treatment, where a county holds a rally, (2) first degree exposure, where a county borders one that holds a rally, (3) second degree, where a county is one county away from one that holds a rally and (4) third degree, where a county is two counties away from one that holds a rally. In each case, we found no positive and significant set of estimates over the outcome window 10–40 d after a rally for each level of exposure to a rally, with non-significant outcome-window-wide averages for the death rate (ATT = −0.333 per million persons, 95% CI = −6.157–5.07, $P = 0.983$, BF = 0.144) and case rate (ATT = 10.636 per

https://doi.org/10.1038/s41562-023-01654-1

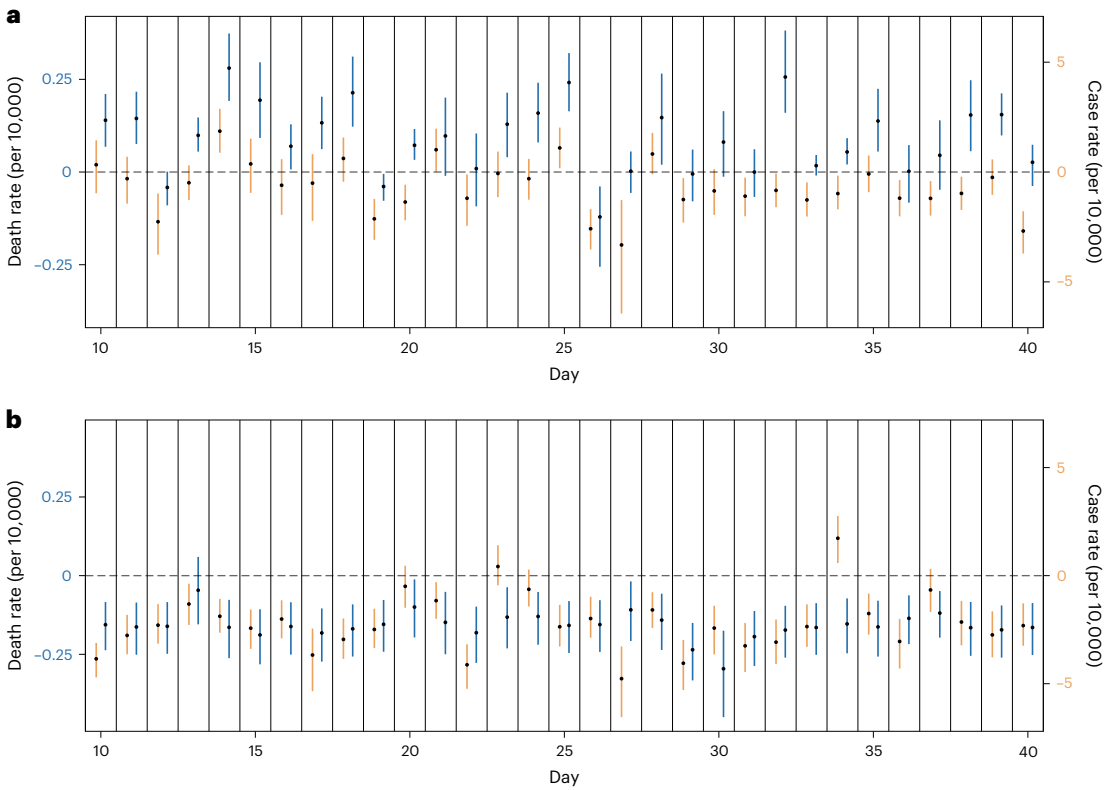


Fig. 4 | Impact of the GA special election and the NJ and VA gubernatorial elections combined on COVID-19 mortality and case rates. a, Overall ATT estimates for the GA election, representing the average difference in the change in death (blue) and case (orange) rates from the day before treatment to 10–40 d after the election. While there is some evidence for a very modest increase in risk of local mortality in the follow-up window, we did not find a statistically significant average effect over this period (Supplementary Table 2); furthermore, we found no evidence for a similar general increase in the case rates. For the death rate, $N = 137$ for each day in the outcome window, where N represents the number of treated counties on a day in the outcome window. For the case rate, $N = 152$. There are 566 and 701 matched counties for the death and case rates, respectively. **b**, Overall ATT estimates for the NJ and VA gubernatorial elections, representing the average difference in the change in death rates from the day before treatment to 10–40 d after the election. There is some evidence for a very modest decline in risk of local mortality and in the case rates and case prevalence on most days in the follow-up window, but we did not find a statistically significant overall average effect over this whole period (Supplementary Table 2). For the death rate, $N = 125$ for each day in the outcome window, where N represents the number of treated counties on a day in the outcome window. For the case rate, $N = 141$. There are 528 and 649 matched counties for the death and case rates, respectively. Covariate balance scores for the results in **a** and **b** are, on average over the matching window, within the threshold of 0.1, indicating sufficient similarity between the treated and matched counties for the estimated ATTs (see Extended Data Fig. 6 and Supplementary Figs. 43–46). Error bars indicate 95% CIs.

million persons, 95% CI = −169.233–213.967, $P = 0.984$, BF = 0.137) for direct treatment (Fig. 5a and Supplementary Table 2). We also examined indirect exposures in neighbouring counties and found no evidence of an association for death and case rates (Fig. 5b). In addition, we stratified by Trump's share of the vote; we supposed that counties that had a majority vote for Trump would be more likely to contain rally attendees and possess a greater likelihood of rally-induced transmission. In each stratum, we found no positive and significant pattern over the outcome window (see Supplementary Figs. 25–28). We note that it is plausible that individuals were infected but did not seek testing after attending Trump rallies, which further underscores death rate as a more reliable metric of interest in this context.

Fifth, we estimated the effect of the BLM protests and found no evidence for an effect of the protests on the mortality rate, with non-significant outcome-window-wise averages for death rate (ATT = −0.442 per million persons, 95% CI = −2.833–1.836, $P = 0.688$, BF = 0.058) and case rate (ATT = 31.380 per million persons, 95% CI = −12.475–85.095, $P = 0.116$, BF = 0.018) (Fig. 6a and Supplementary Table 2). In addition, we stratified by the total crowd size estimate for all protests on a given day in a county, and found no evidence for an effect of the protests across the strata, even for the largest protests (ranging from ~3,000 to 30,000 persons) (Fig. 6b). We also found no clear and significant effects in a model that stratified treated observations on the basis of the number of recent protests (totalling above 800 persons) that occurred within the 3 weeks before the focal protest (see Supplementary Figs. 29 and 30). See Supplementary Information Sections 1 and 2 for more details and for further robustness checks. Extended Data Figs. 2–9 present the estimates before refinement, and with and without calipers for the death rate models.

Political gatherings and local COVID-19 R_t

In addition to analysing the effect of these five event types on the case and mortality rates as discussed above, we also applied our approach to estimates from an independently developed epidemiological model of the transmissibility of SARS-CoV-2, which reconstructed the course of the virus from the observed deaths and cases, and from epidemiologically informed assumptions about the dynamics of the virus (see Methods for further details)[31]. We examined the impact of each of the five event types on the county-level estimated effective reproduction number (R_t). In each situation, we observed no statistically significant impact of large-scale political events on the difference in the change in R_t from the day before an event to between 0 and 20 d afterwards between treated and matched counties (Fig. 7). Note especially the lack of estimated impact on the day of an event, or in the immediate period within 10 d of the event. Furthermore, the effects are close to zero in magnitude in the initial few days, with small confidence intervals, which

https://doi.org/10.1038/s41562-023-01654-1

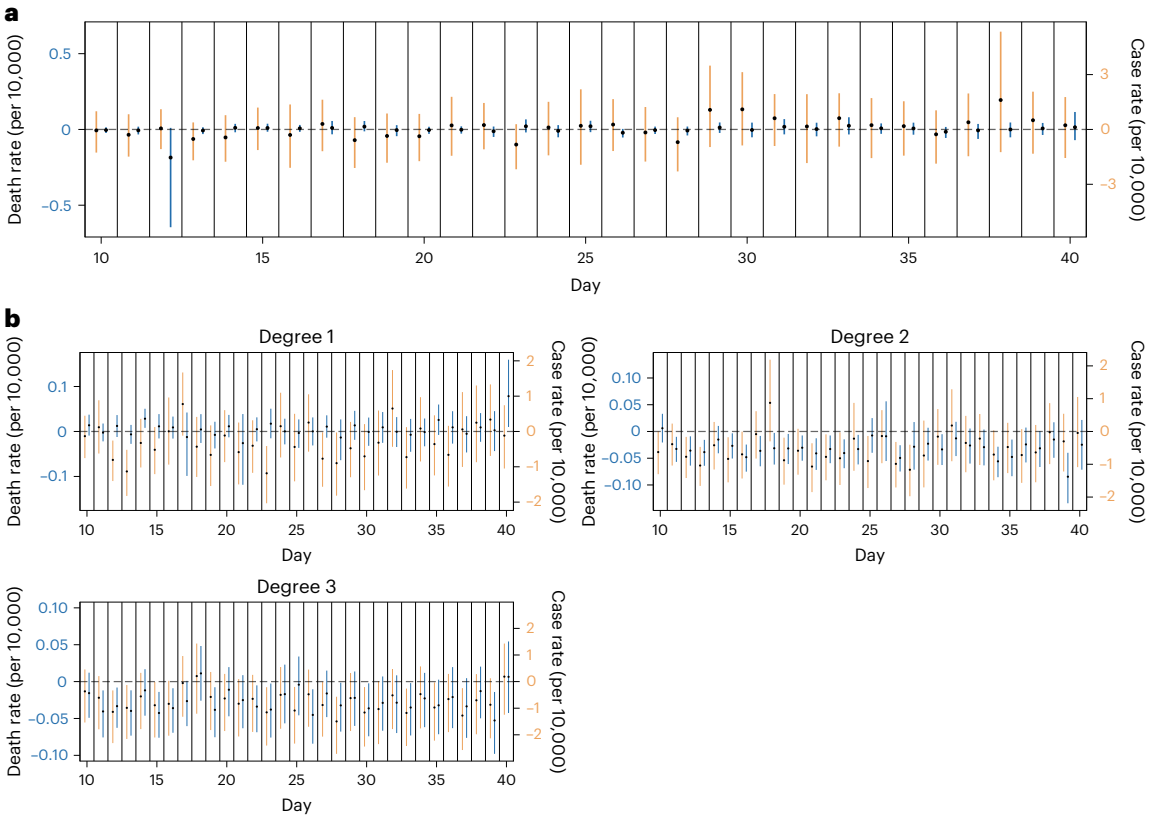


Fig. 5 | Impact of Donald Trump's rallies on COVID-19 mortality and case rates. Generally, we observed no statistically significant increase on average for treated counties from a period of 10–40 d after a political rally was held. Error bars indicate 95% CIs. **a**, Overall ATT estimates for Donald Trump's rallies, representing the average difference in the change in death (blue) and case (orange) rates from the day before treatment to 10–40 d after a rally. For the death rate, $N = 60$ for day 10, 59 for days 11–13 and 58 for days 14–40. For the case rate, $N = 64$ over the whole period. Here, N is the number of treated counties. The average numbers of matched units are 252 and 303 for the death and case rates, respectively. **b**, ATT estimates for counties stratified by indirect exposure to the treatment. Degree 1 represents counties adjacent to a county that held a rally, degree 2 represents a county one county away from one that held a rally, and degree 3 represents a county two counties away from one that held a rally. For the death rate for degrees 1, 2 and 3, $N = 366$, 720 and 1,069 for day 10; 365, 719 and 1,064 for days 11–14; 364, 718 and 1,060 for day 15; 363, 718 and 1,058 for day 16; 363, 718 and 1,056 for day 17; 363, 715 and 1,056 for days 18–19; 363, 715 and 1,054 for day 20; 363, 714 and 1,053 for days 21–22; 363, 713 and 1,053 for days 23–25; 363, 711 and 1,053 for days 26–30; 363, 710 and 1,052 for days 31–38; 363, 709 and 1,049 for day 39; and 363, 708 and 1,047 for day 40. N represents the number of treated counties. Over the whole period, the average numbers of matched counties are 1,679, 3,291 and 4,820 for degrees 1 to 3, respectively. For the case rates for degrees 1, 2 and 3, $N = 384$, 760 and 1,110 for day 10; 383, 760 and 1,105 for days 11–12; 383, 759 and 1,104 for day 13; 382, 759 and 1,104 for day 14; 382, 758 and 1,104 for days 15–16; 381, 758 and 1,102 for days 17–18; 381, 757 and 1,102 for days 19–20; 381, 757 and 1,101 for days 21–22; 380, 757 and 1,101 for days 23–29; 380, 756 and 1,101 for day 30; 380, 755 and 1,101 for days 31–35; 380, 753 and 1,101 for days 36–38; 380, 751 and 1,100 for day 39; and 380, 750 and 1,098 for day 40. Over the whole period, the average numbers of matched counties are 1,840, 3,634 and 5,262 for degrees 1 to 3, respectively. Covariate balance scores for the results in **a** and **b** are, on average over the matching window, within the threshold of 0.1, indicating sufficient similarity between the treated and matched counties for the estimated ATTs (see Extended Data Figs. 7 and 8, and Supplementary Figs. 47 and 48).

further increases our confidence in the absence of an effect of the political events we have studied on the transmissibility of SARS-CoV-2.

Political gatherings and local mobility

Finally, we applied our method to examine the impact of the political gatherings on our three mobility metrics (at fitness and recreation centres, full-service restaurants and grocers) known to be associated with the spread of SARS-CoV-2[28,41]. In each case, we found no significant impact of these political events on visits to these location types over an outcome window of 0–20 d after an event. That is, there were neither statistically significant changes in overall visits to these location types on the day of the events themselves nor a discernible impact on mobility patterns subsequent to the events (Fig. 8).

Discussion

Across a broad array of mass gatherings, including elections, rallies and protests that took place throughout 2020 and 2021, we find no substantial evidence for a material deflection of the local course of the pandemic for a period ranging from 10 to 40 d afterwards. Our analysis here is 'ecological'[56] in the sense that we examine whether political activities affected the course of the pandemic in a county, not whether, at the 'individual' level, going to the polls or participating in a rally or protest affects an individual's risk of contracting the infection.

There is no fixed intrinsic connection between mass gatherings of a specifically political nature and the spread of SARS-CoV-2 per se. Our focus on political gatherings was driven partly by their particular importance and partly by the fact that it was possible to systematically assemble a comprehensive database of all such gatherings, with precise timings (in a way that would not have been as feasible had we focused on other sorts of gatherings, such as sporting events, musical concerts and so on).

Furthermore, the impact of the gatherings on the epidemic, especially in the case of elections, hangs on the particular policies and behaviours found at individual voting sites, for instance, whether masks were used by poll workers and voters, whether most of the time was spent outside and whether individuals were physically distanced

https://doi.org/10.1038/s41562-023-01654-1

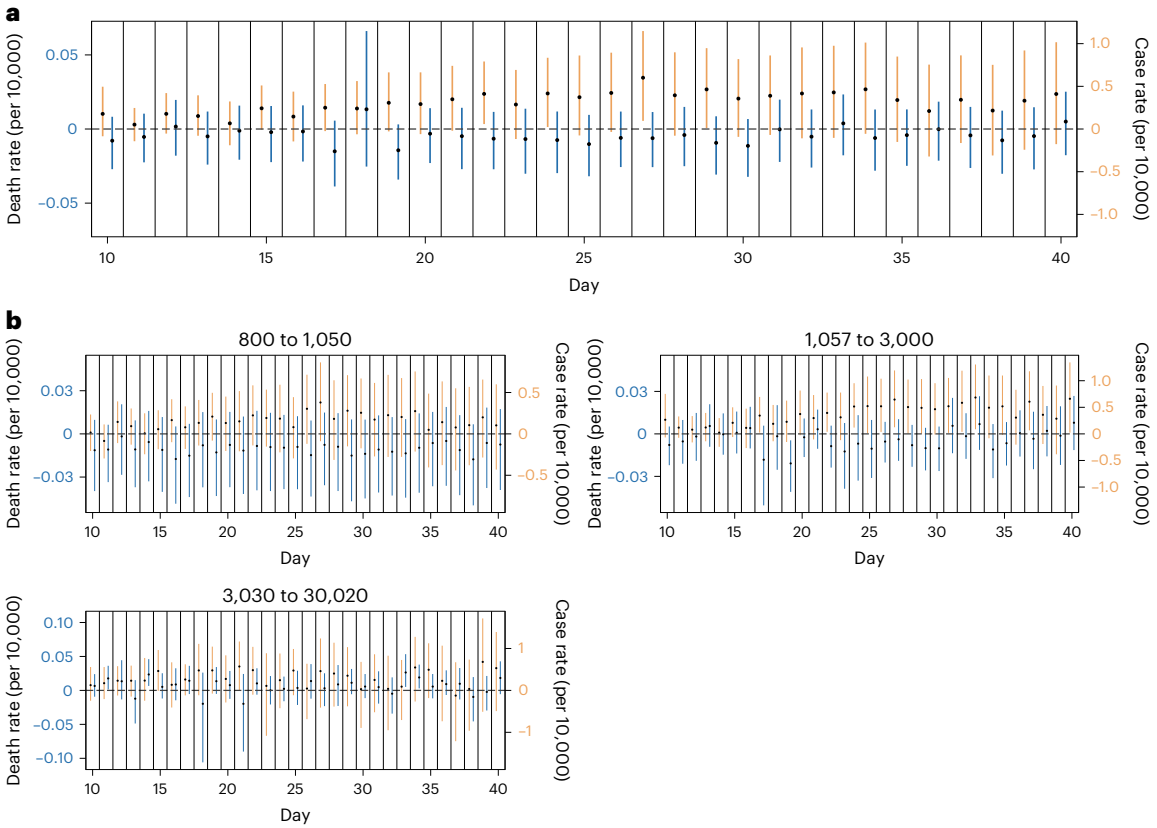


Fig. 6 | Impact of the BLM protests on COVID-19 mortality. Generally, we observed no statistically significant increase for treated counties from a period of 10–40 d after a protest was held. Error bars indicate 95% CIs. **a**, Overall ATT estimates for the protests, representing the average difference in the change in death (blue) and case (orange) rates from the day before treatment to 10–40 d after a protest. For the death rate, $N = 402, 385, 377, 369, 361, 357, 354, 352$ and 350 for days 10–18, 349 for days 19–24, 348 for day 25, 347 for days 26–29, 346 for day 30, 345 for day 31, 344 for days 32–34, 343 for days 35–37 and 342 for days 38–40. N represents the number of treated counties. The average number of matched counties over the period is 1,346. For the case rate, $N = 448, 441, 437, 432, 430, 424, 422$ and 420 for days 10–17, 419 for days 18–21, 418 for days 22–24, 417 for days 25–37 and 416 for days 38–40. The average number of matched counties is 1,764. **b**, ATT estimates stratified by protest size (defined as the total crowd size estimate within the treated county on that day). The strata represent quantiles of the distribution of estimated crowd sizes. For the death rate for each protest size bin, $N = 150, 181$ and 48 for day 10; 145, 172 and 44 for day 11; 142, 168 and 43 for day 12; 139, 167 and 41 for day 13; 137, 164 and 39 for day 14; 134, 161 and 39 for day 15; 132, 160 and 39 for day 16; 132, 159 and 38 for day 17; 132, 157 and 38 for day 18; 132, 156 and 38 for days 19–23; 131, 156 and 38 for day 24; 131, 156 and 37 for day 25; 131, 155 and 37 for days 26–29; 131, 154 and 37 for day 30; 130, 154 and 37 for days 31–32; 129, 154 and 37 for days 33–34; 128, 154 and 37 for days 35–37; and 128, 153 and 37 for days 38–40. N represents the number of treated counties. The average numbers of matched counties are 533, 545 and 130 for the 3 strata, respectively. For the case rate, $N = 170, 210$ and 66 for day 10; 170, 204 and 64 for day 11; 170, 201 and 62 for day 12; 168, 199 and 61 for day 13; 168, 198 and 60 for day 14; 165, 196 and 60 for day 15; 165, 195 and 60 for day 16; 164, 194 and 60 for day 17; 164, 194 and 59 for days 18–21; 164, 194 and 58 for days 22–24; 164, 194 and 57 for days 25–37; and 164, 193 and 57 for days 38–40. The average numbers of matched counties are 699, 778 and 220 for the 3 strata, respectively. Covariate balance scores for the results in **a** and **b** are, on average over the matching window, within the threshold of 0.1, indicating sufficient similarity between the treated and matched counties for the estimated ATTs (see Extended Data Fig. 9 and Supplementary Figs. 49–53).

or quiet while waiting in line. Voting is indeed not ordinarily an intimate activity, and the prevalence of the virus in a given time and place also likely plays a role. Super-spreader events do happen, especially with interpersonal contact[57]. For instance, a single wedding in Maine in August 2020 was linked to 270 cases and eight deaths[58], enough to fundamentally alter the trajectory of COVID-19 in a state that previously had relatively few deaths and cases. However, many of these super-spreading events during the COVID-19 pandemic have been in familial settings associated with close contact and a lack of precautions, or have involved sustained indoor exposure[57,59]. Hence, voting or other political activities may present a level of risk similar to shopping at a grocery store or waiting in line outside a restaurant for a take-out order. In fact, on many days during the outcome window for the NJ and VA gubernatorial elections, we even observed significant and 'negative' estimates, which may reflect the fact that voting could actually be a safer alternative to usual activity patterns then prevalent (which might have been more likely to involve direct, close interpersonal contact).

While we observed a lack of consistent evidence for a spike in mortality associated with the primaries, the GA special election, or the NJ and VA gubernatorial elections, it is important to note several caveats with respect to the safety of voting in general. Most of the primary elections were held under fair-weather conditions, enabling individuals to wait in long lines outside, at lower risk of transmission; however, this was less true for the GA special election (which took place in January) or the NJ and VA elections (in November). It remains possible that the effect could be moderated by the tendency to wear masks or socially distance at the polls. We expect that mass gatherings in close conditions without social distancing, mask use or vaccination with boosters may pose a more serious risk of transmission, especially in cases of high turnout where voting takes place in crammed and enclosed spaces.

Troublingly, while policy arguments about mass gatherings have been politicized during the COVID-19 pandemic, the discussion has largely taken place in the absence of a comprehensive assessment of the public health risks of collective political activities. With respect to the US elections, three studies have been conducted on the Wisconsin

https://doi.org/10.1038/s41562-023-01654-1

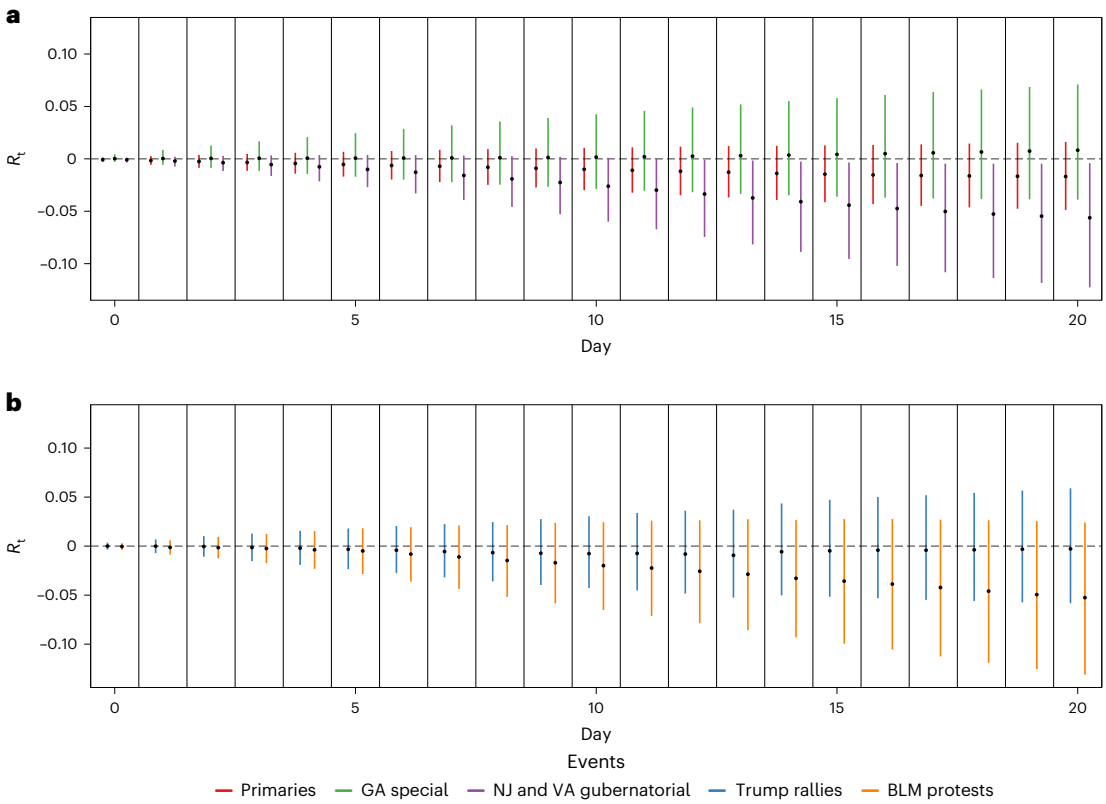


Fig. 7 | Impact of large-scale political events on COVID-19 R_t. The results represent the average difference in the change in the transmissibility of the virus from the day before treatment to the day of a protest, up to 20 d after the protest. We found no evidence of a significant increase in transmissibility stemming from these events, starting on the day of the event (day 0) to within 20 d of the event. Error bars indicate 95% CIs. **a**, Overall ATT estimates for the primary, GA special, and NJ and VA gubernatorial elections. **b**, Overall ATT estimates for the BLM protests and Donald Trump's political rallies. For the primaries, $N = 971$ for days 0–13 and 969 for days 14–20. N represents the number of treated counties. The numbers of matched counties are 4,473 and 4,466 for the same periods; for the GA special, $N = 87$, with 211 matched counties. For the gubernatorial elections, $N = 83$, with 230 matched counties. For the Trump rallies, $N = 65$, with an average of 308 matched counties. For the BLM protests, $N = 466$ for day 0, 460 for day 1, 458 for day 2, 456 for day 3, 543 for day 4, 452 for days 5–6, 451 for day 7, 449 for days 8–10, 448 for day 11, 447 for days 12–14 and 446 for days 15–20. The average number of matched units is 1,902. Covariate balance scores for the results in **a** and **b** are, on average over the matching window, within the threshold of 0.1, indicating sufficient similarity between the treated and matched counties for the estimated ATTs (see Supplementary Figs. 54–59).

primary election, and these reached conflicting conclusions regarding the impact of in-person voting on the subsequent course of the COVID-19 epidemic, although two found no subsequent increase[60–62]. Separately, two studies have been conducted on the effect of the BLM protests, each finding no clear effect[30,63]. Our comprehensive study of the whole USA covering a 2year period thus sheds light on this topic.

Smaller-scale studies in foreign countries also paint a mixed picture, although most agree with our findings. Examining the first and second rounds of the 2020 French municipal elections, one study found a sizeable subsequent increase in hospitalization rates after the first-round election (in March) but did not find any increase associated with the second-round elections (in June) when masks became obligatory[64]. Two further studies[65,66] respectively using a Bayesian mixture model and a sigmoidal mixed effects model found no effect of the same French elections for either round. A study of the October 2020 elections in the Czech Republic found a subsequent increase in cases[67]. Yet another study, of elections in Italy in September 2020, found that a subsequent increase in the spread of COVID-19 was driven mainly by campaign activities leading up to the elections, which allowed indoor events at public and private places with no limit on the size of the event (in contrast to size restrictions imposed on other event types in Italy at that time), rather than by infections at the polls themselves[68]. In still another context, researchers found no effect of voting on the spread of COVID-19 in the 2020 Brazilian municipal elections, which occurred in the context of a broad public health campaign and restrictions, including a mandate that voters wear masks at the polls[69]. Finally, an additional study found no increased mortality among 163,000 candidates in the 2020 French town hall elections[70].

Generally, the studies that have found significant effects of political events are those that either do not account for the nonlinear contagion process[61,71], or do so only indirectly[64,67]. By contrast, studies that directly account for the contagion process, such as the present one, have found no significant effects[60,65,66]. Our study thus adds to the literature by directly accounting for the nonlinear contagion process in a generalized difference-in-differences framework, and avoiding some key difficulties associated with some epidemiological models[72] and many existing causal inference methods[34].

In contrast to voting, the Trump rallies and BLM protests were almost entirely held outdoors in the open air, despite large crowd sizes. Outdoor transmission of COVID-19 is very rare[73]. The Sturgis Motorcycle Rally is often cited as an example of outdoor transmission[74]; however, it was marked by indoor events in addition to the outdoor gathering, with cases stemming from the rally linked to restaurants and workplaces[75]. Consistent with this finding, we found no evidence of an uptick in visits to restaurants or other locations known to yield a high risk of transmission associated with the protests or rallies we studied (Fig. 8 and Supplementary Figs. 31 and 32). Furthermore, it is also possible that the average age of people attending rallies and protests is younger and their underlying health better than the electorate as a whole.

https://doi.org/10.1038/s41562-023-01654-1

Another possible explanation of the observed results with respect to the mass gatherings we have studied is that people in an area with recent political activity, or the individuals participating in such activity themselves, may make compensatory adjustments to lower their 'other' risks of contracting the virus. For instance, elderly people in an area of a rally or protest may decide to stay inside and avoid contact with others for a few weeks after the rally (even if they themselves did not participate)[76]. On this account, the fact that the mass gatherings studied here did not seem to deflect the overall COVID-19 mortality curves in the short run could be driven by compensation for the risks of the events themselves. However, we found no evidence for mobility decreases following the large-scale political activities (at locations known to be at high risk for transmission) that would support such an account (Fig. 8 and Supplementary Figs. 33 and 34).

Our study is potentially limited by unobserved factors, as with any statistical approach to observational data, including previous analyses of COVID-19 risks and responses[28,30,38,41,46,60,61,72,74]. Such inferential problems may indeed be worse for a process that is intrinsically contagious, which is one of the reasons we also sought to match index counties with multiple control counties that had similar previous epidemic trajectories. Furthermore, while the results are consistent across time and event types, the confidence intervals encompass a range of possibilities. Finally, these results, focused on a particular class of mass gatherings, are at odds with some epidemiological studies of NPIs (which have included mass gathering bans as a possible stratagem)[46,77–79]. On the other hand, our results agree with two previous studies of some of the BLM protests[63], and two previous studies of the Wisconsin primary (that is, a single state), as noted earlier[60,62]. Furthermore, they are in line with a more complex picture that is being established as researchers grapple with the set of existing conclusions drawn from traditional statistical methods[34].

We urge careful interpretation of these results, which may be specific to these events in the US context. The full range of potentially confounding factors should be considered in the context of mass political gatherings across the globe, where leaders are deciding whether to hold in-person elections or restrict the right to protest, especially in places characterized by a scarcity of vaccines.

Furthermore, we emphasize the heterogeneous nature of these events themselves, which are associated with differential risk exposure for different groups; for instance, older voters who are at high risk for COVID-19 tended to vote at the highest rates[80], while younger individuals were more likely to attend BLM protests. Such effects may also potentially bias some of our outcome metrics; for instance, those who attend Trump rallies may be less likely to seek testing, and younger people who attend BLM protests may be less likely to die from contracting COVID-19. Hence, while the mortality data are generally taken to be a better reflection of the underlying state of the COVID-19 epidemic[81], we performed a comprehensive analysis that considered case rates and transmissibility as supplements to mortality rate, and observed no evidence for an effect on any of the outcomes.

Another limitation pertains to the potential relationship between the epidemic and turnout at several of the events we studied; it is possible that individuals decided not to turn out to these events for fear of exposure. Our approach helps avoid potential bias due to this potential relationship by matching counties on the basis of characteristics of the epidemic (the death or case rate or transmissibility in the period leading up to the event, along with the date of the first reported case). Nonetheless, because we did not have access to comprehensive data on the list of attendees for all events and were not able to adjust the

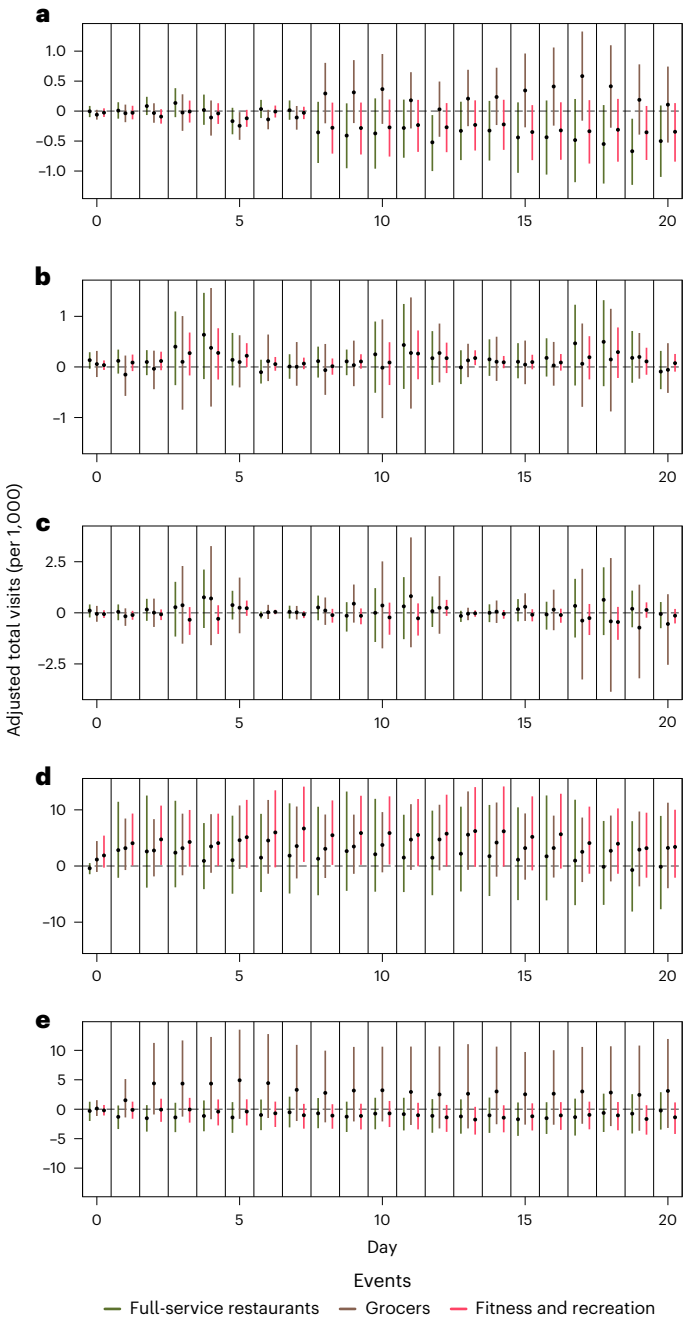


Fig. 8 | Impact of large-scale political events on COVID-19 mobility. The results represent the average difference in the change in the number of adjusted visits to a given location type from the day before treatment to the day of an event, up to 20 d after the event. For each event, we examined visits to full-service restaurants, grocers, and fitness and recreation centres. We found no evidence for a significant change in mobility stemming from these events within 20 d of an event. Error bars indicate 95% CIs. **a**–**e**, Overall ATT estimates for the primaries (**a**), GA special election (**b**), NJ and VA gubernatorial elections (**c**), Donald Trump's political rallies (**d**) and the BLM protests (**e**). For the primaries, the number of treated units was $N = 971$ for days 0–13, 967 for days 14–20, 960 for days 21–31, 957 for days 32–34, 953 for days 35–38 and 952 for days 39–40. There is an average of 4,271 matched units over the whole period. For the GA special, $N = 83$ for each day, with 252 matched units. For the gubernatorial elections, $N = 83$, with 284 matched counties. For the Trump rallies, $N = 32$ for days 0–4, 31 for days 5–16, 30 for days 17–31 and 29 for days 32–40. Over the whole period, there is an average of 284 matched units. For the BLM protests, $N = 188$, 181, 177, 176, 173, 171, 167 and 164 for days 0–7, 163 for days 8–10, 162 for days 11–26, 161 for day 27, 160 for day 28, 159 for days 29–34, 157 for days 35–36 and 156 for days 37–40. The average number of matched counties is 519 over the whole period. Covariate balance scores for the results in **a**–**e** are within the threshold of 0.1, on average over the matching window, indicating sufficient similarity between the treated and matched counties for the estimated ATTs. The balance scores are identical to those for the corresponding models for the death rates.

death counts on the basis of exact attendance, we instead adjusted death counts on the basis of the county population. In addition, we did not include extensive data on health system capacities and constraints. However, given the structural inequities in our society, these are likely to be strongly correlated with the extensive set of matching covariates we did include[82,83], and hence are likely similar between our treated counties and their matched controls. Thus, while this exclusion of data on health system capacities and constraints represents another limitation of our study, it would seem less likely to bias our results. While low attendance in counties with higher rates of COVID-19 might have reduced transmission, we see this as a possible effect modifier, which does not bias the results. Furthermore, we note that these events were large in scale (Extended Data Fig. 10), often involving thousands of individuals at key phases of the pandemic; our analyses that look directly at variation in election turnout and protest size did not yield significant effects.

Policy responses to voting and political expression should be mindful of their potentially disparate impact on otherwise already disenfranchised groups if policymakers make a potentially spurious effort to reduce risks of COVID-19[84–87]. We also note that the decisions individuals make and should make for themselves about whether to attend a political gathering are distinct from the overall collective impact. Still, we consistently find no evidence for a material increase in COVID-19 mortality or other epidemic parameters as a result of mass gatherings for political expression across a wide range of settings in the USA. This lack of evidence stands in contrast to the sizeable effects estimated for some other non-pharmaceutical interventions; as a result, our study may help to underscore the importance of a policy focus on NPIs with demonstrated effects[88]. The complex manifestations of this protean respiratory pathogen should not be underestimated.

Methods

Data

Johns Hopkins COVID-19 mortality (and case) data. County-level COVID-19 deaths were collected from a repository maintained by the Johns Hopkins Coronavirus Resource Center (https://coronavirus.jhu.edu/)[81], compiled from local and state health agencies. This dataset is a standard in the literature and is considered the authoritative source of COVID-19 mortality and case data[89]. Our main analyses used the survey of COVID-19 death counts, while supplementary analyses employed positive test results as an alternative outcome for each estimated model. Our main analyses focused on COVID-19 death counts, rather than positive test results, as our outcome measure due to: (1) limited and changing testing capacity, (2) bias with respect to COVID-19 testing implementation and (3) the lack of available test data at the county level for many areas of the country. With respect to (2), individuals were only likely to be tested if they showed symptoms, or had the means to get tested; consequently, the tests cannot be viewed as a representative sample of the population.

In our analysis, we assumed that entries before the first reported data were zero counts, and we removed observations not attributable to any specific US county. In addition, we retained observations where the cumulative death counts declined from one day to the next; decreases may occur due to reporting error, such as a correction of an overcount. In these cases, we marked the change from one day to the next as a zero.

COVID-19 transmissibility estimates. We drew on estimates of the effective reproduction number of the virus over time at the county level from a COVID-19 epidemiological modelling project. This model uses a Bayesian framework that accounts for reporting delays and variability in case ascertainment to reconstruct the course of the epidemic from observed death and case counts[31,32,33]. Model estimates are freely available online from the project website, or reproducible with 'covidestim', an R package[90]. We drew the daily county mean estimates from the model as an additional outcome to supplement our analyses of case and mortality rates.

The model (as implemented in ref. 31) was fitted using the data on new cases and deaths to model the unobserved transmission rate, from the Johns Hopkins data (described above). The model used four compartments: a- or pre-symptomatic; symptomatic and not severe; symptomatic and severe; and death. Uninfected individuals become asymptomatic and either recover or move to the symptomatic state from which they either recover or move to the severe symptomatic state. Individuals in the severe state either recover or die. Individuals were assumed to be infected only once. The R_t trend was modelled as a log transformed cubic b-spline with knots every 10 d, and with penalties on the first and second differences on the spline parameters to ensure a parsimonious fit. The lags between infection to symptoms, symptomatic cases to severe ones and severe infections, and reporting delays were taken to be fixed, and according to a Gamma distribution, consistent with the literature[46]. Model results were found to be robust under a range of assumptions[31]. Case ascertainment varied substantially over the course of the epidemic; to account for this, the fraction diagnosed in the model was allowed to vary over time. Further details, including a full list of model parameters, are available in ref. 31.

SafeGraph. We applied data from SafeGraph, which collects and aggregates data from a range of mobile phone applications to track population mobility. SafeGraph data, including point-of-interest location data, have been used extensively to study population mobility during COVID-19[41,91,92] and were found to be demographically representative[93]. We used the Places-Patterns dataset, which tracks hourly visits to a set of Core Places, which are points of interest around the USA that contain addresses and North American Industry Classification System (NAICS) codes. We included three business categories previously found to be associated with possible transmission risk: grocers, gyms and fitness clubs, and full-service restaurants[28,41]. We aggregated the place data to the daily and county level for each business category. Each mobility variable represents the total daily visits to locations, classified into the above categories, by users in the SafeGraph database, weighted by the ratio of the number of SafeGraph users resident in that county to the total county population to adjust for the sample size in each county[94]. In addition, we removed all parent points of interest from the dataset to ensure that no location was counted twice in the dataset (for example, counting a mall and the stores therein).

BLM protests. The US Crisis Monitor[95] is a curated dataset, produced in collaboration with the Armed Conflict Location and Event Data Project and the Princeton University Bridging Divides Initiative, that contains comprehensive data on political protests, violence, demonstrations and related events in the USA for 2020. We systematically considered all protest-related events that occurred in the USA between 26 May 2020 and 10 October 2020. This period captures the series of protests that initially started during the aftermath of the murder of George Floyd and extended across the country[95]. We aggregated the individual protest events to the county–day level. Our data included all substantial events that occurred in this period, including some protest events not related to BLM. We observed 658 distinct county-level events, where 94% were classified as part of the BLM movement, 7% as 'Blue Lives Matter' pro-police, pro-Donald Trump and 3% as including direct references to COVID-19. The remaining protests involved various other political causes.

The dataset included accounts from newspaper and traditional media sources on the location, date and size of protest events. Protest crowd size estimates ranged from specific numbers (for example, 300), numeric ranges (for example, 300–500), to descriptive estimates (for example, hundreds, thousands, 'almost a dozen'). For general range descriptions, we took the geometric mean over the possible range of values it could encompass (for example, 'hundreds'

becomes $\sqrt{100 \times 1,000}$). These approximations serve our goal to separate protests by orders of magnitude. Nonetheless, crowd size estimates from newspaper sources have been found to produce generally accurate estimates of crowd sizes[96]. In our data, multiple protests frequently occurred within a county on a single day; in this case, we summed the estimated protest counts over each individual event. Rather than directly relying on protest size estimates, our main analysis applied an event-size threshold to define a binary treatment variable. We only considered a county as 'treated' on a given day if the aggregated protest count was above 800 persons. On the other hand, only counties with protests below 400 persons were eligible to be matched control units to a treated county. The distribution of protest sizes was highly skewed (Supplementary Fig. 35).

US elections. Our elections data consist of the rolling schedule of primary dates. During the 2020 primary election cycle, several states rescheduled (Supplementary Table 1) or cancelled their in-person primary elections. Our primary treatment used the rescheduled election dates, and only elections that were held in-person were considered as treated observations. Counties in which mail-in-only primaries were held were eligible to serve as matched control units. We analysed elections in 1,173 counties drawn from 21 states where applicable elections were held across the mainland USA; we drew on the full set of 3,108 counties in the 48 continental states in mainland USA (including the District of Columbia) as our source for match units. The 2020 primary election cycle ran from 3 February (Iowa (IA)) to 11 August (Connecticut (CT)). In our main analysis, we only considered primaries that occurred after 14 March; we believe that any primary election held before 1 March is not a valid treatment, given the low likelihood of COVID-19 community spread before March in most locations.

We further estimated a model that stratifies by voter turnout. We personally compiled in-person voter turnout directly from state and county election agencies where available. Turnout data were not available for Colorado (CO), IA, Kentucky (KY), Massachusetts (MA), Michigan (MI), Minnesota (MN), Mississippi (MS), Missouri (MO), Montana (MT), Nevada (NV), New York (NY), Oklahoma (OK), Pennsylvania (PA), Rhode Island (RI), South Dakota (SD), Utah (UT) and West Virginia (WV); consequently, these states were estimated as a separate category in models that apply turnout data (these results are below, but not in the main manuscript).

In addition, we drew on elections data for the 5 January 2021 US Senate special election in GA from the US Elections Project[97], including the in-person turnout in each county, the overall turnout rate and the number of mail-in ballots. Turnout data for the NJ gubernatorial election were drawn from the Department of State of New Jersey's Department of Elections[98], and turnout data for the VA gubernatorial election were pulled from the Virginia Department of Elections[99]. In each case, the in-person voter turnout rate was calculated as the number of in-person ballots case divided by the county population.

Donald Trump's campaign rallies. We used a list of Trump rallies from 6 June 2020 to 2 November 2020, compiled directly from traditional media sources, for a total of 67 individual rallies across the country. On some days, multiple rallies were held in multiple locations and some locations were repeated across the data over time. In each case, we handled a county as treated on a particular day if a rally occurred. Because it is likely that many rally attendees travelled across county lines from rural more Republican counties to the comparatively urban and central rally locations, we also employed US county adjacency data from the US Census[100] to account for spillover effects to neighbouring counties.

US census. We drew county-level demographic data from the 2014–2018 Five-Year American Community Survey from the US Census Bureau[101]. Specifically, we constructed estimates of the percentage of African American, percentage of Hispanic, log median income, population density (persons per square mile), percentage of those at least 65 years old and Donald Trump's share of the vote in the 2016 general election.

The New York Times masking behaviour survey. The county-level tendency to wear a mask was drawn from a survey conducted by The New York Times[102] between 2 and 14 July 2020, consisting of over 250,000 individual responses weighted by age and gender. Respondents were asked "How often do you wear a mask in public and when you expect to be within six feet of another person?" and given the option to answer 'never', 'rarely', 'sometimes', 'frequently' or 'always'. We constructed the tendency to never or rarely wear a mask as the sum of 'never' and 'rarely' percentages by county. We did not make use of this data for covariate matching in our general analyses since the survey was conducted amid the primary elections. Specifically, we used this data for a supplementary analysis of the GA special election.

Software and code. Analysis was conducted in the Julia programming language[103] (v.1.7.1) with the 'TSCSMethods' Julia package[104] (which we are pleased to release). Additional paper replication materials are available on GitHub as the 'COVIDPoliticalEvents' Julia package. Data cleaning, preparation and additional analysis were carried out with the R programming language[105] (v.4.0.3). The map in Fig. 1 was made with the 'urbnmapr' R package[106].

Generalized difference-in-differences estimation for COVID-19

Modellers of COVID-19 and interventions that affect its spread face a methodological dilemma: whether to apply an epidemiological or causal statistical method. The first includes compartmental models (for example, Susceptible-Exposed-Infectious-Recovered models), which face serious difficulties with identification[72] and reliance upon estimated epidemiological parameters that may change as the pandemic unfolds, increasing the complexity of the process model. In such models, the estimated parameters may identify descriptive quantities, but they may not identify the causal parameters that would undergird an effective policy response.

The second brushes against the limitations of many common statistical approaches. Existing statistical analyses of the impact of NPIs on the spread of COVID-19 apply either interrupted time-series design[38] or difference-in-differences estimation[71]. Both of these approaches rely on the assumption that without the intervention, the spread of COVID-19 could have been predicted from the model fit to the process from the pre-intervention period, ruling out other time-varying factors that could confound the relationship between the interventions and the spread. Furthermore, these approaches assume that the outcome is a linear process, which is often inappropriate in the context of an infectious disease that is characterized by a nonlinear trend. In practical terms, these methods may falsely attribute natural changes in the course of the epidemic to a particular intervention.

Specifically, interrupted time-series methods typically (1) do not control for an underlying trend not caused by the intervention, or rely on specification of its correct functional form[107], such as the underlying contagion process of an infectious disease and (2) simply rule out the possibility of unmeasured factors causing a given effect. Specifically, refs. 37,38 attempted to resolve (1) through fixed-effects models that indirectly account for the contagion process as reduced-form models justified on the assumption that the proportion of susceptible individuals approaches unity. However, we did not find this assumption justified in our setting, given the absence of reliable testing data in the USA (which severely underestimates the number of active cases)[31] and a study time-horizon that ranges from March 2020 to November 2021, over every phase of the pandemic.

On the other hand, difference-in-differences methods do control for an underlying trend not caused by the intervention, by relying on the common trend assumption[108]. In addition, more advanced

difference-in-differences strategies rely on regression adjustment or matching to allow for unobserved time-invariant factors[109–111]. However, traditionally, they do not account for time-varying imbalances, or the existence of a nonlinear data-generating process. The approach here is more in line with recent approaches for difference-in-differences methods in less restrictive settings[109,112]. Instead, both difference-in-differences and interrupted time series applied to COVID-19 often assume that cases are log-linear[30,37,38,61,71,74,113], which induces bias, especially when the mean counts are low and over-dispersed, which holds true in our county-level daily death data. Our data are also strongly zero-inflated, limiting the utility of traditional count models. In addition, the fact that many counties have zero deaths over portions of our study time-horizon precludes the estimation of traditional epidemiological models to assess the impact of our interventions. Practically, these methods induce bias that might lead to a spurious assignment of causal pathways, which might in turn prompt inappropriate policy responses.

To avoid relying on parametric models for the dynamics of COVID-19 and in the wake of the above difficulties, we modified and applied a non-parametric, generalized difference-in-differences estimator with a matching procedure for time-series cross-sectional data to estimate the average effect of the treatment on the treated (ATT). We did so by extending an approach from ref. 35 in a novel way. In particular, we defined different causal parameters that are more meaningful for this setting and implemented a sliding time window to find matches for units that can be considered 'treated' at each timepoint. Our approach may be applied in any setting governed by a nonlinear contagion process, where the 'treatment' occurs in a limited time frame or is repeated over time (time-varying treatment) and its occurrence varies across the population. Furthermore, each of the event types we analysed presents different challenges that require alteration of our estimation procedure: in the case of the Trump rallies, we assumed the presence of spillover effects, and in the case of the BLM rallies, we had many units with multiple treatment-like protests within a narrow time window.

Our difference-in-differences approach measures the average effect of an event on the observations where the event actually occurred (ATT), by comparing the outcome change over time across matched units in different treatment groups. The difference-in-differences strategy requires that in the absence of the event, the outcome would have followed the same trend in the treated and control arms. Nonetheless, our approach makes less stringent assumptions than many traditional approaches for the estimation of causal effects with panel data. As opposed to the standard difference-in-differences estimator, this method relies on a parallel trend assumption only after conditioning on both baseline and time-varying covariates before the intervention, including the pre-treatment outcome history. In addition, covariate adjustment is conducted using a matching procedure. This approach also allows for time-varying treatments that can occur multiple times over the observed window. Our approach is presumably novel in applying such a method to epidemic data and could constitute a new general strategy for research on the effects of single or intermittent events on a contagion.

In our analysis, a unit is the entire time series of a US county (or county equivalent). A unit is treated on a specific day if it holds an event of interest (for example, an election, rally or protest) on that day. For such a given treated observation, we matched on that county's characteristics for 30 d before treatment, up to 1 d before treatment. We contrasted the treated and matched unit's outcome for each day in an outcome window defined as 10–40 d after an event, which is the epidemiologically informed period in which we would expect an effect on the mortality rate from an event where contagion may occur. Given the underlying contagion process, it was not reasonable to assume that the effect of any intervention on the epidemic reverses itself, but rather alters the epidemic trajectory in perpetuity. However, we allowed units with previous treatments to serve as eligible controls for a treatment event, under specific conditions. For each day that we estimated on the period from 10 to 40 d, matched counties must have a similar treatment history to the treated unit during the pre-treatment crossover window and have no treatment during the post-treatment crossover window (detailed below). Consequently, we separately matched units to a treated observation for each day in the outcome window.

We did not match on treatment history before the pre-treatment crossover period since we considered deaths after day 40 to be endogenous to the contagion process. The outcome window period represents the first wave of deaths that would be expected to occur due to first-hand infection at the event. Since these first-wave deaths are caused by an exogenous shock (for example, an election, rally or protest), they cannot be predicted from the observed death rate alone. However, deaths that occurred beyond 40 d are likely second-wave deaths, which are predictable from the observed death rate. Thus, we considered deaths past 40 d to be endogenous to the contagion process, allowing similarity matching based on the death rate to account for the effects of treatment events that occurred before the pre-treatment crossover period.

In addition, we restricted our outcome window to the expected first wave of deaths because estimation over a longer outcome window risks confounding by events that may occur and thereby separately alter the epidemic trajectory in our treated or control counties. Consequently, there arises a risk of misattributing the effect of such events to the treatment considered. We thus face a trade-off between capturing the long-run effects of an intervention (that is, an election, rally or protest) and applying a shorter outcome window to avoid post-treatment confounding. Hence, we restricted our procedure to the period in which 90% of the first-wave deaths would occur due to transmission on the day of treatment.

Thus, we took the treatments to influence the outcome from 10 to 40 d after a given treatment, corresponding to the 5th and 95th percentiles of the distribution of times from infection to death. The distribution of times from infection until death comes from empirical estimates in the epidemiological literature and is modelled as the sum of two gamma distributions (Fig. 1):

$$\pi \sim \text{Gamma}(5.1, 0.86) + \text{Gamma}(17.8, 0.45) \quad (1)$$

Identification of the ATT

We defined $X_{i,t}$ to be the treatment for unit i *on day* t, valued $X_{i,t} = 1$ when the unit is treated on day t (it holds an event of interest), and 0 otherwise. On the basis of this outcome window, we assumed that the potential outcome of each unit i at day τ depends on the event history in a carry-over window, that is, $F_{\max}$ days to $F_{\min}$ days before τ, with $F_{\max} > F_{\min}$. That is, the outcome, $Y_{i,\tau}$, at τ depends on interventions $X_{i,q}$, with $q \in \{\tau - F_{\max}, \ldots 0, \ldots, \tau - F_{\min}\}$, that is, $Y_{i,t}\left(\{X_{i,q}\}_{q=\tau-F_{\max}}^{\tau-F_{\min}}\right)$.

We then contrasted the average potential outcome at time $t+F$, for a treatment at time t, with the average potential outcome at time $t+F$ without treatment at time t, where $F \in \{F_{\min}, F_{\min}+1, F_{\min}+2, \ldots, F_{\max}\}$. In particular, our interest lies in the following causal quantity, $\delta(F; F_{\min}, F_{\max})$, the ATT:

$$\begin{aligned}\delta(F; F_{\min}, F_{\max}) = E\Big[Y_{i,t+F}\Big(X_{i,t}=1, \{X_{i,t-l_1}\}_{l_1=1}^{F_{\max}-F}, \{X_{i,t+l_2}\}_{l_2=1}^{F-F_{\min}}\Big) \\ - Y_{i,t+F}\Big(X_{i,t}=0, \{X_{i,t-l_1}\}_{l_1=1}^{F_{\max}-F}, \{X_{i,t+l_2}\}_{l_2=1}^{F-F_{\min}}=0\Big) \,|\, X_{i,t}=1\Big]\end{aligned} \quad (2)$$

with $F \in \{F_{\min}, F_{\min}+1, F_{\min}+2, \ldots, F_{\max}\}$. The parameter $\delta(F; F_{\min}, F_{\max})$ is the causal effect of a treatment event on day t on the death rate F days after t, compared to not having a treatment event on day t and in the post-treatment carry-over window affecting day $t+F$, when the sequence of treatment events are the same in the pre-treatment carry-over window, marginalized over the distribution of the treatment history after the treatment event at day t conditional on having the

treatment at time t. E refers to the expectation, and l_1 and l_2 iterate over the pre- and post-treatment carry-over windows, respectively.

We relied on two key assumptions:

1. Non-interference between units. We assumed that there was no interference between counties over the study period. We assumed that an event that took place in a single county did not affect the death rate in other counties, at least over the time frame we considered.
2. Parallel trends, conditional on treatment, covariate and outcome histories for F_{max} days before the day of the elections. This assumption relaxes sequential ignorability, the condition that the treatment assignment is unconfounded, conditional on the covariate and outcome history up to $t - F_{max}$. Instead, we allowed for the presence of unobserved confounding variables. The parallel trend assumption holds conditionally on the treatment, covariate and outcome histories for F_{max} days before treatment administration.

The ATT was estimated over a range of days $F \in \{ F_{min}, F_{min} + 1, F_{min} + 2, \ldots, F_{max} \}$, corresponding to the outcome (carry-over) window over which we expected treatment effects to occur. We chose $F_{min} = 10$ and $F_{max} = 40$ as the 5th and 95th percentiles, respectively, of the distribution of the number of days from infection to death. This window captures the first wave of deaths expected from an exogenous shock (that is, an election, rally or protest).

County matching procedure

The matching procedure contained three steps: restriction on the treatment history of each county to determine the set of allowable matches, distance matching based on chosen covariates and matching refinement from the set of allowable matches.

Selection of eligible matches based on treatment history. The definition of our ATT implies that for treatment day t and outcome day $t + F$, treated units (that had the treatment on that day) should be compared with control units that had no treatment events on day t or during the post-treatment crossover window for $t + F$, that is, $(t, t + F - F_{min})$. During the pre-treatment crossover window $(t + F - F_{max}, t - 1)$, units were matched on their treatment histories. In the case of a single treatment per unit (the elections), this means that eligible matches must have zero treatments over the pre-treatment crossover period, as in the treated unit. In the multiple treatment scenarios (the rallies and the protests), we performed an inexact match on the basis of the similarity of number of treatments that occurred in the pre-treatment crossover period between a potential match and the treated unit. In particular, we discretized the number of treatments in the pre-treatment crossover period and allowed a match when the potential match and the treated unit fell into the same treatment category. For the BLM protests, we defined the categories as: both have 0 previous treatments, both have 1–2 previous treatments, both have 3–9 previous treatments and both have 10 or more previous treatments. Separately, for the rallies, categories were defined similarly, but with additional accounting for the level of treatment exposure to neighbouring counties (see 'Selection of eligible matches based on treatment history with spillover effects').

Matching refinement via distance matching based on chosen covariates. After applying these conditions on the treatment history, we conducted distance matching on the basis of a set of county-level time-invariant demographic characteristics, time-varying social distancing mobility data, and features of the epidemic at the county level relevant to behaviour and the spread of COVID-19.

Generally, the length of the matching period for time-varying covariates is subject to a bias–variance trade-off: choosing a longer period on which to match decreases the possible bias of the model, insofar as treated and matched counties are increasingly similar with respect to the overall COVID-19 trend. However, a longer lag period leads to greater difficulty in matching, increasing the variance of the estimator.

After selecting eligible control counties, we looked for counties that are similar to the treated county with respect to demographic and epidemiological characteristics from 30 d before, up to 1 d before the primary election. We picked this fixed period to ensure that we always matched on a 30 d period, regardless of F. We applied the fixed window since the covariate histories additionally reflect general information about a county's tendency to have interpersonal contact (for the mobility covariates) and information about where the county falls with respect to its epidemic trajectory (in the case of the cumulative death rate). We did not use the pre-treatment crossover period as the window for covariate matching because this shortens as F increases. Using the pre-treatment crossover window as the sole basis for covariate matching would imply that unit similarity over the fixed-length period decreases in importance as F increases.

In the matching procedure, we included movement specifically at full-service restaurants, grocers, and fitness and recreation facilities because they have been identified as high-risk locations for COVID-19 transmission[28,41]. We did not explicitly include indicators for NPIs (for example, stay-at-home orders or their retractions) or other large-scale events. We note that there is no comprehensive data source on closures or their lifting at the county level. Moreover, these orders were haphazard and not consistently enforced. That is, a stay-at-home order in one county may have been highly effective but ignored in another[114] (these trends probably relate to political and demographic characteristics such as racial composition, socio-economic status and political views, for which we did adjust). Furthermore, adherence to the NPIs is crucial to their role in altering transmission. Consequently, it has been argued that mobility patterns are more directly linked to transmissibility than the NPIs and may be used in place of the NPIs themselves[91] or to evaluate their effectiveness[115]. Mobility is frequently used in models of COVID-19 transmissibility[116]. In line with this, numerous studies have found a strong link between transmissibility, NPI measures and mobility patterns[42,117–120]. Mobility is not a perfect measure of transmissibility and has the strongest relationship to transmissibility during the earlier phases of the pandemic when NPIs were being aggressively implemented[33,92,121]. Consequently, in light of these considerations, we believe we have effectively adjusted (to some extent) for confounders that could arise from NPIs or other large events, especially in tandem with our other covariates. Specifically, we focused on broad mobility categories that have been demonstrated to be linked to COVID-19 transmission: visits to restaurants, fitness and recreation facilities, and grocery stores[41], which we believe are relevant to our primary analysis of large-scale gatherings in general.

In addition, we used the following county-level socio-demographic variables: (log) median income, percentage of people 65 years old and above, percent African American, percent Hispanic, population density and Donald Trump's share of the vote in the 2016 presidential election. These demographic characteristics may be relevant to how people behave, which in turn may affect the spread of the virus: for example, counties in which Trump had a higher share of the vote may be less likely to practice social distancing; counties with a higher percentage of African Americans or that are lower in terms of median income may have a greater share of essential workers[11,122]. In addition, to adjust for the features of the epidemic, we adjusted for the date of the first reported infection in that county, as well as one time-varying feature, namely, the cumulative county-level death rate. Note that matching on the cumulative death rate implicitly contains information about the whole history of the epidemic trend, adjusting not only for the daily deaths over the lag period, but also for the total deaths accrued over the reporting window. We think that this adjustment better matches counties in terms of their COVID-19 trends. Finally, we additionally matched on the county-level propensity to wear a mask for the GA

special election, and the NJ and VA gubernatorial elections, which took place after this survey data were collected, to further adjust attitudes and behaviour regarding COVID-19.

Under matching refinement, we selected at most five units that are the most similar to the treated unit during the matching period in terms of Mahalanobis distance:

$$S_{i,t}(i') = \frac{1}{L}\sum_{l=1}^{L}\sqrt{\left(V_{i,t-l-V_{i',t-l}}\right)^{T}\Sigma_{i,t-l}^{-1}\left(V_{i,t-l}-V_{i',t-l}\right)} \quad (3)$$

where the beginning of the matching window before a treatment is given by $L = 30$. We note that $i' \in M_{i,t}$ is a unit in the set of potential matches to a treated unit i, $V_{i,t'}$ is the set of time-varying covariates for which we adjusted in the pre-treatment lag period and $\Sigma_{i,t'}$ is the diagonal of the sample covariance matrix. The distance was computed for each potential match for each day in the pre-treatment lag period, which was then averaged over that period (from $t-30$ up to $t-1$).

Subsequently, we calculated the covariate balance as the standardized mean difference between the treated and control units across all counties used in the analysis:

$$B_{i,t}(j,l) = \frac{V_{i,t-l,j} - \sum_{i' \in M_{i,t}} w_{i,t}^{i'} V_{i',t-l,j}}{\sqrt{\frac{1}{N_1-1}\sum_{t'=1}^{N}\sum_{t'=L+1}^{T} D_{i',t'}\left(V_{i',t'-l,j} - \bar{V}_{t'-l,j}\right)^2}} \quad (4)$$

where N_1 is the total number of treated observations and $V_{i,t,j}$ represents the jth covariate of unit i on day t. The quantity $w_{i,t}^{i'}$ represents the weight given to match i' for control unit i at t. Note that $w = 0$ for each i' that is not a match for the control unit. Finally, $\bar{B}(j,l)$ is simply the average balance over the treated units:

$$\bar{B}(j,l) = \frac{1}{N_1}\sum_{i=1}^{N}\sum_{t=L+1}^{T} D_{i,t}B_{i,t}(j,l) \quad (5)$$

where N is the total number of units, T is the total number of periods and $D_{i,t}$ indicates whether unit i is treated at t.

Caliper matching. In addition, we imposed a caliper to ensure sufficient covariate balance for estimation. We calculated the standardized Euclidean distance between each pair of matching covariates given by the treated and control units over the matching period and took the matching period average. We then applied a caliper threshold to a chosen set of covariates, rejecting a potential match unit if its distance is above the threshold on a chosen covariate. Caliper values are specified in terms of standardized Euclidean distances between the values of specific matching covariates.

In particular, calipers were applied individually to specific matching covariates as needed, to ensure that covariate (im)balance remains below 0.1 on average over the matching period. In addition, all models included a caliper of 0.25 on the cumulative death rate (or cumulative case rate as appropriate to the model), since this is the crucial covariate to ensure that the epidemic trajectories are similar between the treated units and their matches. The value of 0.25 is a standard recommendation in the matching literature[123–125]; however, we checked the sensitivity to stricter calipers in Supplementary Information Section 2.5. This procedure removes potential matches since they may fail to satisfy the distance requirement. When a treated observation retained zero matches, it was removed entirely from estimation. Thus, a caliper might meaningfully reduce the number of treated units. For each set of model results, we present the number of treated observations that remained after the application of a caliper, out of the initial number (see below for details). In cases where we were unable to find a set of matches to the treated units with a sufficiently low balance score (below 0.1 on average over the pre-treatment matching period), we dropped the mobility covariates and matched on the remaining variables. Plots that display the balance scores for each of the models in the main text are available in Supplementary Information Section 3. The mobility models had the same balance scores as the respective death rate models for each event.

This was the case for ten models: the Trump vote share, in-person turnout, and mask behaviour stratified models and the overall model for the GA election; the NJ and VA gubernatorial election model, the model for Donald Trump's rallies accounting for spillover effects; the model for Donald Trump's rallies stratified by Donald Trump's share of the vote in 2016 and accounting for spillover effects; the number-of-recent-protests stratified model, the protest-size stratified model, and the overall model for the protests; and the transmissibility model for the GA special election. In addition, the date of the first case was dropped from the NJ and VA election analyses since these occurred in a much later phase of the pandemic; also, visits to full-service restaurants was the only included mobility variable in this model since we were unable to find sufficient balance with the full set of matching covariates.

Difference-in-difference estimator and Bayes factors

For N units and T total days in a dataset, our difference-in-difference estimator of the ATT is given by

$$\hat{\delta}(F) = \frac{1}{\sum_{i=1}^{N}\sum_{t=1}^{T} D_{i,t}}\sum_{i=1}^{N}\sum_{t=1}^{T} D_{i,t}\left\{\left(Y_{i,t+F} - Y_{i,t-1}\right) - \sum_{i' \in M_{i,t}} w_{i,t}^{i'}\left(Y_{i',t+F} - Y_{i',t-1}\right)\right\} \quad (6)$$

which captures the average of the average difference in the change in the daily death rate ($Y_{i,t}$), from the day before treatment ($t-1$) to a post-treatment timepoint ($t+F$) between a treated unit and its control counties. $D_{i,t} = 1$ defines a unit i treated at t with at least one matched control unit $i' \in M_{i,t}$. $w_{i,t}^{i'} = \frac{1}{|M_{i,t}|}$ is the match weight, with the number of matches in the set of matches to the treated observation denoted by $|M_{i,t}|$. We constructed confidence intervals with a weighted block-bootstrap procedure appropriate for panel data with a fixed number of matches, which accounts for the number of times that a unit is used as a match and samples units rather than observations[126] (details presented below). All confidence levels presented are at the 95% significance level.

The overall estimates presented for each of the main models are simple arithmetic means of the individual estimates for each day in the post-treatment outcome window. The confidence intervals for the overall estimates were taken by aggregating the bootstrap distribution across the individual estimates (see 'Bootstrap procedure' below). In addition, for key models, we present Bayes factors to provide more information about the relative evidence for the null and alternative hypotheses in our analyses. In contrast to P values, the Bayes factor is able to quantify evidence in favour of the null hypothesis[127]. In each case, we calculated the Bayes factor for the alternative hypothesis (B_{10}), where increasingly small numbers below 1 indicate stronger evidence for the null hypotheses over the alternative. We calculated the Bayes factors using Gaussian quadrature, using the estimated mean and variance from the bootstrap distribution for the ATT (see below) to calculate a statistic for a classical t-test[128,129].

Spillover effects

Voting is a fundamentally 'local' activity and individuals typically do not travel outside of their neighbourhood to cast a ballot, let alone across county lines. Furthermore, mobility in the spring and summer of 2020 was below usual levels, in the wake of stay-at-home orders and an increase in the number of individuals working from home. With respect to the BLM protests, we note that larger protests are more likely to occur in highly urban areas, with a larger concentration of young individuals whose politics tend to align with the protests themselves.

In contrast, surrounding suburban areas tend to be more Republican leaning (Supplementary Fig. 36) and less likely to have high concentrations of protesters. Thus, we assumed that the protests are also an activity that is local in nature. Consequently, we believe that it is unlikely for an election or protest held in a single county to yield an increase in infections in other counties over the study follow-up period we considered.

While assumption (1) (above in 'Identification of the ATT') is clearly not justified in general for the spread of COVID-19, we assumed that it holds for the election and protest events over the time horizon we used. Specifically, we assumed that it does not hold in the case of Trump rallies, as individuals may have travelled from surrounding counties to attend the rallies. This is especially plausible given that the rallies themselves tended to be in more Democratic-leaning areas than surrounding counties (Supplementary Fig. 36). We allowed for the presence of spillover effects up to three counties away from the location of a Trump rally, on the logic that individuals may travel up to several hours to attend a rally.

We generalized the ATT to account for interference from neighbouring counties (Supplementary Fig. 37). We denoted by $G_{i,t}$ the exposure variable representing whether unit i at time t is directly exposed to treatment or indirectly exposed to the treatment of neighbouring counties. Specifically, when $G_{i,t} = 1$, a unit i is directly treated at t, having an event within its county borders at time t, regardless of the treatment of neighbouring counties; when $G_{i,t} = 2$, county i is not directly treated but it is the first degree neighbour of a county that holds an event at t (geographically adjacent); when $G_{i,t} = 3$, county i is not directly treated and does not have adjacent counties treated, but it is the second degree neighbour of a county that holds an event at t (one county away); when $G_{i,t} = 4$, county i is the third degree neighbour of a county that holds an event at t (two counties away). $G_{i,t} = 0$ when a county i at day t is not directly treated and is at least three counties away from one that holds an event at t.

Consequently, we took $G_{i,t}$ as the treatment exposure for unit i *on day* t, valued $G_{i,t} > 0$ when the unit is treated on day t (it holds or is near an event of interest), and 0 otherwise. On the basis of this outcome window, we assumed that the potential outcome of each unit i at day τ depends on the event history in a carry-over window, that is, $F_{\max}$ days to $F_{\min}$ days before T, with $F_{\max} > F_{\min}$. That is, the outcome at τ depends on interventions $G_{i,q}$, with $q \in \{\tau - F_{\max}, \ldots 0, \ldots, \tau - F_{\min}\}$, that is, $Y_{i,\tau}\left(\{G_{i,q}\}_{q=\tau-F_{\max}}^{\tau-F_{\min}}\right)$.

In this setting, the ATT becomes:

$$\delta(g, F; F_{\min}, F_{\max}) = E[\, Y_{i,t+F}(G_{i,t} = g, \{G_{i,t-l_1}\}_{l_1=1}^{F_{\max}-F}, \{G_{i,t+l_2}\}_{l_2=1}^{F-F_{\min}}) - Y_{i,t+F}(G_{i,t} = 0, \{G_{i,t-l_1}\}_{l_1=1}^{F_{\max}-F}, \{G_{i,t+l_2}\}_{l_2=1}^{F-F_{\min}} = 0) | G_{i,t} = g\,] \quad (7)$$

where $g \in \{1, 2, 3, 4\}$ represents the treatment exposure category. Here, 1 represents direct exposure to treatment; values from 2 to 4 represent indirect exposure via neighbouring treatments.

While estimation of the total effect included spillover from treated units onto untreated units[130], we ignored this possibility, simply averaging within different exposures among treated units. This assumption is plausible if we assume that many attendees travel from Republican-leaning counties to the rally locations and that attendees do not typically attend more than one rally on a single day. After matching, four sets of ATTs were estimated, one set for each level of exposure.

Selection of eligible matches based on treatment history with spillover effects. As above ('Selection of eligible matches based on treatment history'), we discretized the number of treatments in the pre-treatment crossover period and allowed a match when the potential match and the treated unit fell into the same treatment category. For the rallies, categories were defined similarly, but with additional accounting for the level of treatment exposure to neighbouring counties. For treated units, eligible matches were selected on the basis of treatment history criteria that accounted for exposure. Specifically, we allowed a match if it is in the same category as the treatment for each exposure type, where the categories are defined as having zero versus one or more previous treatments within 30 d of the treatment.

Stratification

We additionally stratified our estimation on the basis of properties of the treated observations (for example, population density, in-person turnout rate). These models applied the standard matching procedure but separated treated observations and their matches into different strata; the ATTs in each stratum were estimated separately.

Bootstrap procedure

The entire time series of a unit was resampled to account for within-unit dependence over time, in a weighted block-bootstrap design. This method accounts for matching with a fixed number of units and the use of a matched unit for more than one treated observation, through the calculation of observation weights[35]:

$$W_{i,t}^{*} = \sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i',t'} v_{i,t}^{i',t'} \quad (8)$$

where

$$v_{i,t}^{i',t'} = \begin{cases} 1 \text{ if } (i,t) = (i', t' + F) \\ -1 \text{ if } (i,t) = (i', t' - 1) \\ -w_{i,t'}^{i} \text{ if } i \in M_{i',t'} \,\&\, (i', t' + F) \\ w_{i,t'}^{i} \text{ if } i \in M_{i',t'} \,\&\, (i', t' - 1) \\ 0, \text{ otherwise} \end{cases}$$

An observation's weight was assigned to 1 if it is a treated observation and fell within the outcome window. It was valued at −1 in the case of the period immediately before the focal treatment. An observation that is not treated was assigned a non-zero weight $w_{i,t}^{i'} = \frac{1}{|M_{i,t}^{*}|}$ when $i' \in M_{i,t}^{*}$, that is, when it fell into the outcome window, or the period preceding the treatment for a focal treatment observation. Reference [35] notes that the estimator may alternatively be calculated using the observation weights as:

$$\frac{\sum_{i=1}^{N}\sum_{t=1}^{T} W_{i,t}^{*} Y_{i,t}}{\sum_{i=1}^{N}\sum_{t=1}^{T} D_{i,t}} \quad (9)$$

Consequently, in the bootstrap, we sampled units and then calculated this quantity for each bootstrap sample. We took our confidence intervals from percentiles 2.5 and 97.5 of the distribution formed from 10,000 resamples.

Reporting summary

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

Data availability

The data on COVID-19 case and death counts are available from the Johns Hopkins Coronavirus Resource Center (https://coronavirus.jhu.edu/data). US Census data are also publicly available (https://www.census.gov/programs-surveys/acs). The mobility tracking data are available from SafeGraph, Inc. and are freely available to academic researchers (https://www.safegraph.com/products/places). The elections turnout data are available directly from state governmental

election agencies. The protest event data are available as the US Crisis Monitor dataset from the Armed Conflict Location & Event Data Project (https://acleddata.com/special-projects/us-crisis-monitor/). County-level masking data are available from The New York Times GitHub repository (https://raw.githubusercontent.com/nytimes/covid-19-data/master/mask-use/mask-use-by-county.csv).

Code availability

Data cleaning and preparation was carried out with the R programming language[106]. All analysis was conducted in the Julia programming language[105] with the 'TSCSMethods' Julia package[104] (https://github.com/human-nature-lab/TSCSMethods.jl). In addition, paper replication materials are available on GitHub as the 'COVIDPoliticalEvents' Julia package (https://github.com/human-nature-lab/COVIDPoliticalEvents.jl).

References

1. Christakis, N. A. *Apollo's Arrow: The Profound and Enduring Impact of Coronavirus on the Way We Live* (Little, Brown Spark, 2020).
2. Enten, H. Trump is creating an untraditional partisan divide on vote by mail. *CNN* https://www.cnn.com/2020/07/27/politics/vote-by-mail-partisan-divide-analysis/index.html (2020).
3. Milosh, M., Painter, M., Sonin, K., Dijcke, D. V. & Wright, A. L. Unmasking partisanship: polarization undermines public response to collective risk. *J. Public Econ.* **204**, 104538 (2021).
4. Easley, J. & Miller, M. Pandemic sparks partisan brawl over voting by mail. *The Hill* https://thehill.com/homenews/campaign/491862-pandemic-sparks-partisan-brawl-over-voting-by-mail (2020).
5. Perrett, C. Fauci says 'there's no reason' in-person voting shouldn't be safe with masks and proper social distancing. *Business Insider* https://www.businessinsider.com/fauci-says-in-person-election-with-distancing-masks-is-safe-2020-8 (2020).
6. Reichmann, D. As virus surges, Trump rallies keep packing in thousands. *AP NEWS* https://apnews.com/article/donald-trump-rallies-virus-surges-50e79fabd46472c51ecc1444184082de (2021).
7. Cillizza, C. Donald Trump is (still) totally obsessed with crowd size. *CNN* https://www.cnn.com/2020/10/27/politics/donald-trump-crowds-rallies/index.html (2020).
8. Buchanan, L., Bui, Q. & Patel, J. K. Black Lives Matter may be the largest movement in U.S. history. *The New York Times* (2020).
9. Madrigal, A. C. & Meyer, R. America is giving up on the pandemic. *The Atlantic* https://www.theatlantic.com/science/archive/2020/06/america-giving-up-on-pandemic/612796/ (2020).
10. Palus, S. Public health experts say the pandemic is exactly why protests must continue. *Slate* https://slate.com/technology/2020/06/protests-coronavirus-pandemic-public-health-racism.html (2020).
11. Yancy, C. W. COVID-19 and African Americans. *JAMA* **323**, 1891–1892 (2020).
12. Agarwal, V, Pokharel, K. & Mandhana, N. As Covid-19 surged, Indian teachers died after working elections. *Wall Street Journal* https://www.wsj.com/articles/as-covid-19-surged-indian-teachers-died-after-working-elections-11620558475 (2021).
13. *A Guide to 2021 Latin American Elections* https://www.as-coa.org/content/guide-2021-latin-american-elections (Americas Society/Council of the Americas, 2021).
14. Keating, C. Connecticut House OKs absentee voting expansion despite Republican concerns. *Hartford Courant* https://www.courant.com/2022/03/16/connecticut-house-oks-absentee-voting-expansion-despite-republican-concerns/ (2022).
15. Pasieka, C. How the pandemic might affect voter turnout, and Ford's chances in June. *CBC News* https://www.cbc.ca/news/canada/toronto/ontario-election-turnout-2022-1.6388401 (2022).
16. Kong, K. Seoul's full cafes, Apple store lines show mass testing success. *BNN Bloomberg* https://www.bloomberg.com/news/articles/2020-04-18/seoul-s-full-cafes-apple-store-lines-show-mass-testing-success#xj4y7vzkg (2020).
17. Spinelli, A. *Managing Elections Under the COVID-19 Pandemic: The Republic of Korea's Crucial Test* (International Institute for Democracy and Electoral Assistance, 2020).
18. Robles, F. Cubans denounce 'misery' in biggest protests in decades. *The New York Times* https://www.nytimes.com/2021/07/11/world/americas/cuba-crisis-protests.html (2021).
19. COVID 'superspreader'? India warns against mass farmer protests. *Al Jazeera* https://www.aljazeera.com/economy/2021/5/24/india-warns-farmer-protests-can-be-covid-super-spreader (2021).
20. Balmforth, T. Russia cracks down on home front as its troops invade Ukraine. *Reuters* https://www.reuters.com/world/europe/russia-cracks-down-home-front-its-troops-invade-ukraine-2022-02-24/ (2022).
21. Sonabend, R. et al. Non-pharmaceutical interventions, vaccination, and the SARS-CoV-2 Delta variant in England: a mathematical modelling study. *Lancet* **398**, 1825–1835 (2021).
22. Sharma, M. et al. Understanding the effectiveness of government interventions against the resurgence of COVID-19 in Europe. *Nat. Commun.* **12**, 5820 (2021).
23. Callaway, E. Delta coronavirus variant: scientists brace for impact. *Nature* **595**, 17–18 (2021).
24. Earnest, R. et al. Comparative transmissibility of SARS-CoV-2 variants Delta and Alpha in New England, USA. *Cell Rep. Med.* **3**, 100583 (2022).
25. Kamp, J. & Abbott, B. Delta surge of Covid-19 recedes, leaving winter challenge ahead. *Wall Street Journal* https://www.wsj.com/articles/delta-surge-of-covid-19-recedes-leaving-winter-challenge-ahead-11635672600 (2021).
26. Ferré-Sadurní, L. & Goldstein, J. 1st vaccination in U.S. is given in New York, hard hit in outbreak's first days. *The New York Times* https://www.nytimes.com/2020/12/14/nyregion/coronavirus-vaccine-new-york.html (2020).
27. *COVID-19 Vaccinations in the United States* (CDC, 2022).
28. Spiegel, M. & Tookes, H. Business restrictions and COVID-19 fatalities. *Rev. Financ. Stud.* **34**, 5266–5308 (2021).
29. Sabin, S. 2 in 3 voters at least somewhat concerned about voting in person during coronavirus. *Morning Consult* https://morningconsult.com/2020/03/20/voting-options-poll-coronavirus/ (2020).
30. Dave, D. M., Friedson, A. I., Matsuzawa, K., Sabia, J. J. & Safford, S. *Black Lives Matter Protests and Risk Avoidance: The Case of Civil Unrest During a Pandemic* (NBER, 2020); https://www.nber.org/papers/w27408
31. Chitwood, M. H. et al. Reconstructing the course of the COVID-19 epidemic over 2020 for US states and counties: results of a Bayesian evidence synthesis model. *PLoS Comput. Biol.* **18**, e1010465 (2022).
32. Petrone, M. E. et al. Combining genomic and epidemiological data to compare the transmissibility of SARS-CoV-2 variants Alpha and Iota. *Commun. Biol.* **5**, 439 (2022).
33. Kishore, N. et al. Evaluating the reliability of mobility metrics from aggregated mobile phone data as proxies for SARS-CoV-2 transmission in the USA: a population-based study. *Lancet Digit. Health* **4**, e27–e36 (2022).
34. Weill, J., Stigler, M., Deschenes, O. & Springborn, M. *Researchers' Degrees-of-Flexibility and the Credibility of Difference-in-Differences Estimates: Evidence From the Pandemic Policy Evaluations* (NBER, 2021); http://www.nber.org/papers/w29550.pdf
35. Imai, K., Kim, I. S. & Wang, E. H. Matching methods for causal inference with time-series cross-sectional data. *Am. J. Polit. Sci.* https://doi.org/10.1111/ajps.12685 (2021).

36. Imbens, G. W. & Rubin, D. B. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction* (Cambridge Univ. Press, 2015).
37. Fowler, J. H., Hill, S. J., Levin, R. & Obradovich, N. Stay-at-home orders associate with subsequent decreases in COVID-19 cases and fatalities in the United States. *PLoS ONE* **16**, e0248849 (2021).
38. Hsiang, S. et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature* **584**, 262–267 (2020).
39. Jia, J. S. et al. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* **582**, 389–394 (2020).
40. Crawford, F. W. et al. Impact of close interpersonal contact on COVID-19 incidence: evidence from 1 year of mobile device data. *Sci. Adv.* https://doi.org/10.1126/sciadv.abi5499 (2022).
41. Chang, S. et al. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2021).
42. Badr, H. S. et al. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *Lancet Infect. Dis.* **20**, 1247–1254 (2020).
43. Zhang, J. et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* **368**, 1481–1486 (2020).
44. Kang, Y. et al. Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic. *Sci. Data* **7**, 390 (2020).
45. Verity, R. et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.* **20**, 669–677 (2020).
46. Flaxman, S. et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584**, 257–261 (2020).
47. Phillips, A. What you need to know about the Georgia Senate runoff elections. *Washington Post* https://www.washingtonpost.com/politics/interactive/2020/georgia-senate-runoff-guide/ (2021).
48. McCartney, A. Turnout hits historic highs in contentious Georgia Senate races. *Bloomberg* https://www.bloomberg.com/graphics/2021-georgia-senate-runoff/ (2021).
49. Constantino, A. K. Virginia election sees highest turnout in recent history, fueling Glenn Youngkin's victory. *CNBC* https://www.cnbc.com/2021/11/03/virginia-election-sees-highest-turnout-in-recent-history-fueling-glenn-youngkins-victory.html (2021).
50. Enten, H. 2021 shows Republicans shouldn't fear high voter turnout. *CNN* https://www.cnn.com/2021/11/07/politics/turnout-republicans-analysis/index.html (2021).
51. Petherick, A. et al. *A Worldwide Assessment of COVID-19 Pandemic-Policy Fatigue* (SSRN, 2021); https://papers.ssrn.com/abstract=3774252
52. Crabtree, S. Americans' social distancing habits have tapered since July. *Gallup* https://news.gallup.com/poll/322064/americans-social-distancing-habits-tapered-july.aspx (2020).
53. *Gen Z and the Toll of the Pandemic* (AP-NORC, 2021); https://apnorc.org/projects/gen-z-and-the-toll-of-the-pandemic/
54. *The Red/Blue Divide in COVID-19 Vaccination Rates* (KFF, 2021); https://www.kff.org/policy-watch/the-red-blue-divide-in-covid-19-vaccination-rates/
55. Hughes, M. M. et al. County-level COVID-19 vaccination coverage and social vulnerability — United States, December 14, 2020–March 1, 2021. *MMWR Morb. Mortal. Wkly Rep.* **70**, 431–436 (2021).
56. King, G., Rosen, O. & Tanner, M. A. *Ecological Inference: New Methodological Strategies* (Cambridge Univ. Press, 2004).
57. Shridhar, S. V., Alexander, M. & Christakis, N. A. Characterizing super-spreaders using population-level weighted social networks in rural communities. *Phil. Trans. R. Soc. Math. Phys. Eng. Sci.* **380**, 20210123 (2022).
58. Mahale, P. Multiple COVID-19 outbreaks linked to a wedding reception in rural Maine — August 7–September 14, 2020. *MMWR Morb. Mortal. Wkly Rep.* **69**, 1686–1690 (2020).
59. Hamner, L. High SARS-CoV-2 attack rate following exposure at a choir practice — Skagit County, Washington, March 2020. *MMWR Morb. Mortal. Wkly Rep.* **69**, 606–610 (2020).
60. Leung, K., Wu, J. T., Xu, K. & Wein, L. M. No detectable surge in SARS-CoV-2 transmission attributable to the April 7, 2020 Wisconsin election. *Am. J. Public Health* **110**, 1169–1170 (2020).
61. Cotti, C., Engelhardt, B., Foster, J., Nesson, E. & Niekamp, P. The relationship between in-person voting and COVID-19: evidence from the Wisconsin primary. *Contemp. Econ. Policy* **39**, 760–777 (2021).
62. Berry, A. C., Mulekar, M. S. & Berry, B. B. Increase in daily new COVID-19 cases not seen following the Wisconsin primary election April 2020. *J. Infect. Dis. Epidemiol.* https://doi.org/10.23937/2474-3658/1510148 (2020).
63. Neyman, G. & Dalsey, W. Black Lives Matter protests and COVID-19 cases: relationship in two databases. *J. Public Health* https://doi.org/10.1093/pubmed/fdaa212 (2020).
64. Cassan, G. & Sangnier, M. The impact of 2020 French municipal elections on the spread of COVID-19. *J. Popul. Econ.* **35**, 963–988 (2022).
65. Duchemin, L., Veber, P. & Boussau, B. Bayesian investigation of SARS-CoV-2-related mortality in France. *Peer Community J.* **2**, e6 (2022).
66. Zeitoun, J.-D. et al. Reciprocal association between participation to a national election and the epidemic spread of COVID-19 in France: nationwide observational and dynamic modeling study. *Eur. J. Public Health* https://doi.org/10.1093/eurpub/ckab140 (2021).
67. Palguta, J., Levínský, R. & Škoda, S. Do elections accelerate the COVID-19 pandemic? *J. Popul. Econ.* **35**, 197–240 (2022).
68. Cipullo, D. & Le Moglie, M. To vote, or not to vote? Electoral campaigns and the spread of COVID-19. *Eur. J. Polit. Econ.* **72**, 102118 (2022).
69. Tarouco, G. *Covid-19 and the Brazilian 2020 Municipal Elections* (International IDEA, 2020).
70. Bach, L., Guillouzouic, A. & Malgouyres, C. Does holding elections during a Covid-19 pandemic put the lives of politicians at risk? *J. Health Econ.* **78**, 102462 (2021).
71. Bernheim, B. D., Buchmann, N., Freitas-Groff, Z. & Otero, S. *The Effects of Large Group Meetings on the Spread of COVID-19: The Case of Trump Rallies* (SSRN, 2020); https://doi.org/10.2139/ssrn.3722299
72. Korolev, I. Identification and estimation of the SEIRD epidemic model for COVID-19. *J. Econ.* **220**, 63–85 (2021).
73. Weed, M. & Foad, A. Rapid scoping review of evidence of outdoor transmission of COVID-19. Preprint at *medRxiv* https://doi.org/10.1101/2020.09.04.20188417 (2020).
74. Dave, D., McNichols, D. & Sabia, J. J. The contagion externality of a superspreading event: the Sturgis Motorcycle Rally and COVID-19. *South. Econ. J.* **87**, 769–807 (2021).
75. Javid, B. et al. Should masks be worn outdoors? *Brit. Med. J.* **373**, n1036 (2021).
76. Jia, J. S., Yuan, Y., Jia, J. & Christakis, N. A. Risk perception and behaviour change after personal vaccination for COVID-19 in the USA. Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/afyv8 (2022).
77. Brauner, J. M. et al. Inferring the effectiveness of government interventions against COVID-19. *Science* https://doi.org/10.1126/science.abd9338 (2021).
78. Li, Y. et al. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. *Lancet Infect. Dis.* **21**, 193–202 (2021).
79. Haug, N. et al. Ranking the effectiveness of worldwide COVID-19 government interventions. *Nat. Hum. Behav.* **4**, 1303–1312 (2020).

80. Election demographics and voter turnout. *Bloomberg Government* https://about.bgov.com/brief/election-demographics-and-voter-turnout/ (2022).
81. Dong, E. et al. The Johns Hopkins University Center for Systems Science and Engineering COVID-19 Dashboard: data collection process, challenges faced, and lessons learned. *Lancet Infect. Dis.* **22**, e370–e376 (2022).
82. Chin, T. et al. US-county level variation in intersecting individual, household and community characteristics relevant to COVID-19 and planning an equitable response: a cross-sectional analysis. *BMJ Open* **10**, e039886 (2020).
83. Miller, I. F., Becker, A. D., Grenfell, B. T. & Metcalf, C. J. E. Disease and healthcare burden of COVID-19 in the United States. *Nat. Med.* **26**, 1212–1217 (2020).
84. Cottrell, D., Herron, M. C. & Smith, D. A. Vote-by-mail ballot rejection and experience with mail-in voting. *Am. Polit. Res.* **49**, 577–590 (2021).
85. Shino, E., Suttmann-Lea, M. & Smith, D. A. Determinants of rejected mail ballots in Georgia's 2018 General Election. *Polit. Res. Q.* **75**, 231–243 (2022).
86. Merkley, E., Bergeron, T., Loewen, P. J., Elias, A. & Lapp, M. Communicating safety precautions can help maintain in-person voter turnout during a pandemic. *Elect. Stud.* **75**, 102421 (2022).
87. Fernandez-Navia, T., Polo-Muro, E. & Tercero-Lucas, D. Too afraid to vote? The effects of COVID-19 on voting behaviour. *Eur. J. Polit. Econ.* **69**, 102012 (2021).
88. In praise of replication studies and null results. *Nature* **578**, 489–490 (2020).
89. Gardner, L., Ratcliff, J., Dong, E. & Katz, A. A need for open public data standards and sharing in light of COVID-19. *Lancet Infect. Dis.* **21**, e80 (2021).
90. Chitwood, M. H. et al. covidestim: COVID-19 nowcasting (CDC, CSTE, 2021); https://covidestim.org/
91. Unwin, H. J. T. et al. State-level tracking of COVID-19 in the United States. *Nat. Commun.* **11**, 6189 (2020).
92. Nouvellet, P. et al. Reduction in mobility and COVID-19 transmission. *Nat. Commun.* **12**, 1090 (2021).
93. Athey, S., Blei, D., Donnelly, R., Ruiz, F. & Schmidt, T. Estimating heterogeneous consumer preferences for restaurants and travel time using mobile location data. *AEA Pap. Proc.* **108**, 64–67 (2018).
94. Squire, R. F. Advanced methods for correlating safegraph patterns with other datasets across time. *SafeGraph* https://colab.research.google.com/drive/16BELpcum4TKoH-5wg8Xym_CGgIGgpu1I?usp=sharing#scrollTo=fZ0aE3PBEafp (2020).
95. *Demonstrations and Political Violence in America: New Data for Summer 2020* (ACLED, 2020); https://acleddata.com/2020/09/03/demonstrations-political-violence-in-america-new-data-for-summer-2020/
96. Sobolev, A., Chen, M. K., Joo, J. & Steinert-Threlkeld, Z. C. News and geolocated social media accurately measure protest size variation. *Am. Polit. Sci. Rev.* **114**, 1343–1351 (2020).
97. *Georgia Early Voting Statistics - 2021 Senate Run-off Election* (United States Elections Project, 2021); https://electproject.github.io/Early-Vote-2020G/GA_RO.html
98. *Election Information and Results Archive* (NJ Division of Elections, 2021); https://nj.gov/state/elections/election-information.shtml
99. *Election Results* (Virginia Department of Elections, 2021); https://www.elections.virginia.gov/resultsreports/election-results/
100. *County Adjacency File* (The United States Census Bureau, 2018); https://www.census.gov/geographies/reference-files/2010/geo/county-adjacency.html
101. *American Community Survey 5-Year Data (2009–2018)* (The United States Census Bureau, 2018); https://www.census.gov/data/developers/data-sets/acs-5year.html
102. The New York Times. *Mask-Wearing Survey Data* (GitHub, 2020); https://github.com/nytimes/covid-19-data
103. Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. B. Julia: a fresh approach to numerical computing. *SIAM Rev.* **59**, 65–98 (2017).
104. Feltham, E. M. TSCSMethods.jl. *GitHub* https://github.com/human-nature-lab/TSCSMethods.jl (2022).
105. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2021).
106. Strochak, S., Ueyama, K. & Williams, A. urbnmapr: state and county shapefiles in sf and tibble format. *GitHub* https://github.com/UrbanInstitute/urbnmapr (2023).
107. Bernal, J. L., Cummins, S. & Gasparrini, A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int. J. Epidemiol.* **46**, 348–355 (2017).
108. O'Neill, S., Kreif, N., Grieve, R., Sutton, M. & Sekhon, J. S. Estimating causal effects: considering three alternatives to difference-in-differences estimation. *Health Serv. Outcomes Res. Methodol.* **16**, 1–21 (2016).
109. Abadie, A. Semiparametric difference-in-differences estimators. *Rev. Econ. Stud.* **72**, 1–19 (2005).
110. Zeldow, B. & Hatfield, L. A. Confounding and regression adjustment in difference-in-differences studies. *Health Serv. Res.* **56**, 932–941 (2021).
111. Lechner, M. The estimation of causal effects by difference-in-difference methods. *Found. Trends Econ.* **4**, 165–224 (2011).
112. Taddeo, M. M., Amorim, L. D. & Aquino, R. Causal measures using generalized difference-in-difference approach with nonlinear models. *Stat. Interface* **15**, 399–413 (2022).
113. Siedner, M. J. et al. Social distancing to slow the US COVID-19 epidemic: longitudinal pretest–posttest comparison group study. *PLoS Med.* **17**, e1003244 (2020).
114. Murray, T. Stay-at-home orders, mobility patterns, and spread of COVID-19. *Am. J. Public Health* **111**, 1149–1156 (2021).
115. Buckee, C. O. et al. Aggregated mobility data could help fight COVID-19. *Science* **368**, 145–146 (2020).
116. Bhattacharjee, S., Liao, S., Paul, D. & Chaudhuri, S. Inference on the dynamics of COVID-19 in the United States. *Sci. Rep.* **12**, 2253 (2022).
117. Abouk, R. & Heydari, B. The immediate effect of COVID-19 policies on social-distancing behavior in the United States. *Public Health Rep.* **136**, 245–252 (2021).
118. Bonvini, M., Kennedy, E. H., Ventura, V. & Wasserman, L. Causal inference for the effect of mobility on COVID-19 deaths. *Ann. Appl. Stat.* **16**, 2458–2480 (2022).
119. Kraemer, M. U. G. et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).
120. Zhang, M. et al. Human mobility and COVID-19 transmission: a systematic review and future directions. *Ann. GIS* **28**, 501–514 (2022).
121. Wikle, N. B. et al. SARS-CoV-2 epidemic after social and economic reopening in three U.S. states reveals shifts in age structure and clinical characteristics. *Sci. Adv.* **8**, eabf9868 (2022).
122. Goldman, N., Pebley, A. R., Lee, K., Andrasfay, T. & Pratt, B. Racial and ethnic differentials in COVID-19-related job exposures by occupational standing in the US. *PLoS ONE* **16**, e0256085 (2021).
123. Rosenbaum, P. R. & Rubin, D. B. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am. Stat.* **39**, 33–38 (1985).
124. Cochran, W. G. & Rubin, D. B. Controlling bias in observational studies: a review. *Sankhya Indian J. Stat. A* **35**, 417–446 (1973).
125. Austin, P. C. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm. Stat.* **10**, 150–161 (2011).

126. Abadie, A. & Imbens, G. W. On the failure of the bootstrap for matching estimators. *Econometrica* **76**, 1537–1557 (2008).
127. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
128. Morey, R. D. & Rouder, J. N. Bayes factor approaches for testing interval null hypotheses. *Psychol. Methods* **16**, 406–419 (2011).
129. Morey, R. et al. BayesFactor (CRAN, 2022); https://richarddmorey.github.io/BayesFactor/
130. Forastiere, L., Airoldi, E. M. & Mealli, F. Identification and estimation of treatment and interference effects in observational studies on networks. *J. Am. Stat. Assoc.* **116**, 901–918 (2021).

Acknowledgements

We thank B. Snyder for help with data development and G. King for helpful comments. This research was supported by the Robert Wood Johnson Foundation. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

E.F., L.F., M.A. and N.A.C. conceptualized the project. E.F., L.F., M.A. and N.A.C. developed the methodology. E.F. and L.F. performed statistical analysis. N.A.C. acquired funding. E.F., L.F., M.A. and N.A.C. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41562-023-01654-1.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-023-01654-1.

Correspondence and requests for materials should be addressed to Eric Feltham.

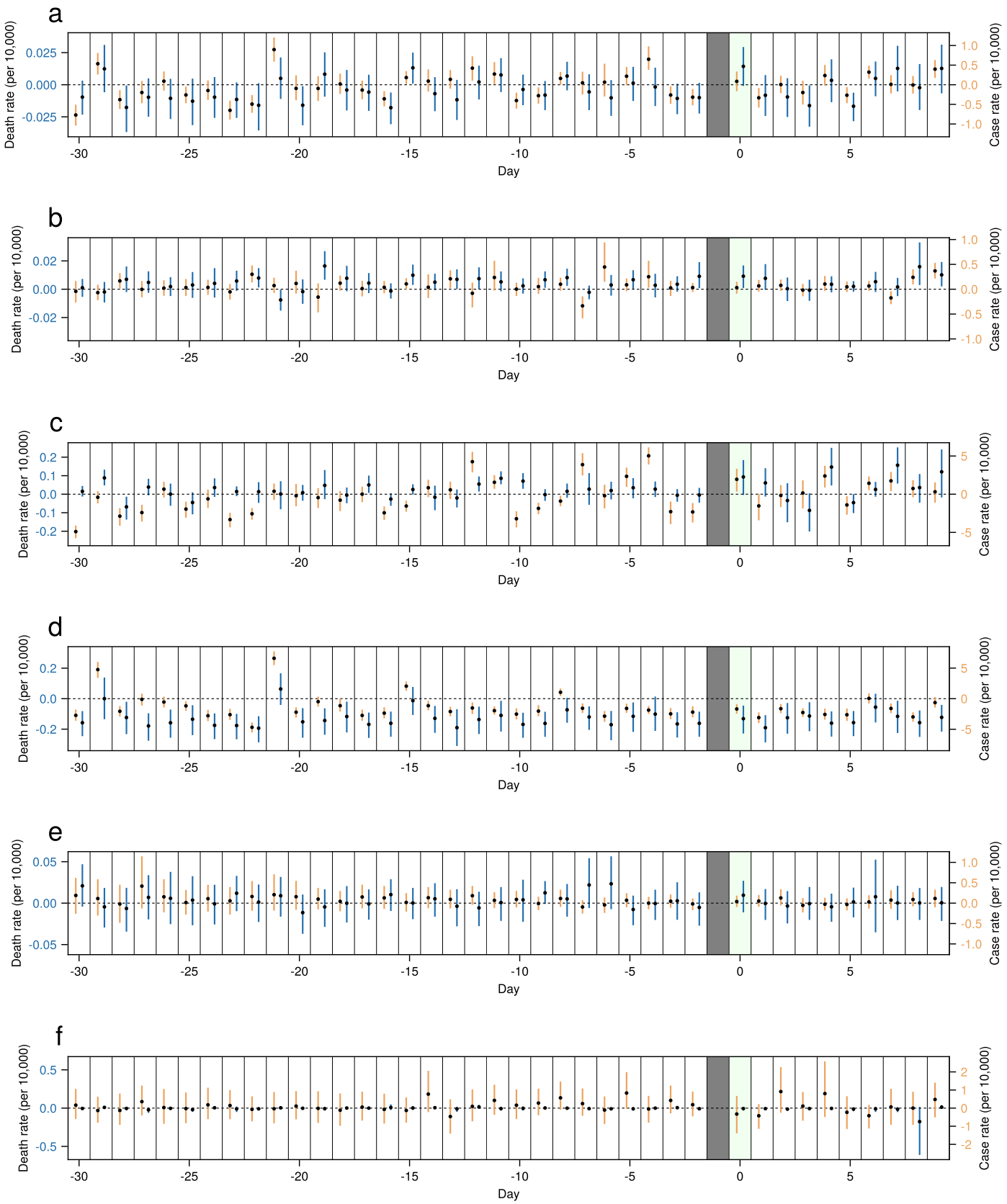
Peer review information *Nature Human Behaviour* thanks Engy Ziedan, Yesim Tozan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

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

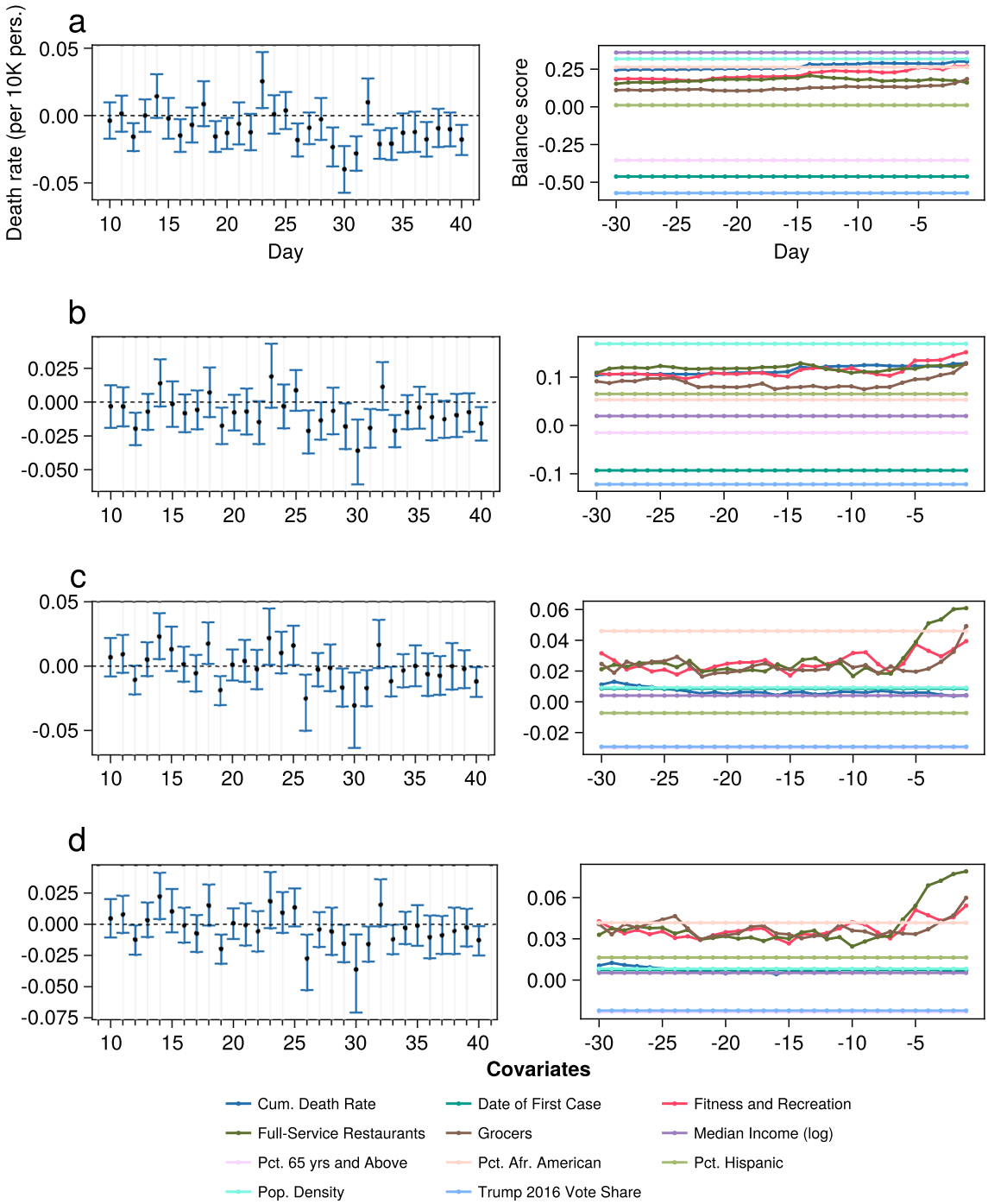
[1]Yale Institute for Network Science, Yale University, New Haven, CT, USA. [2]Department of Sociology, Yale University, New Haven, CT, USA. [3]Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA. [4]Frank H. Netter MD School of Medicine, Quinnipiac University, North Haven, CT, USA. [5]Department of Statistics and Data Science, Yale University, New Haven, CT, USA. [6]Department of Medicine, Yale School of Medicine, New Haven, CT, USA. ✉e-mail: eric.feltham@aya.yale.edu

https://doi.org/10.1038/s41562-023-01654-1



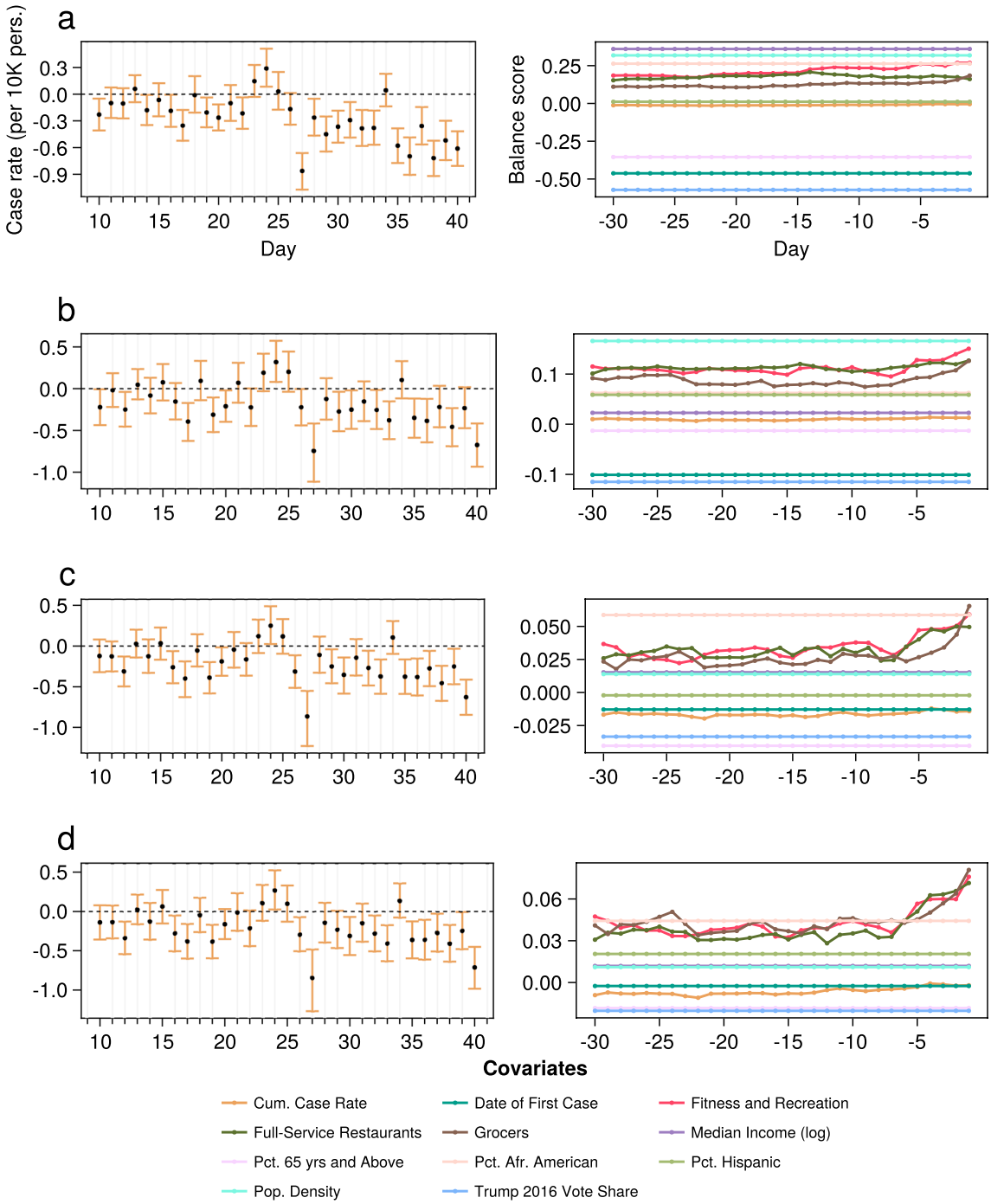
Extended Data Fig. 1 | Pre-outcome-window ATTs. ATTs for the death and case rates, from 30 days before treatment up to the 9 days after treatment. The light green cell is the day of treatment, and the dark grey cell is the reference day for the ATT calculation (see Methods). In each panel, the error bars indicate the 95% CIs. The panels present the 'average treatment effect on the treated' (ATT) estimates for the **(a)** omnibus analysis, **(b)** the primary elections, **(c)** GA special election, **(d)** NJ & VA gubernatorial elections, **(e)** Donald Trump's political rallies, and **(f)** the BLM protests. In each case, we observe results that are similar to those for the corresponding main analysis; that is, generally non-significant results, which is expected for the period prior to treatment.

https://doi.org/10.1038/s41562-023-01654-1



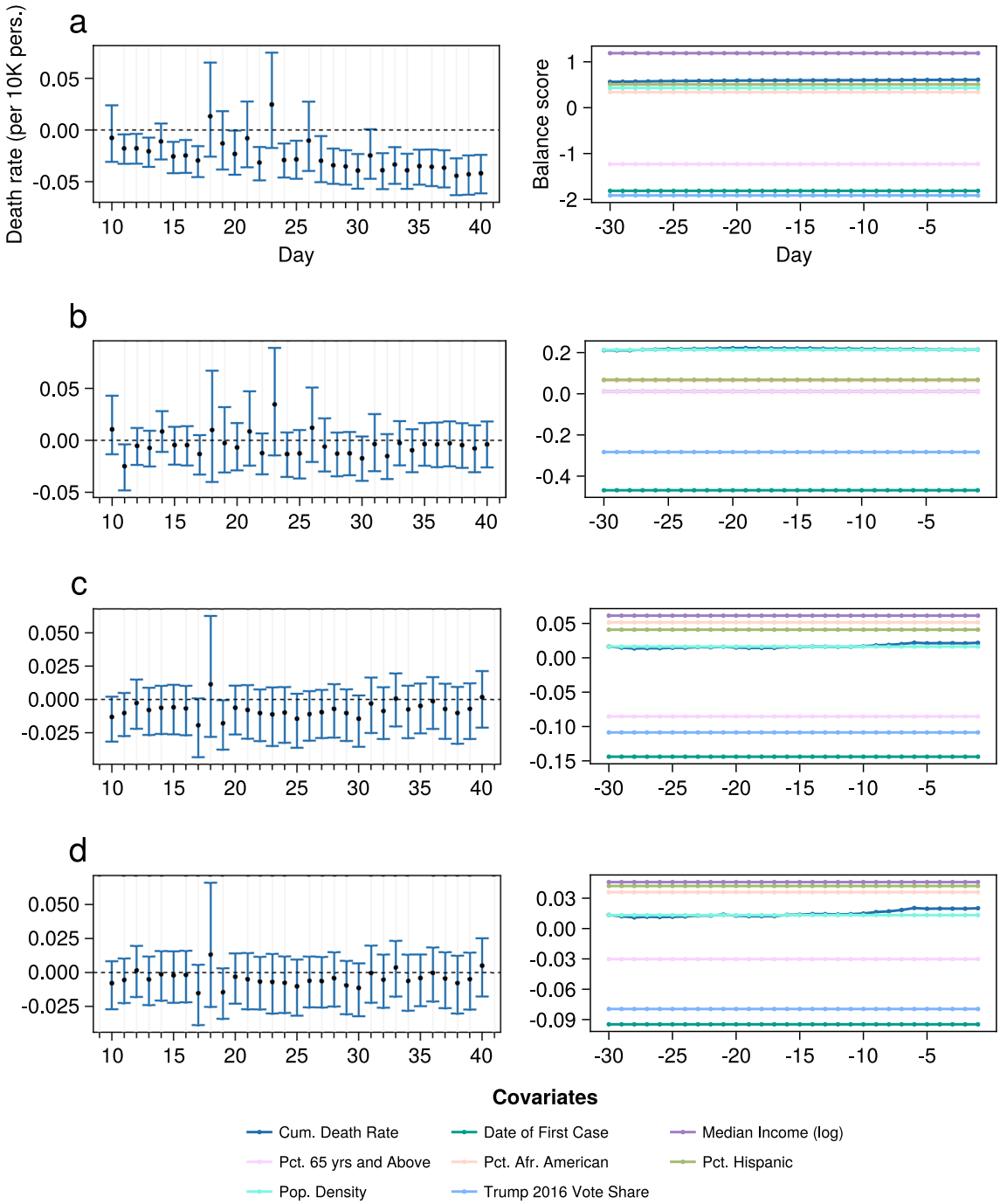
Extended Data Fig. 2 | Omnibus estimates for the effect of political events on the death rate. (a) Overall ATT estimates and covariate balance before matching refinement. (The ATTs represent the average difference in the change in death rates, from the day before treatment to 10 to 40 days after an election. The pre-refinement covariate balance for each matching covariate. All covariates are measurements at the county level. The balance score is the average standardized mean difference between the treated and control units, over a matching period from 30 days before to 1 day before treatment. There are 2135 treated units present. **(b)** Overall ATT estimates and covariate balance after matching refinement, to no more than the five best matches to each treated county. **(c)** Overall ATT estimates and covariate balance before matching refinement and after the application of a caliper. **(d)** Overall ATT estimates and covariate balance after the application of a caliper, and after matching refinement, the observed balance scores are, on average over the matching window, within the threshold of 0.1, indicating sufficient similarity between the treated and matched counties. On average, for estimates over the outcome window, 1449.4 treated units remain, with 6084.5 matches.

https://doi.org/10.1038/s41562-023-01654-1



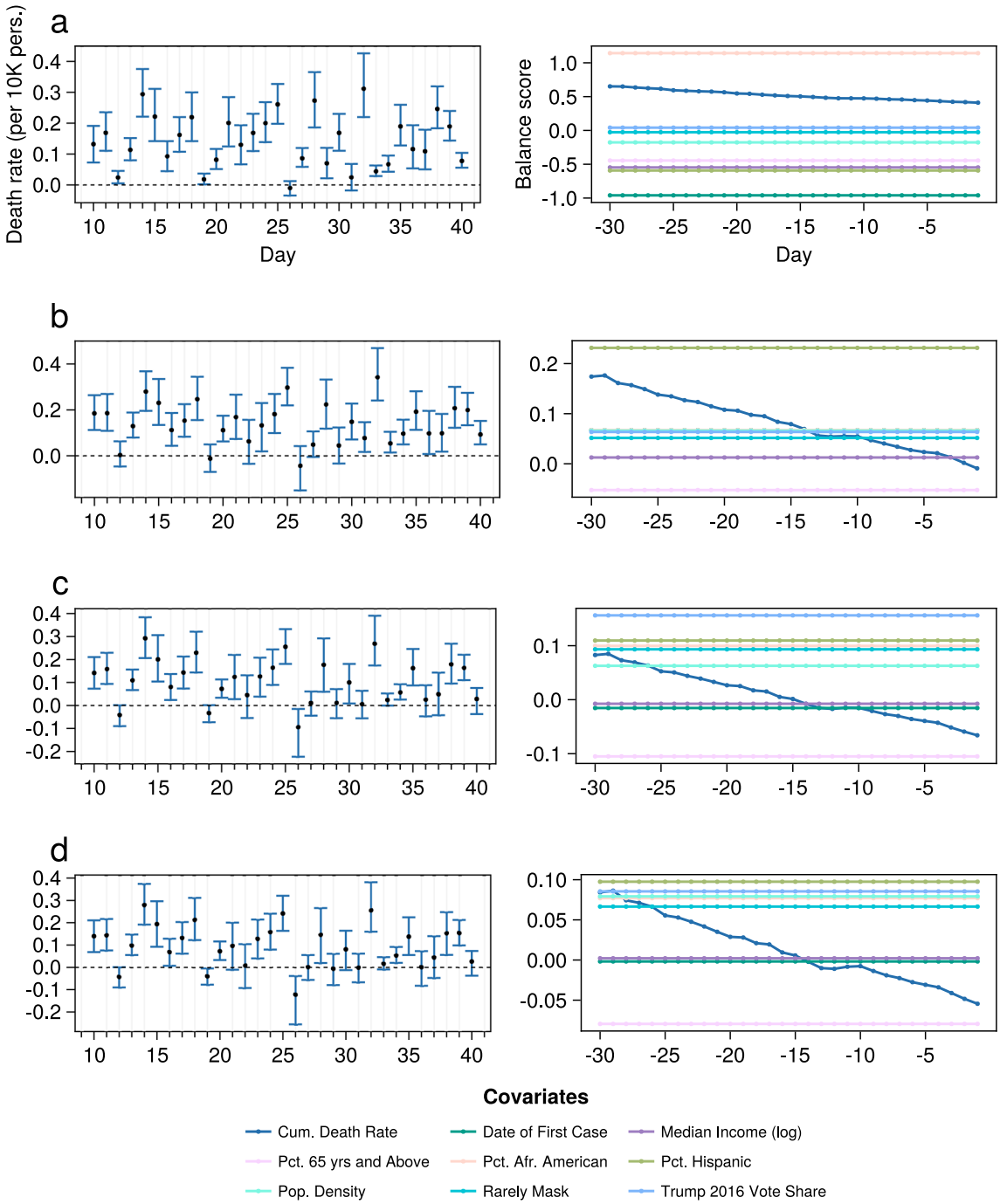
Extended Data Fig. 3 | Omnibus estimates for the effect of political events on the case rate. (a) Overall ATT estimates and covariate balance before matching refinement. (The ATTs represent the average difference in the change in case rates, from the day before treatment to 10 to 40 days after an election. The pre-refinement covariate balance for each matching covariate. All covariates are measurements at the county level. The balance score is the average standardized mean difference between the treated and control units, over a matching period from 30 days before to 1 day before treatment. There are 2135 treated units present. **(b)** Overall ATT estimates and covariate balance after matching refinement, to no more than the five best matches to each treated county. **(c)** Overall ATT estimates and covariate balance before matching refinement and after the application of a caliper. **(d)** Overall ATT estimates and covariate balance after the application of a caliper, and after matching refinement, the observed balance scores are, on average over the matching window, within the threshold of 0.1, indicating sufficient similarity between the treated and matched counties. On average, for estimates over the outcome window, 1473.5 treated units remain, with 6305.4 matches.

https://doi.org/10.1038/s41562-023-01654-1



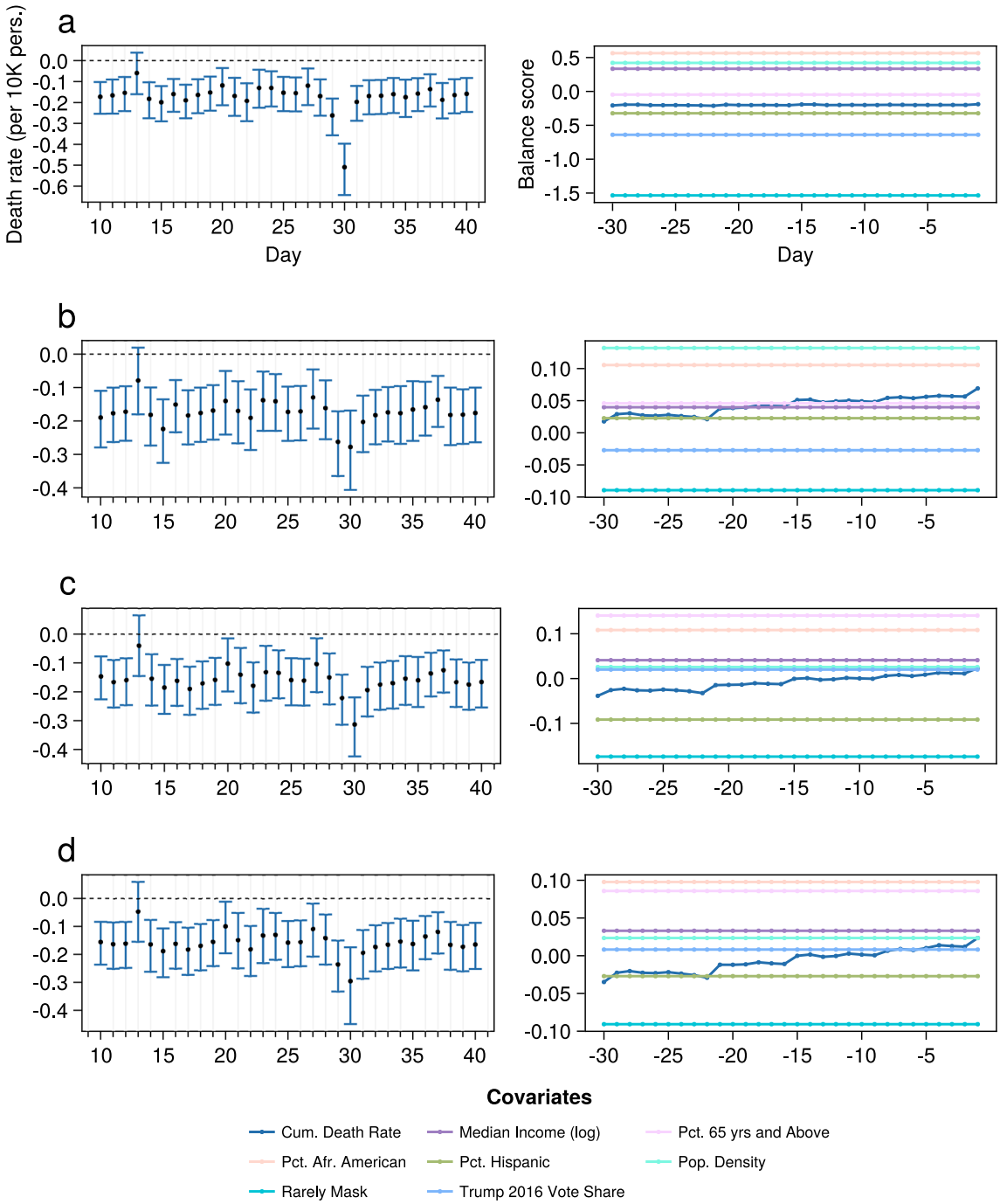
Extended Data Fig. 4 | Overall estimates for the primary elections. In each panel, the error bars indicate the 95% CIs. **(a)** Overall ATT estimates and covariate balance before matching refinement. (The ATTs represent the average difference in the change in death rates, from the day before treatment to 10 to 40 days after an election. The pre-refinement covariate balance for each matching covariate. All covariates are measurements at the county level. The balance score is the average standardized mean difference between the treated and control units, over a matching period from 30 days before to 1 day before treatment. There are 1173 treated units present. **(b)** Overall ATT estimates and covariate balance after matching refinement, to no more than the five best matches to each treated county. **(c)** Overall ATT estimates and covariate balance before matching refinement and after the application of a caliper. **(d)** Overall ATT estimates and covariate balance after the application of a caliper, and after matching refinement, the observed balance scores are, on average over the matching window, within the threshold of 0.1, indicating sufficient similarity between the treated and matched counties. On average, for estimates over the outcome window, 961.2 treated units remain, with 4283.8 matched units.

https://doi.org/10.1038/s41562-023-01654-1



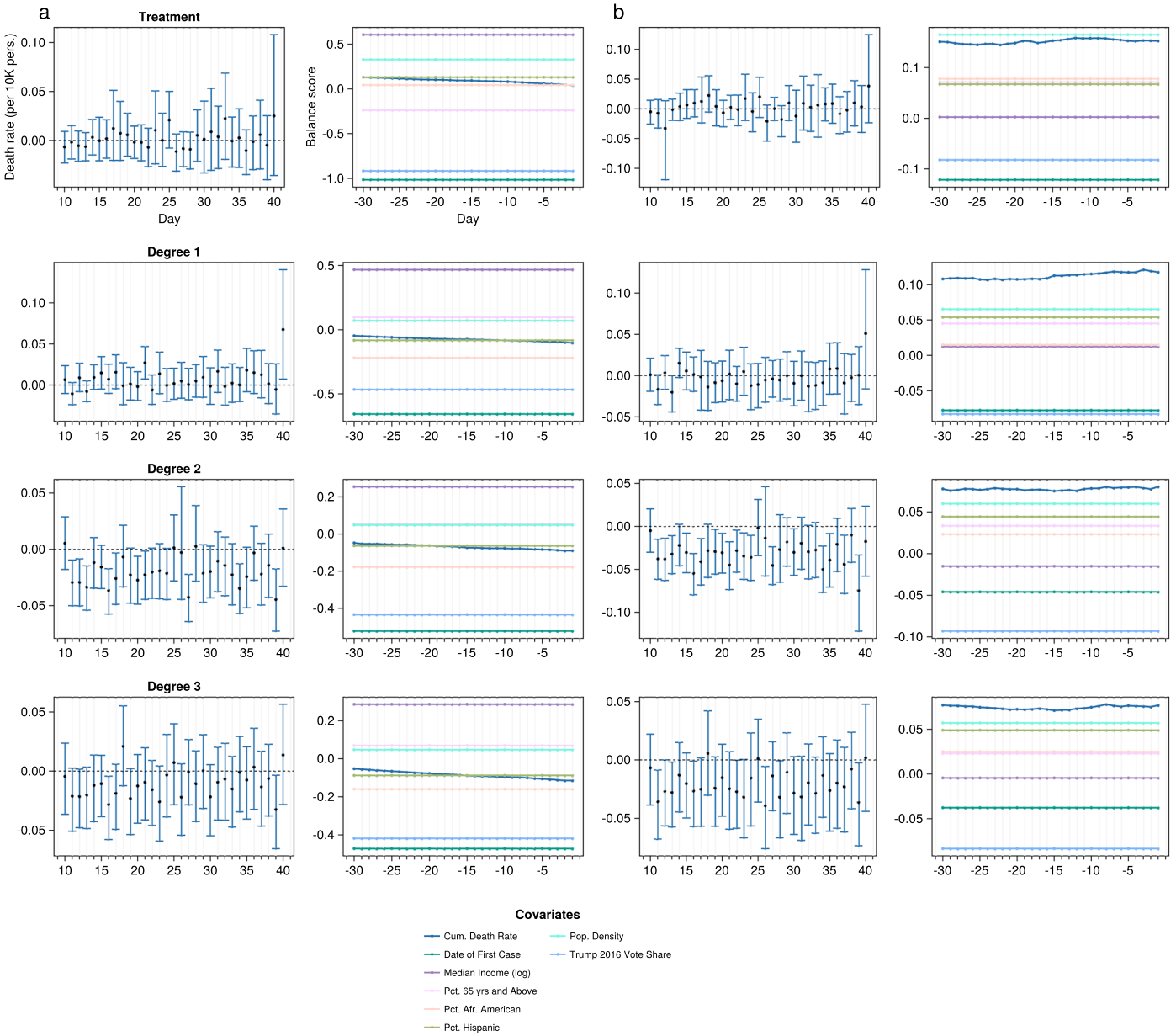
Extended Data Fig. 5 | Overall estimates for the GA elections. In each panel, the error bars indicate the 95% CIs. **(a)** Overall ATT estimates and covariate balance before matching refinement. (The ATTs represent the average difference in the change in death rates, from the day before treatment to 10 to 40 days after an election. The pre-refinement covariate balance for each matching covariate. All covariates are measurements at the county level. The balance score is the average standardized mean difference between the treated and control units, over a matching period from 30 days before to 1 day before treatment. There are 159 treated units present. **(b)** Overall ATT estimates and covariate balance after matching refinement, to no more than the five best matches to each treated county. **(c)** Overall ATT estimates and covariate balance before matching refinement and after the application of a caliper. **(d)** Overall ATT estimates and covariate balance after the application of a caliper, and after matching refinement, the observed balance scores are, on average over the matching window, within the threshold of 0.1, indicating sufficient similarity between the treated and matched counties. On average, for estimates over the outcome window, 137 treated units remain, with 566 matched units.

https://doi.org/10.1038/s41562-023-01654-1



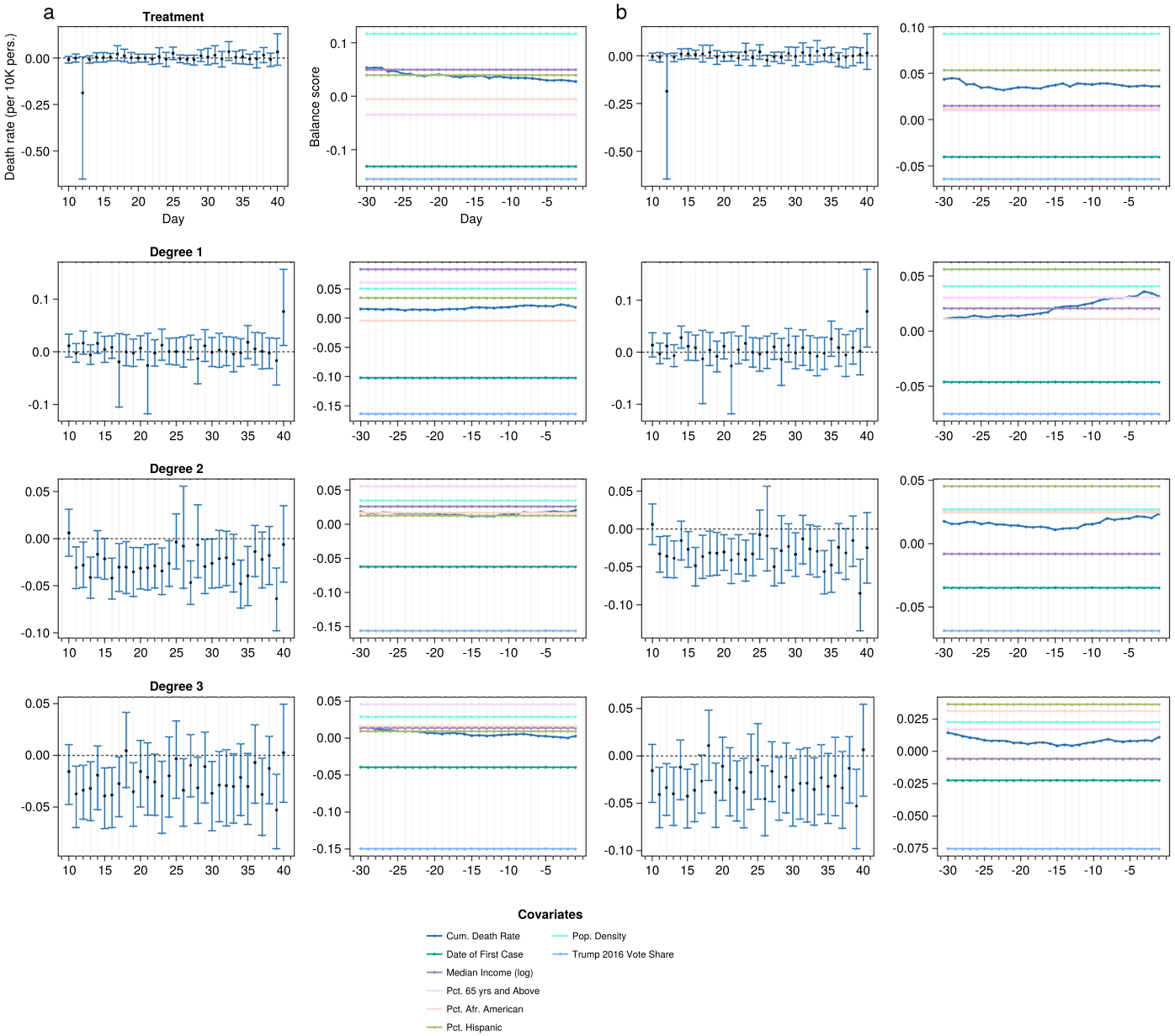
Extended Data Fig. 6 | Overall estimates for the NJ and VA gubernatorial elections. In each panel, the error bars indicate the 95% CIs. **(a)** Overall ATT estimates and covariate balance before matching refinement. (The ATTs represent the average difference in the change in death rates, from the day before treatment to 10 to 40 days after an election. The pre-refinement covariate balance for each matching covariate. All covariates are measurements at the county level. The balance score is the average standardized mean difference between the treated and control units, over a matching period from 30 days before to 1 day before treatment. There are 154 treated units present. **(b)** Overall ATT estimates and covariate balance after matching refinement, to no more than the five best matches to each treated county. **(c)** Overall ATT estimates and covariate balance before matching refinement and after the application of a caliper. **(d)** Overall ATT estimates and covariate balance after the application of a caliper, and after matching refinement, the observed balance scores are, on average over the matching window, within the threshold of 0.1, indicating sufficient similarity between the treated and matched counties. On average, for estimates over the outcome window, 141 treated units remain, with 649 matched units.

https://doi.org/10.1038/s41562-023-01654-1



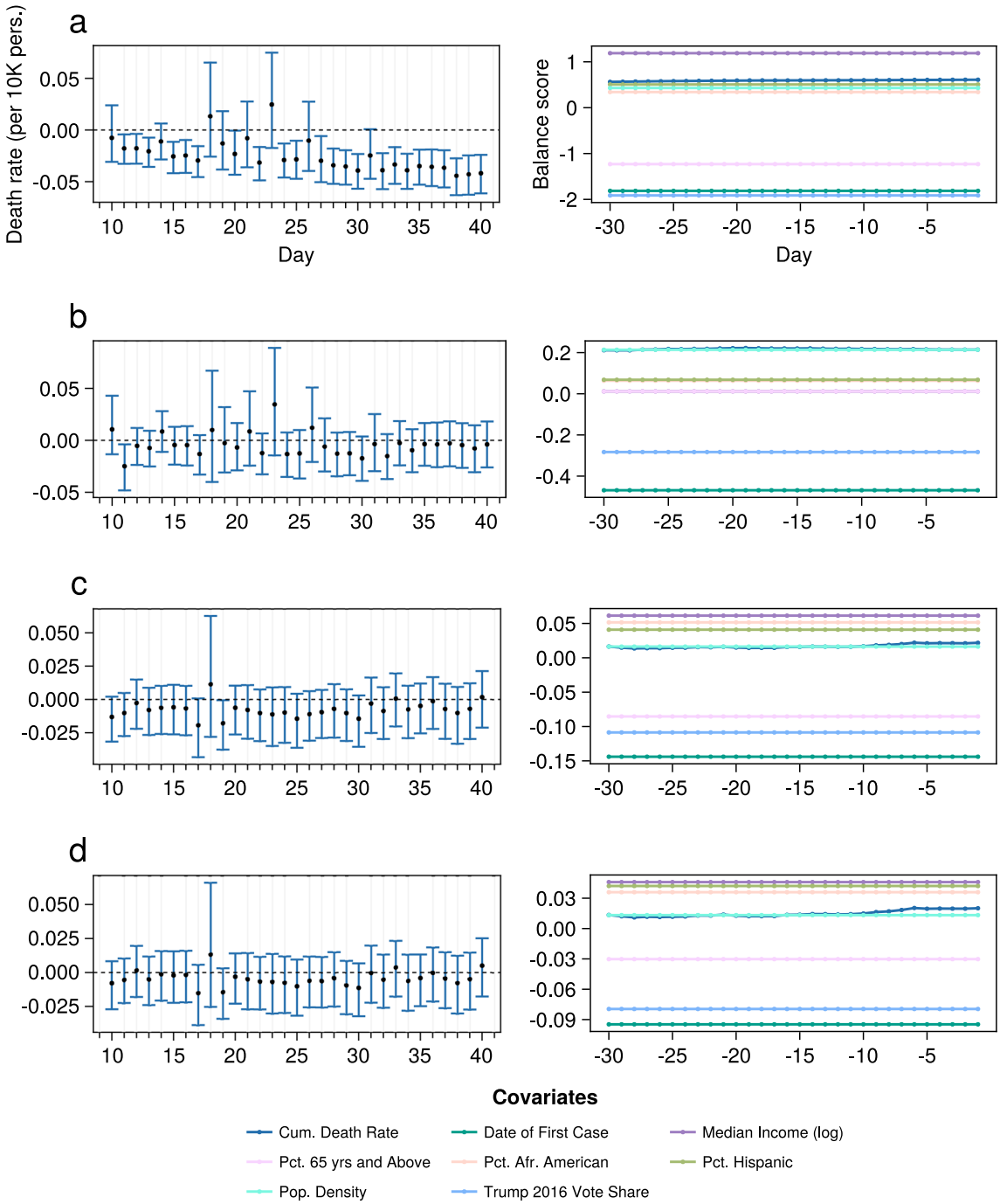
Extended Data Fig. 7 | Estimates for Donald Trump's rallies, stratified by exposure (without caliper). (a) Overall ATT estimates and covariate balance before matching refinement, for each stratum. (The ATTs represent the average difference in the change in death rates, from the day before treatment to 10 to 40 days after an election. The pre-refinement covariate balance for each matching covariate. All covariates are measurements at the county level. The balance score is the average standardized mean difference between the treated and control units, over a matching period from 30 days before to 1 day before treatment. The number of treated units in each stratum are on average, over the outcome window. For Treatment, there are 67 treated units; for Degree 1, there are 397 treated units; for Degree 2, there are 788 treated units; for Degree 3, there are 1156 treated units. **(b)** Overall ATT estimates and covariate balance after matching refinement, to no more than the five best matches to each treated county, for each stratum.

https://doi.org/10.1038/s41562-023-01654-1



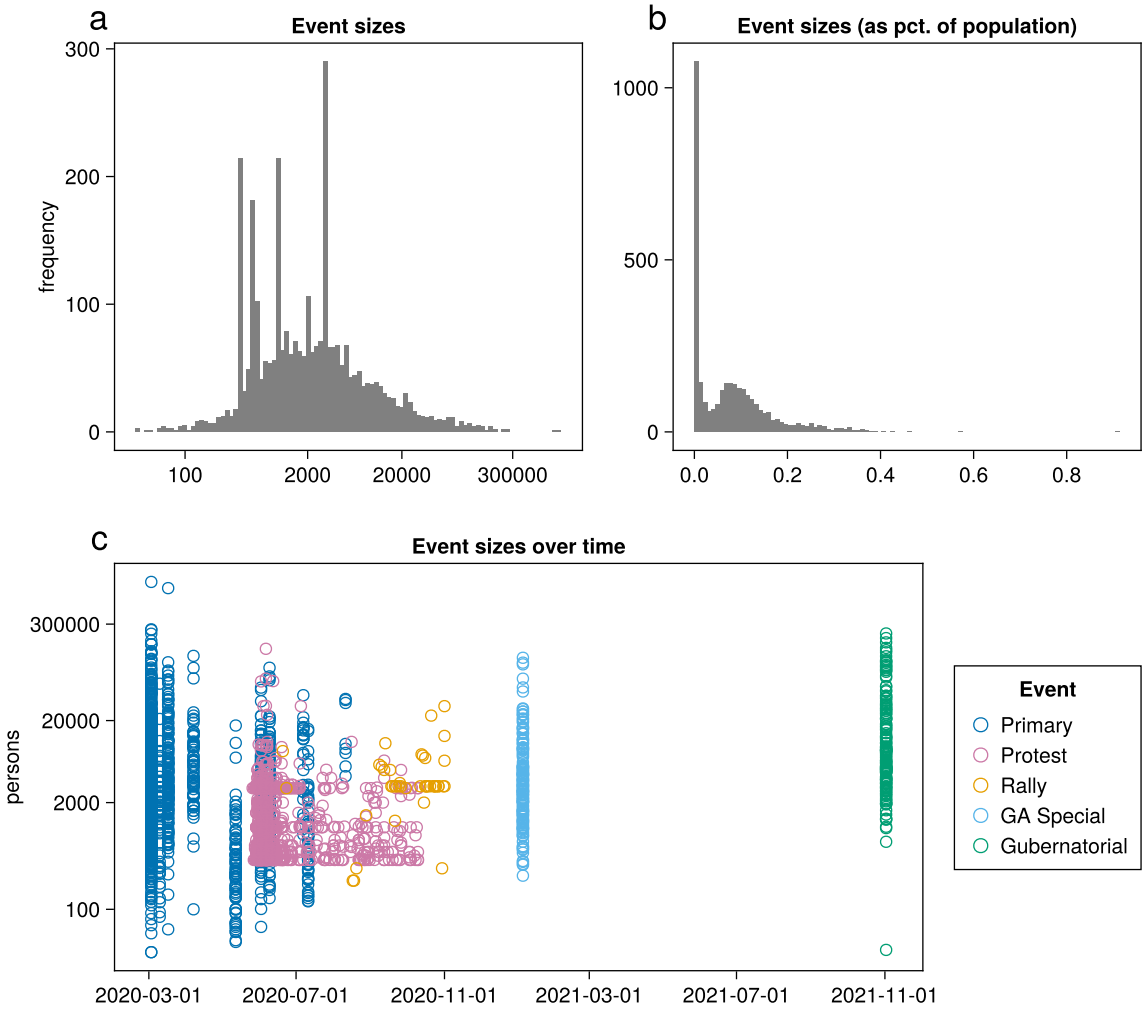
Extended Data Fig. 8 | Estimates for Donald Trump's rallies, stratified by exposure (with caliper). In both panels, the error bars indicate the 95% CIs. **(a)** Overall ATT estimates and covariate balance before matching refinement, for each stratum. (The ATTs represent the average difference in the change in death rates, from the day before treatment to 10 to 40 days after an election. The pre-refinement covariate balance for each matching covariate. All covariates are measurements at the county level. The balance score is the average standardized mean difference between the treated and control units, over a matching period from 30 days before to 1 day before treatment. The number of treated units in each stratum are on average, over the outcome window. For Treatment, 58.2 treated units remain; for Degree 1, 363.4 treated units remain; for Degree 2, 713.3 treated units remain; for Degree 3, 1055.1 treated units remain. Respectively, with 251.8, 1679, 3290.9, 4280.4 matches. **(b)** Overall ATT estimates and covariate balance after matching refinement, to no more than the five best matches to each treated county, for each stratum.

https://doi.org/10.1038/s41562-023-01654-1



Extended Data Fig. 9 | Overall estimates for the BLM protests. In each panel, the error bars indicate the 95% CIs. In each panel, the error bars indicate the 95% CIs. **(a)** Overall ATT estimates and covariate balance before matching refinement. (The ATTs represent the average difference in the change in death rates, from the day before treatment to 10 to 40 days after an event. The pre-refinement covariate balance for each matching covariate. All covariates are measurements at the county level. The balance score is the average standardized mean difference between the treated and control units, over a matching period from 30 days before to 1 day before treatment. On average, for estimates over the outcome window, 658 treated units are present. **(b)** Overall ATT estimates and covariate balance after matching refinement, to no more than the five best matches to each treated county. **(c)** Overall ATT estimates and covariate balance before matching refinement and after the application of a caliper. **(d)** Overall ATT estimates and covariate balance after the application of a caliper, and after matching refinement, the observed balance scores are, on average over the matching window, within the threshold of 0.1, indicating sufficient similarity between the treated and matched counties. On average, for estimates over the outcome window, 450.5 treated units remain, with 1901.7 matched units.

https://doi.org/10.1038/s41562-023-01654-1



Extended Data Fig. 10 | Political Event sizes. (a) Overall distribution of event sizes in the data, across event type. Large frequencies for specific values reflect the thresholding procedure used to estimate crowd sizes from different reports (see Methods). **(b)** Overall distribution of event sizes in the data, across each event type, represented as the percentage of the county population in which the event takes place. **(c)** Event sizes over the roughly two-year period that constitutes our study horizon, colored by event type. Event sizes are plotted on the natural log scale, labelled on the original scale (persons at event).

nature portfolio

Corresponding author(s): Eric Martin Feltham

Last updated by author(s): May 29, 2023

Reporting Summary

Nature Portfolio 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 Portfolio 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	Except for the SafeGraph mobility data, all data used is freely available in the public domain. In each case, data was extracted from the relevant websites and cleaned with custom R or Julia code written for this project. All code used to obtain the data, and the links to the original data, are available in a Github repository COVIDPoliticalEvents (https://github.com/human-nature-lab/COVIDPoliticalEvents.jl).
Data analysis	Data analysis was conducted in the Julia programming language, v 1.7.1, using two packages developed specifically for this analysis: TSCSMethods v 1.0 (https://github.com/human-nature-lab/TSCSMethods.jl); and COVIDPoliticalEvents v 1.0 (https://github.com/human-nature-lab/COVIDPoliticalEvents.jl). Data cleaning was carried out with the R programming language, v. 4.0.3. The epidemiological models were estimated with the covidestim R package v. 0.1.1 (https://pkg.covidestim.org).

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The raw data required for this work is publicly available, but had to be assembled. The data on COVID-19 case and death counts are available from the Johns Hopkins Coronavirus Resource Center (https://coronavirus.jhu.edu/data). US Census data is available also publicly available (https://www.census.gov/programs-surveys/acs). Supplementary data, the USA County adjacency files and Census Region specifications, are also available from the US Census website. County-level masking data is available from the New York Times Github repository (https://raw.githubusercontent.com/nytimes/covid-19-data/master/mask-use/mask-use-by-county.csv) The elections turnout data is available directly from state governmental election agencies. Data on the BLM protests is available from the Armed Conflict Location & Event Data Project (https://acleddata.com/), as the US Crisis Monitor dataset (https://acleddata.com/special-projects/us-crisis-monitor/). The mobility tracking data is available from SafeGraph, Inc. and is freely available to academic researchers through an academic license (https://www.safegraph.com/products/places). Details and data files are contained the COVIDPoliticalEvents GitHub repository (https://github.com/human-nature-lab/COVIDPoliticalEvents.jl).

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

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

Behavioural & social sciences study design

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

Study description	Epidemic disease can spread during mass gatherings. We assessed the impact on the local-area trajectory of the COVID-19 epidemic of a type of mass gathering about which comprehensive data were available. Here, we examined five types of political events in 2020 and 2021: the US primary elections; the US Senate special election in Georgia; the gubernatorial elections in New Jersey and Virginia; Donald Trump's political rallies; and the Black Lives Matter protests. Our study period encompassed over 700 such mass gatherings during multiple phases of the pandemic. We used data from the 48 contiguous states, representing 3,218 counties, and we implemented a novel extension of a recently developed non-parametric, generalized difference-in-difference estimator with a (high-quality) matching procedure for panel data to estimate the average effect of the gatherings on local mortality and other outcomes. There were no statistically significant increases in cases, deaths, or a measure of epidemic transmissibility (Rt) in a 40-day period following large-scale political activities. We estimated quantitatively small and insignificant effects, corresponding to an average difference of -0.0567 deaths (95% CI: -0.319, 0.162), and 8.275 cases (95% CI: -1.383, 20.7), on each day, for counties that held mass gatherings for political expression compared to matched control counties. In sum, there is no statistical evidence of a material increase in local COVID-19 deaths, cases, or transmissibility after mass gatherings for political expression during the first two years of the pandemic in the USA. This may relate to the specific manner in which such activities are typically conducted.
Research sample	The units of study consist of all mainland US counties. We analyze all major elections (held at the state level) for which there is treatment variation (excluding the national general election) over 2020 and 2021. We consider all major protest events occurring in the summer of 2020 (May 26 to October 10), covering the major timeline of the BLM protest movement. We consider all of Donald Trump's major campaign rallies in 2020.
Sampling strategy	We did not employ a sampling strategy, since we include the whole population of mainland US Counties in our analysis. Further, we only analyze quantitative observational data.

Data collection	Our data was collected from authoritative, publicly available sources in all but one case (our mobility data comes from a private company, SafeGraph). The study is non-experimental, so there was no procedure for blinding.
Timing	The most recent versions of our data were downloaded on May 10, 2022.
Data exclusions	Counties (and county equivalents) outside of the mainland US (AK, HI) were excluded from the analysis. Out of the 3,243 counties in the USA (excluding territories) we the include 3,119 counties (outside of AK, HI) as eligible match units in our statistical analyses. Further, include the full set of covered elections over the period of study, all of the Trump rallies, and all BLM protest events larger than 800 persons (events smaller than 400 persons were eligible to be match units, events between 400 and 800 were excluded from the analysis to ensure that any potentially significant event were not included as match units). Data exclusions were established prior to the analysis of the data, and are described in detail in the Methods section of the paper.
Non-participation	We did not work with human subjects for this analysis.
Randomization	We did not make use of a randomized experimental design. However, we adjusted for an extensive set of covariates using a statistical matching procedure designed for panel data with multiple treatments. More specifically, covariate adjustment is conducted using a matching procedure whereby the five best matches are selected and used for each treated county. For each model, the quality of the five best matches was examined with reference to a balance score, calculated based on the standardized distance between a matched unit and its best matches in terms of covariate similarity. In each case, acceptable balance scores were obtained, on average, for each matching covariate. This approach also allows for time-varying treatments that can occur multiple times over the observed window. We used a range of variables describing the counties, including demographic, political, and mobility variables. Our mobility data tracks daily visits to points of interest at individual locations across the US in census block groups, typically covering between 600-3,000 individuals, aggregated to our 3,119 counties. Mobility data has previously been linked to the spread of COVID-19. Many existing studies use general movement patterns, such as overall time spent away from home, which does not necessarily indicate whether individuals engage in the sort of close interpersonal contact that is heavily responsible for the spread of COVID-19. However, more recent studies have used finer-grained mobility metrics in line with our approach. We track movement specifically at full-service restaurants, grocers, and fitness and recreation facilities, some of which are known to be high-risk locations, and directly include these measures in the matching procedure.

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
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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