The Effects of Large Group Meetings on the Spread of COVID-19: The Case of Trump Rallies

B. Douglas Bernheim Nina Buchmann Zach Freitas-Groff
Sebastián Otero*

January 20, 2021

Abstract

We investigate the effects of large group meetings on the spread of COVID-19 by studying the impact of eighteen Trump campaign rallies. To capture the effects of subsequent contagion within the pertinent communities, our analysis encompasses up to ten post-rally weeks for each event. Our method is based on a collection of regression models, one for each event, that capture the relationships between post-event outcomes and pre-event characteristics, including demographics and the trajectory of COVID-19 cases, in similar counties. We explore a total of 24 procedures for identifying sets of matched counties. For the vast majority of these variants, our estimate of the average treatment effect across the eighteen events implies that they increased subsequent confirmed cases of COVID-19 by more than 250 per 100,000 residents. Extrapolating this figure to the entire sample, we conclude that these eighteen rallies ultimately resulted in more than 30,000 incremental confirmed cases of COVID-19. Applying county-specific post-event death rates, we conclude that the rallies likely led to more than 700 deaths (not necessarily among attendees).

*Department of Economics, Stanford University, Stanford, CA 94305-6072. We would like to thank Mark Duggan and Guido Imbens for valuable comments and suggestions. All remaining errors are our own.

1 Introduction

As of this writing, more than 8.7 million Americans have contracted COVID 19, resulting in more than 225,000 deaths (Dong et al., 2020). The CDC has advised that large in-person events, particularly in settings where participants do not wear masks or practice social distancing, pose a substantial risk of further contagion (Centers for Disease Control and Prevention, 2020). There is reason to fear that such gatherings can serve as "superspreader events," severely undermining efforts to control the pandemic (Dave et al., 2020).

The purpose of this study is to shed light on these issues by studying the impact of election rallies held by President Donald Trump's campaign between June 20th and September 30th, 2020. Trump rallies have several distinguishing features that lend themselves to this inquiry. First, they involved large numbers of attendees. Though data on attendance is poor, it appears that the number of attendees was generally in the thousands and sometimes in the tens of thousands. Because the available information about the incidence of COVID-19 is at the county level, the effects of smaller meetings would be more difficult to detect using our methods. Second, the set of major Trump campaign events is easily identified. We know whether and when the Trump campaign held a rally in each county. This property allows us to distinguish between "treated" and "untreated" counties. Third, the events occurred on identifiable days. They neither recurred within a given county nor stretched across several days. This feature allows us to evaluate the effects of individual gatherings. Fourth, rallies were not geographically ubiquitous. As a result, we always have a rich set of untreated counties we can use as comparators. Fifth, at least through September 2020, the degree of compliance with guidelines concerning the use of masks and social distancing was low (Sanchez, 2020), in part because the Trump campaign downplayed the risk of infection (Bella, 2020). This feature heightens the risk that a rally could become a "superspreader event."

Despite these favorable characteristics, the task of evaluating the effects of Trump rallies on the spread of COVID-19 remains challenging for the reasons detailed in Section 3. Briefly, our approach involves a separate analysis for each of eighteen Trump rallies. We identify a set of counties that are comparable to the event county at the pertinent point in time, based at least in part on the trajectory of confirmed COVID-19 cases prior to the rally date. We then estimate the statistical relationship among those counties between subsequent COVID-19 cases and various conditions, such as pre-existing COVID-19 prevalence and pandemic-related restrictions, along with demographic characteristics. We use this relationship to predict the post-event incidence of new confirmed COVID-19 cases for the event county. The difference between the actual incidence and the predicted incidence is an estimate of the treatment effect. Because the standard error of each prediction is large, we combine estimates across events to obtain an average treatment effect. We also perform the same analysis for the event counties focusing on a "placebo event" occurring 10 weeks before the actual event. This exercise allows us to determine whether our

method systematically mispredicts the outcomes for event counties, and to evaluate the possible existence of pre-event trend in "unexplained" cases occurring prior to the event (which, in principle, could produce spurious treatment effects after the event).

Although our methods involve prediction models, it is important to understand that the nature of these predictions differ in critical ways from the types of predictions generated by epidemiological models. In the typical epidemiological analysis, predictions for a collection of jurisdictions between a fixed point in time, t (often the present), and some subsequent point in time, t', are based only on data pertaining to period t and earlier periods (e.g., current and past data). In contrast, our approach is to predict outcomes in an event jurisdiction between periods t and t' based not only on that jurisdiction's history up until t, but also on the complete histories of comparable counties through period t'. In other words, our predictions employ "future" data for comparable counties, whereas epidemiological models do not. Consequently, in contrast to the epidemiologists, we are not forecasting into an unobserved period of time. Rather, we are observing outcomes within the forecast period for comparable counties, and making inferences about the county of interest based on similarities.

For the vast majority of county matching procedures we employ, our estimate of the average treatment effect across the eighteen rallies implies that they increased subsequent confirmed cases of COVID-19 by more than 250 per 100,000 residents. In contrast, the pseudo-treatment effects for the placebo events are small, slightly negative, and statistically insignificant. The striking contrast between the estimated treatment effects for the actual events and the pseudo-treatment effects for the placebo events underscores the reliability of our results. Extrapolating the average treatment effects to the entire sample, we conclude that these eighteen rallies ultimately resulted in more than 30,000 incremental confirmed cases of COVID-19. Applying county-specific post-event death rates, we conclude that the rallies likely led to more than 700 deaths (not necessarily among attendees).

We are aware of a small handful of related analyses. Dave et al. (2020) focus on the Tulsa rally. Based on a synthetic control involving comparable counties, they find no elevation in new cases or deaths. A problem with focusing on a single event is that COVID-19 outcomes are highly variable, as indicated by the magnitudes of the standard errors of the forecasts in our analysis. In such settings, measuring the average treatment effect over multiple events, as in our study, produces more reliable results. Like us, Waldrop and Gee (2020) follow the strategy of focusing on a collection of rallies,[1] but their analysis simply asks whether cases in the three weeks following the rally were above or below pre-existing trends. Our analysis involves more elaborate forecasts and encompasses up to 10 weeks of post-event data. The latter difference may be particular important, in that the effects of a superspreader event may snowball over time. Even so, the study's conclusions

[1] Because they focused on a shorter post-event window, they employed data on 22 rally events, whereas we study 18.

corroborate ours: "...the Trump rallies are often followed by increased community spread of the coronavirus...". Analysis in Nayer (2020) points to a similar conclusion.

The paper is organized as follows. Section 2 describes the data used in our analysis, Section 3 details our methods and presents our main results, Section 4 presents additional analyses of highly impacted counties, and Section 5 concludes.

2 Data

2.1 Trump rallies

We focus on rallies held between June 20th and September 22nd. While the Trump campaign held many rallies after September 22nd, we do not include them in our analysis for two reasons. First, because the effects of an event may grow substantially over time as incremental infections spread, we required at least four weeks of post-event data (excluding the week of the rally). Second, there are some indications that compliance with public health guidelines, such as the use of masks, improved at later rallies. While it would be worth evaluating the diminution of treatment effects resulting from greater compliance, we currently lack sufficient compliance data to conduct that investigation.

We obtain a list of general election Tump rallies via a regularly-updated Wikipedia list based on local and national news reports. We verify the date and county of the rally from local news reports. We manually record from news reports whether the rally was indoors.

Table 1 lists the rallies we include in our analysis, their dates, and whether they were indoor or outdoor.

2.2 Data on COVID 19

Our data on the incidence of COVID-19 comes from the COVID-19 Data Repository maintained by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (Dong et al., 2020). The CSSE collects case and death reports from the CDC and U.S. state and local governments. The CSSE time series data includes total confirmed cases and deaths from COVID-19 associated with each Federal Information Processing Standards county code beginning January 22nd, 2020. The frequency of the data is daily, but we use it to construct weekly data series to reduce noise. We obtain measures of new cases and deaths for each week by differencing these series. In a small number of cases, the resulting measure of new cases or deaths is negative; we treat those increments as zero. The CSSE also provides the latitude and longitude for each county, which we use to compute the distance to the nearest rally county where relevant.[2]

[2]Through exploratory analysis, we found no elevation of COVID-19 cases in neighboring counties. However, to reduce the threat of measurement error from spillovers, we drop all counties within 50 kilometers of an event county.

Table 1: List of rallies included in the analysis

City	Date	Indoors	City	Date	Indoors
Tulsa	6/20/2020	Yes	Henderson	9/13/2020	Yes
Phoenix	6/23/2020	Yes	Mosinee	9/17/2020	No
Mankato	8/17/2020	No	Bemidji	9/18/2020	No
Oshkosh	8/17/2020	No	Fayetteville	9/19/2020	No
Yuma	8/18/2020	No	Swanton	9/21/2020	No
Old Forge	8/20/2020	No	Vandalia	9/21/2020	No
Londonberry	8/28/2020	No	Pittsburgh	9/22/2020	No
Latrobe	9/3/2020	No	*Jacksonville*	9/24/2020	No
Winston-Salem	9/8/2020	No	*Newport News*	9/25/2020	No
Freeland	9/10/2020	No	*Middletown*	9/26/2020	No
Minden	9/12/2020	No			

Notes: Rallies with only three weeks of post-event data in italics.

2.3 Other data

In addition to data on rallies and the spread of COVID-19, we use data on testing, COVID-related policies, and county-level demographic and election data, which we obtain from a variety of sources. State-level testing data is provided by the COVID Tracking Project at *The Atlantic*, which collects information from state departments of public health. We obtain county-level testing data for Wisconsin from the state departments of public health. We also use HealthData.gov's dataset of COVID-19 State and County Policy Orders to obtain county-level policies, such as shelter-in-place orders and mask mandates. Finally, we extract several county-level demographics such as racial and socio-economic compositions from The Census Bureau's Annual County Resident Population Estimates and 2016 election results from the MIT Election Data and Science Lab's 2018 Election Analysis Dataset.

3 Measurement of Average Treatment Effects

3.1 Methods

Efforts to measure the effects of Trump rallies on the spread of COVID-19 must overcome a number of significant challenges.

First, the dynamics of COVID-19 are complex. Even the most superficial examination of the data reveals that the process governing the spread of COVID-19 differs across counties and changes

over time. To make matters even more challenging, there appears to be substantial cross-county heterogeneity with respect to the manner in which the dynamic process has evolved over time. Accordingly, familiar econometric specifications with time and county fixed effects are not able to accommodate the first-order patterns in the data. Specifications that include interactions between time fixed effects and a handful of county characteristics perform only marginally better. An entirely different approach is required.

Second, the effects of rallies are likely heterogeneous. Treatment effects will depend on whether the rally is indoors or outdoors, the infection rate among attendees, the degree to which infected individuals were "shedding" the virus, the distribution of infected individuals among rally attendees, the fraction of rally attendees wearing masks, the degree of social distancing practiced at the rally, the size of the rally, and precautions taken by attendees after leaving the rally. While the first of these characteristics (indoor/outdoor) is known, the others are not.[3] An additional consideration is that superspreading likely occurs when circumstances align, which means that the distribution of treatment effects is likely right-skewed.

In this section, we describe the method we deploy to overcome these challenges. In recognition of the dimensions of heterogeneity mentioned above, our approach involves a separate analysis for each rally. First we identify "similar" counties for each event using objective criteria (see Rubin (2006) or Abadie et al. (2010)). Then we recover the cross-sectional relationship between post-event outcomes and county characteristics, including pre-existing levels of COVID-19, demographics, and policy measures. We then use the estimated relationship to predict the outcome for the county in which the rally occurred. The difference between the actual outcome and the prediction is the estimated treatment effect. Given the noisiness of this measure for individual rallies, we focus our attention on the average treatment effect.

3.1.1 Actual events and placebo events

We will use $\mathcal{N} \equiv \{1, ..., N\}$ to denote the set of counties, and $\mathcal{T} \subset \mathcal{N}$ to denote the set of counties in which Trump rallies took place. For $i \in \mathcal{T}$, an *actual event* consists of the pair (i, t), where t is the week in which the county i rally occurred. Let $\mathcal{E}$ denote the set of actual events. A *placebo event* consists of a pair (i, t) such that $(i, t + 10) \in \mathcal{E}$. Let $\mathcal{P}$ denote the set of placebo events. We use the term *event* to reference either an actual event or a placebo event.

In other words, the placebo event associated with county i always occur 10 weeks before the actual event. We use these placebo events to determine whether our method implies the existence of pseudo-treatment effects in the ten weeks leading up to each event. Were we to find such effects, our measured treatment effects might be attributable to pre-existing trends.

[3] As we have noted, there is no consistent and reliable data source on rally attendance.

3.1.2 Matched samples

For each pair of counties $i, j \in \mathcal{N}$ and week t, we compute a similarity index, $s_{ijt} \geqq 0$. Then, for each event $(i,t) \in \mathcal{E} \cup \mathcal{P}$, we define the set S_{it} to consist of the counties with the M smallest values of s_{ijt}, excluding $j \in \mathcal{T}$. In other words, S_{it} consists of the M counties that are most comparable to county i in week t according to the chosen similarity index. In our empirical analysis, we examine $M \in \{100, 200\}$. Accordingly, our matched samples with $M = 100$ represent roughly 3.2% of all counties, and those with $M = 200$ represent roughly 6.4% of all counties.

We explore robustness with respect to multiple measures of similarity. The most important dimension of comparability is the pre-event trajectory of COVID-19 cases. Letting y_{it} denote new cases in county i at time t, we define the following class of similarity indexes:

$$s_{ijt}^{\rho} = \sum_{k=1}^{L} \rho^{k-1} \left(y_{i,t-k} - y_{j,t-k} \right)^2$$

For the special case of $\rho = 1$, this index is the (square of) the Euclidean distance between $(y_{i,t-1}, ..., y_{i,t-L})$ and $(y_{j,t-1}, ..., y_{j,t-L})$. For $\rho < 1$, it weights more recent outcomes more heavily. We explore robustness with respect to the following values: $\rho \in \{0.25, 0.5, 0.75, 0.9, 1\}$ and $L \in \{5, 10\}$.

We also examine the impact of incorporating additional dimensions of comparability. Suppose we want to ensure comparability across a set of variables $(x_{1it}, ..., x_{Rit})$. These variables could capture fixed demographic characteristics such as the educational composition of the population, or alternatively some of them could represent time-varying characteristics. For instance, we might have $x_{rit} = y_{i,t-r}$, in which case the matching variables are just the first r lags of new cases, as above. Our general strategy is to employ similarity indexes of the form

$$s_{ijt} = \sum_{r=1}^{R} \alpha_{rit} \left(x_{rit} - x_{rjt} \right)^2,$$

where the α_{rit} are weights. We construct the weights as follows:

$$\alpha_{rit} = \left[\sum_{j \in \mathcal{N} \setminus \mathcal{T}} (x_{rit} - x_{rjt})^2 \right]^{-1}$$

Intuitively, this index is a weighted average of the squared discrepancies, where the weight for the squared discrepancy in each matching variable is inversely proportional to its sample variation across all counties (other than the event counties). It therefore attaches equal importance to a one-standard-deviation discrepancy between counties i and j for each matching variables. See, for example, Rubin (2006) for similar procedures.

3.1.3 Regressions

Because the counties in S_{it} are not perfect matches for county i at time t, we estimate regressions to adjust for their differences. Specifically, for each event $(i,t) \in \mathcal{T} \cup \mathcal{P}$ and matched set S_{it}, we use counties in S_{it} to estimate an OLS regression relating an outcome variable to a collection of predictors.

For the county j outcome, we use the total number of new cases following the event or placebo event (i,t) within an outcome window, t through $t + w_{it}$:

$$Y_{jt}^{i} = \sum_{w=0}^{w_{it}} y_{j,t+w}$$

For the actual events, the duration of the outcome window (w_{it}) is equal to the number of weeks for which we have data, truncated at 10. For the placebo events, we set $w_{it} = 10$, so that Y_{jt} measures the total number of new cases in the ten weeks between the placebo event and the event.

Potential predictors include: $y_{j,t-k}$ (new confirmed COVID-19 cases) for $k > 0$, new COVID-19 deaths in periods $t-k$ for $k > 0$, indicators for restrictive policies (mask mandates, shelter-in-place mandates) in periods $t-k$ for $k > 0$, population, percent female, percent 65 and over, percent 29 and younger, percent Black/Indican American/Asian/Native American/Hispanic/White, Trump vote share in 2016, Clinton vote share in 2016, percent foreign born, median household income, percent unemployed, percent less than high school education, percent less than college education, and percent rural.

Running OLS regressions with either 100 or 200 observations and such a large collection of predictors is potentially problematic from the perspective of overfitting. We therefore use the week t data for all counties in $\mathcal{N} \setminus \mathcal{T}$ to estimate LASSO regressions relating Y_{jt}^{i} to the complete set of predictors. We adjust the penalty parameter until LASSO selects 10 predictors, and until it selects 20 predictors. For all regressions associated with event (i,t) involving 100 observations, we use the first group of (10) predictors, while for those involving 200 observations we use the second group of (20) predictors.

3.1.4 Evaluation of treatment effects

For each event (i,t) and matching counties S_{it}, we use the resulting regression to compute the fitted value of cumulative new cases within the outcome window for county i, $\hat{Y}_{it}^{i}$. We then compute the standard error of the forecast using the conventional formula:

$$\sigma_{fit} = \sigma_{it}\sqrt{1 + x_{i,it}(X_{it}'X_{it})^{-1}x_{t,it}},$$

where σ_{fit} is the standard error of the forecast for the county i between t and $t + w_{it}$, σ_{it} is the standard error of the regression for event (i, t), $x_{i,it}$ is the vector of the predictors used in the (i, t) regression for county i, and X_{it} is the matrix of predictors used in the (i, t) regression.

The final step is to compute the average treatment effect across actual events and, separately, across placebo events, along with the associated standard errors. Following standard practice (see, for example, Borenstein et al. (2011)), we attach greater weight to counties for which we have more precise predictions. Accordingly, we obtain the average treatment effect for the actual events, μ_A, as follows:

$$\mu_A = \left[\sum_{i \in \mathcal{T}} \frac{\left(Y_{it}^i - \hat{Y}_{it}^i\right)}{\sigma_{fit}^2}\right] \left[\sum_{i \in \mathcal{T}} \frac{1}{\sigma_{fit}^2}\right]^{-1}$$

Noting that county populations are non-overlapping, and that there is only limited overlap between the data samples used for the various regressions, it is reasonable to treat the forecasts as roughly independent. Accordingly, the standard error of the estimate is given by

$$\sigma_A = \left[\sum_{i \in \mathcal{T}} \frac{1}{\sigma_{fit}^2}\right]^{-\frac{1}{2}}$$

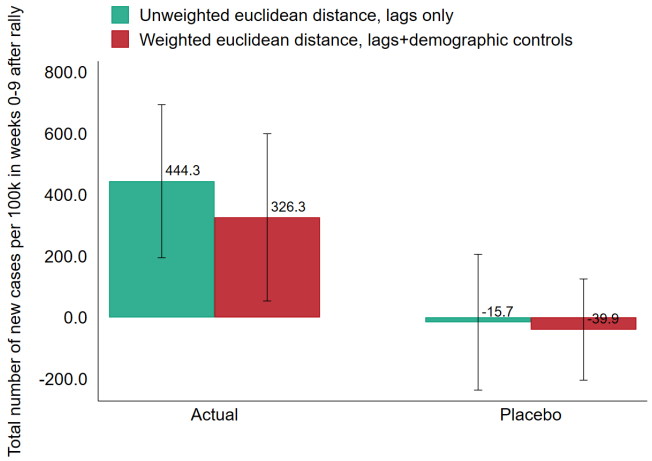
So, for example, in the case where the variance of the forecast error is the same for all events (σ_f), we would have $\sigma_A = \frac{1}{\sqrt{N}}\sigma_f$. The calculation of the mean and standard deviation of the treatment effect for the placebo events, μ_P and σ_P, proceeds analogously.

3.2 Main results

Figure 1 illustrates our results for a base-case variant of our method (shown in blue), as well as one of the variants we explored (shown in red). Estimated average treatment effects for the actual events appear on the left, and estimated average treatment effects for the placebo events appear on the right. The height of each bar represents the incremental number of confirmed cases per hundred-thousand residents and the whiskers represent 95% confidence intervals. For the base case, we use 100 matched counties selected on the basis of simple Euclidean distance for ten weekly lags of confirmed cases per capita. This criterion ensures a close match between the pre-event COVID-19 trajectories in each event county and each county to which we compare it. As the figure shows, the estimated average treatment effect is 332 cases per hundred-thousand residents. Because the standard error of the estimate is just under 90, the 95 percent confident interval ranges from roughly 150 to just over 500 cases per hundred-thousand residents. While this is a broad range, it plainly excludes zero. In contrast, the placebo effect (-49.7) is negative, much smaller in magnitude, and statistically indistinguishable from zero.

For the second set of estimates shown in Figure 1, we identify matching counties based not only on ten weekly lags of confirmed cases per capita, but also on three demographic characteristics: total population, percent less than college educated, and Trump vote share in 2016.[4] Placing weight on other variables inevitably reduces the similarity between the pre-event trajectories of COVID-19 cases between event counties and matched counties. However, we still find evidence of strong treatment effects. As the figure shows, the estimated average treatment effect is 261 confirmed cases per hundred-thousand residents. While the standard errors are a bit larger, zero still lies well outside the 95% confidence interval. In contrast, the placebo effect (-64.8) is once again negative, much smaller in magnitude, and statistically indistinguishable from zero.

Figure 1: Total average treatment effects for the rally events and placebo events, by matching algorithm



Notes: The figure shows the average treatment effect among the 18 treatment counties with at least 4 weeks of data after the rally, either for the actual rally event or for a placebo event 10 weeks before the actual event. Treatment effects are calculated through comparisons with the 100 closest control county matches and when matched either by the unweighted euclidean distance of the 10 most recent lags in terms of new confirmed cases per 100,000 or the weighted distance in terms of cases and demographic characteristics. Error bounds mark the 95% confidence intervals.

Table 2 contains results for 24 variants of our base-case method. As indicated in the first column, we vary the number of weekly lags of cases per capita (5 and 10) used to match counties, the inclusion of demographic variables in the calculation of the similarity index (yes or no), the

[4]We have experimented with other combinations of demographic variables. For example, we tried using the three variables that were most often selected in a collection of exploratory LASSO regressions (one for each event data) based on data from all counties. These variables were percent female, percent foreign born, and percent 29 or under. The results were generally similar.

weighting across weekly lags of cases per capita, and the number of matched counties (100 and 200). On the whole, the estimates of average treatment effects (shown in column (1)) are similar, and in nearly all cases we can reject the null of no effect with 95 percent confidence (based on comparisons with the corresponding standard errors in column (2)). In no case do we find a positive or significant placebo effect (see columns (5) and (6)).[5]

For each variant of our method shown in Table 2, we multiply each county's population by the average treatment effect and divide by 100,000 to obtain an estimate of total incremental confirmed cases for that county. Multiplying by the post-event death rate for the same county,[6] we obtain an estimate of the total increase in deaths. Summing over counties then yields the estimates of total incremental confirmed cases shown in column (3), and of total incremental deaths in column (4). Our results suggest that the rallies resulted in more than 30,000 incremental cases and likely led to more than 700 deaths.

A possible issue is that the same death rate might not apply to baseline cases and incremental cases. For example, if the increase in confirmed cases were entirely due to a rally-induced increase in the scope of testing, then total deaths might not rise at all. Of course, in that case, the death rate (measured as a fraction of cases) would decline. A differences-in-differences calculation reveals that the discrepancy between the average before-to-after-event change in death rates for event counties and matched counties is statistically insignificant.[7]

[5] As a further robustness test, we also run a variant in which we drop the counties with the highest and lowest treatment effect. In this variant, we estimate an average treatment effect of 213 confirmed cases per 100,000 (SE: 100.6) and a placebo effect of -78 confirmed cases per 100,000 (SE: 119.4).

[6] When we have ten weeks of post-event data, we calculate the post-event death rate using seven-week totals, where total confirmed cases are based on weeks through $t + 6$, and total deaths are based on weeks $t + 3$ through $t + 9$. When we have less post-event data, we proceed analogously, but with shorter periods for cases and deaths.

[7] The differences-in-differences estimate is 0.002, with a standard error of 0.006.

Table 2: Average treatment effect in terms of confirmed cases per 100,000

	Actual				Placebo	
	ATE	SD of ATE	Total incr. cases	Total incr. deaths	ATE	SD of ATE
	(1)	(2)	(3)	(4)	(5)	(6)
Euclidean, 10 lags, M=100	295.27	100.03	33424	610	17.38	91.77
- $\rho = 0.9$	304.08	96.50	34422	629	24.51	92.66
- $\rho = 0.75$	311.95	104.88	35312	645	-20.01	95.90
- $\rho = 0.5$	303.60	109.10	34367	628	-33.72	96.71
- $\rho = 0.25$	338.15	127.84	38278	699	-1.96	92.90
- Euclidean, weighted	284.42	122.97	32196	588	-43.33	71.73
- 10 lags + Dems.	273.54	123.36	30965	566	-38.41	72.17
Mahalanobis, 10 lags, M=100	323.77	111.85	36650	669	-2.94	84.98
- 10 lags + Dems., weighted	345.62	124.05	39124	715	-23.89	79.41
Euclidean, 5 lags, M=100	323.51	111.33	36621	669	-47.39	93.96
- $\rho = 0.9$	294.63	110.87	33352	609	-33.95	94.78
- $\rho = 0.75$	273.35	111.86	30943	565	-17.79	94.26
- $\rho = 0.5$	296.35	113.45	33546	613	-29.59	97.10
- $\rho = 0.25$	428.37	138.31	48491	886	1.80	93.03
- Euclidean, weighted	286.24	124.30	32402	592	-26.36	72.96
- 5 lags + Dems.	274.50	123.22	31073	568	-58.51	72.31
Mahalanobis, 5 lags, M=100	188.45	89.38	21332	390	-19.33	89.99
- 5 lags + Dems., weighted	189.72	130.97	21476	392	-0.58	69.85
Euclidean, 10 lags, M=200	366.25	121.72	41459	757	-16.44	101.47
- $\rho = 0.9$	263.65	119.37	29844	545	1.10	99.83
- $\rho = 0.75$	254.13	116.50	28767	525	28.21	102.30
- $\rho = 0.5$	218.22	116.23	24702	451	-1.96	107.43
- $\rho = 0.25$	269.50	137.80	30507	557	8.12	109.53
- Euclidean, weighted	247.05	126.76	27965	511	-7.83	73.03
- 10 lags + Dems.	245.51	126.56	27791	508	-5.66	73.01
Mahalanobis, 10 lags, M=200	315.69	122.76	35736	653	-16.10	104.78
- 10 lags + Dems., weighted	285.72	141.55	32343	591	-37.45	70.24
Euclidean, 5 lags, M=200	268.92	118.72	30442	556	55.94	103.29
- $\rho = 0.9$	238.68	120.01	27018	493	65.88	103.07
- $\rho = 0.75$	221.01	118.75	25018	457	56.74	106.15
- $\rho = 0.5$	209.46	124.36	23710	433	-15.07	107.20
- $\rho = 0.25$	284.62	137.91	32219	588	16.80	109.63
- Euclidean, weighted	236.34	127.58	26753	489	1.74	73.65
- 5 lags + Dems.	284.11	126.24	32161	587	-6.71	72.01
Mahalanobis, 5 lags, M=200	257.71	131.95	29172	533	-12.03	97.32
- 5 lags + Dems., weighted	222.37	149.43	25172	460	-13.26	77.93
Entire sample	349.23	166.28	39533	722	242.07	121.66

Notes: The table shows the average treatment effect (the standardized mean across counties) in terms of total number of new confirmed cases per 100,000 in weeks 0-9 after the rally, for both the actual and placebo events. The sample includes all counties with at least 10 weeks of data after the rally. The total incremental confirmed cases is the sum of the incremental cases across all treatment counties. The incremental deaths for county i is the product of the total number of incremental cases in county i and the death rate in county i after the rally (week 0-9). The total incremental deaths is the sum of the incremental deaths across all treatment counties.

Table 3: Average treatment effect in terms of confirmed cases per 100,000

	Actual				Placebo	
	ATE (%)	SD of ATE	Total incr. cases (%)	Total incr. deaths (%)	ATE	SD of ATE
	(1)	(2)	(3)	(4)	(5)	(6)
Euclidean, 10 lags, M=100	30	100.03	19	23	17.38	91.77
- $\rho = 0.9$	31	96.50	20	24	24.51	92.66
- $\rho = 0.75$	33	104.88	20	25	-20.01	95.90
- $\rho = 0.5$	30	109.10	20	24	-33.72	96.71
- $\rho = 0.25$	31	127.84	22	27	-1.96	92.90
- Euclidean, weighted	24	122.97	18	22	-43.33	71.73
- 10 lags + Dems.	23	123.36	18	22	-38.41	72.17
Mahalanobis, 10 lags, M=100	29	111.85	21	25	-2.94	84.98
- 10 lags + Dems., weighted	29	124.05	22	27	-23.89	79.41
Euclidean, 5 lags, M=100	32	111.33	21	25	-47.39	93.96
- $\rho = 0.9$	30	110.87	19	23	-33.95	94.78
- $\rho = 0.75$	28	111.86	18	21	-17.79	94.26
- $\rho = 0.5$	29	113.45	19	23	-29.59	97.10
- $\rho = 0.25$	37	138.31	28	34	1.80	93.03
- Euclidean, weighted	23	124.30	19	23	-26.36	72.96
- 5 lags + Dems.	22	123.22	18	22	-58.51	72.31
Mahalanobis, 5 lags, M=100	26	89.38	12	15	-19.33	89.99
- 5 lags + Dems., weighted	17	130.97	12	15	-0.58	69.85
Euclidean, 10 lags, M=200	30	121.72	24	29	-16.44	101.47
- $\rho = 0.9$	23	119.37	17	21	1.10	99.83
- $\rho = 0.75$	24	116.50	16	20	28.21	102.30
- $\rho = 0.5$	22	116.23	14	17	-1.96	107.43
- $\rho = 0.25$	26	137.80	17	21	8.12	109.53
- Euclidean, weighted	20	126.76	16	19	-7.83	73.03
- 10 lags + Dems.	20	126.56	16	19	-5.66	73.01
Mahalanobis, 10 lags, M=200	27	122.76	20	25	-16.10	104.78
- 10 lags + Dems., weighted	22	141.55	19	22	-37.45	70.24
Euclidean, 5 lags, M=200	26	118.72	17	21	55.94	103.29
- $\rho = 0.9$	23	120.01	15	19	65.88	103.07
- $\rho = 0.75$	22	118.75	14	17	56.74	106.15
- $\rho = 0.5$	20	124.36	14	16	-15.07	107.20
- $\rho = 0.25$	27	137.91	18	22	16.80	109.63
- Euclidean, weighted	19	127.58	15	19	1.74	73.65
- 5 lags + Dems.	23	126.24	18	22	-6.71	72.01
Mahalanobis, 5 lags, M=200	23	131.95	17	20	-12.03	97.32
- 5 lags + Dems., weighted	17	149.43	14	17	-13.26	77.93
Entire sample	26	166.28	23	27	242.07	121.66

Notes: The table shows the average treatment effect (the standardized mean across counties) in terms of total number of new confirmed cases per 100,000 in weeks 0-9 after the rally, for both the actual and placebo events. The sample includes all counties with at least 10 weeks of data after the rally. The total incremental confirmed cases is the sum of the incremental cases across all treatment counties. The incremental deaths for county i is the product of the total number of incremental cases in county i and the death rate in county i after the rally (week 0-9). The total incremental deaths is the sum of the incremental deaths across all treatment counties.

Table 4: Average treatment effect in terms of confirmed cases per 100,000 - INCLUDING RECENT

	Actual				Placebo	
	ATE	SD of ATE	Total incr. cases	Total incr. deaths	ATE	SD of ATE
	(1)	(2)	(3)	(4)	(5)	(6)
Euclidean, 10 lags, M=100	211.38	80.67	63024	1056	11.90	51.12
- $\rho = 0.9$	231.40	78.65	68993	1156	-3.93	51.53
- $\rho = 0.75$	215.84	83.63	64354	1078	-20.19	53.79
- $\rho = 0.5$	202.83	86.87	60476	1013	3.97	56.54
- $\rho = 0.25$	191.98	97.14	57241	959	27.96	56.85
- Euclidean, weighted	156.86	88.54	46770	784	-32.02	43.16
- 10 lags + Dems.	149.16	88.61	44472	745	-30.19	43.19
Mahalanobis, 10 lags, M=100	224.74	83.57	67009	1123	-9.43	46.14
- 10 lags + Dems., weighted	187.98	93.91	56048	939	-26.05	50.04
Euclidean, 5 lags, M=100	224.93	85.82	67064	1123	2.47	55.25
- $\rho = 0.9$	219.98	85.77	65588	1099	11.16	55.99
- $\rho = 0.75$	214.04	86.83	63819	1069	15.82	56.10
- $\rho = 0.5$	206.71	89.15	61632	1032	11.46	57.05
- $\rho = 0.25$	227.54	101.55	67844	1137	28.12	56.97
- Euclidean, weighted	147.64	92.04	44021	737	-39.74	42.78
- 5 lags + Dems.	149.22	89.43	44492	745	-32.51	43.46
Mahalanobis, 5 lags, M=100	168.81	74.21	50333	843	26.86	50.76
- 5 lags + Dems., weighted	147.70	97.13	44039	738	-1.57	46.55
Euclidean, 10 lags, M=200	188.98	90.42	56347	944	-8.38	56.46
- $\rho = 0.9$	121.24	89.00	36150	606	-3.72	57.00
- $\rho = 0.75$	109.85	88.09	32753	549	17.78	58.05
- $\rho = 0.5$	73.18	89.94	21818	366	20.63	59.97
- $\rho = 0.25$	83.55	102.27	24913	417	-7.27	61.04
- Euclidean, weighted	34.83	89.35	10385	174	-12.30	44.98
- 10 lags + Dems.	33.78	89.31	10071	169	-11.13	44.95
Mahalanobis, 10 lags, M=200	157.37	88.33	46921	786	-11.04	55.47
- 10 lags + Dems., weighted	79.26	99.31	23633	396	-2.32	46.81
Euclidean, 5 lags, M=200	113.49	90.11	33837	567	36.76	59.10
- $\rho = 0.9$	87.40	90.80	26059	437	55.75	59.11
- $\rho = 0.75$	103.17	90.77	30760	515	53.23	60.87
- $\rho = 0.5$	77.88	94.35	23219	389	9.79	60.56
- $\rho = 0.25$	92.99	102.44	27726	464	-4.15	61.10
- Euclidean, weighted	51.80	91.46	15444	259	-24.82	45.14
- 5 lags + Dems.	69.16	89.18	20622	345	-14.59	44.94
Mahalanobis, 5 lags, M=200	87.39	95.15	26056	437	-8.65	55.51
- 5 lags + Dems., weighted	33.63	104.09	10027	168	0.59	49.76
Entire sample	166.18	121.60	49548	830	78.10	73.23

Notes: The table shows the average treatment effect (the standardized mean across counties) in terms of total number of new confirmed cases per 100,000 in weeks 0-9 after the rally, for both the actual and placebo events. The sample includes all counties with at least 10 weeks of data after the rally. The total incremental confirmed cases is the sum of the incremental cases across all treatment counties. The incremental deaths for county i is the product of the total number of incremental cases in county i and the death rate in county i after the rally (week 0-9). The total incremental deaths is the sum of the incremental deaths across all treatment counties.

Table 5: Average treatment effect in terms of confirmed cases per 100,000 - TRIMMED

	Actual				Placebo	
	ATE	SD of ATE	Total incr. cases	Total incr. deaths	ATE	SD of ATE
	(1)	(2)	(3)	(4)	(5)	(6)
Euclidean, 10 lags, M=100	269.53	121.46	29309	535	-26.82	101.81
- $\rho = 0.9$	224.41	101.48	24619	438	-20.11	103.01
- $\rho = 0.75$	251.03	111.23	27539	490	-82.49	104.20
- $\rho = 0.5$	249.77	120.44	27310	488	-81.80	101.94
- $\rho = 0.25$	316.63	143.94	34621	618	-39.03	97.71
- Euclidean, weighted	224.41	135.52	14857	256	-82.08	80.69
- 10 lags + Dems.	208.89	135.77	13830	238	-78.91	80.57
Mahalanobis, 10 lags, M=100	218.92	116.88	24016	427	-34.10	95.23
- 10 lags + Dems., weighted	230.80	128.83	25320	451	-47.19	91.33
Euclidean, 5 lags, M=100	256.10	119.68	28095	500	-64.26	99.38
- $\rho = 0.9$	230.72	120.11	25311	450	-54.40	100.01
- $\rho = 0.75$	202.80	121.10	22248	396	-50.55	99.43
- $\rho = 0.5$	245.46	123.73	26928	479	-79.86	103.03
- $\rho = 0.25$	336.96	149.92	36843	658	-33.03	97.92
- Euclidean, weighted	204.54	138.12	13542	233	-87.86	84.37
- 5 lags + Dems.	185.97	136.29	12313	212	-100.32	81.32
Mahalanobis, 5 lags, M=100	151.35	93.64	16604	295	-42.76	94.03
- 5 lags + Dems., weighted	210.03	146.01	22965	410	-27.16	77.74
Euclidean, 10 lags, M=200	271.44	129.38	30175	551	-30.66	106.83
- $\rho = 0.9$	198.35	126.47	21760	387	-24.64	103.03
- $\rho = 0.75$	222.54	123.77	24413	434	-2.26	108.62
- $\rho = 0.5$	156.43	124.43	17161	305	-28.47	115.88
- $\rho = 0.25$	235.26	145.30	26153	478	-7.33	117.65
- Euclidean, weighted	201.05	137.92	22277	408	-35.72	79.71
- 10 lags + Dems.	199.89	137.73	22149	406	-33.66	79.71
Mahalanobis, 10 lags, M=200	146.72	129.20	16340	297	-35.13	110.05
- 10 lags + Dems., weighted	229.91	152.56	25475	467	-75.85	78.85
Euclidean, 5 lags, M=200	180.29	127.30	19779	352	17.81	111.33
- $\rho = 0.9$	181.71	128.80	19934	355	26.99	111.34
- $\rho = 0.75$	157.68	126.59	17298	308	14.14	115.54
- $\rho = 0.5$	145.60	131.09	16186	296	-5.35	115.94
- $\rho = 0.25$	242.82	145.63	26993	493	1.47	117.66
- Euclidean, weighted	184.94	140.15	12515	225	-24.54	81.04
- 5 lags + Dems.	216.71	135.26	23696	423	-33.79	78.79
Mahalanobis, 5 lags, M=200	171.90	143.10	17073	300	-35.75	103.37
- 5 lags + Dems., weighted	226.00	165.61	24711	441	-23.05	83.60
Entire sample	344.09	186.63	35973	667	145.72	132.42

Notes: The table shows the average treatment effect (the standardized mean across counties) in terms of total number of new confirmed cases per 100,000 in weeks 0-9 after the rally, for both the actual and placebo events. The sample includes all counties with at least 10 weeks of data after the rally. The total incremental confirmed cases is the sum of the incremental cases across all treatment counties. The incremental deaths for county i is the product of the total number of incremental cases in county i and the death rate in county i after the rally (week 0-9). The total incremental deaths is the sum of the incremental deaths across all treatment counties.

Table 6: Average treatment effect in terms of confirmed cases per 100,000 - MASKS

	Actual			
	ATE	SD of ATE	Total incr. cases	Total incr. deaths
	(1)	(2)	(3)	(4)
Euclidean, 10 lags, M=100	290.23	100.49	17944	340
- $\rho = 0.9$	316.45	96.39	19565	371
- $\rho = 0.75$	346.00	101.85	21393	406
- $\rho = 0.5$	325.64	106.84	20134	382
- $\rho = 0.25$	268.17	127.75	16580	314
- Euclidean, weighted	228.22	129.48	14111	268
- 10 lags + Dems.	146.95	111.18	9086	172
Mahalanobis, 10 lags, M=100	324.46	115.44	20061	380
- 10 lags + Dems., weighted	175.26	133.87	10836	205
Euclidean, 5 lags, M=100	326.00	107.02	20156	382
- $\rho = 0.9$	309.05	107.46	19108	362
- $\rho = 0.75$	299.07	107.92	18491	351
- $\rho = 0.5$	325.19	110.99	20106	381
- $\rho = 0.25$	311.81	135.75	19279	366
- Euclidean, weighted	193.12	131.47	11940	226
- 5 lags + Dems.	139.33	114.76	8615	163
Mahalanobis, 5 lags, M=100	165.92	91.10	10258	194
- 5 lags + Dems., weighted	194.53	136.40	12028	228
Euclidean, 10 lags, M=200	386.11	123.06	23873	453
- $\rho = 0.9$	297.80	119.84	18413	349
- $\rho = 0.75$	243.21	116.48	15037	285
- $\rho = 0.5$	181.74	115.07	11237	213
- $\rho = 0.25$	195.89	138.13	12112	230
- Euclidean, weighted	199.17	134.97	12314	233
- 10 lags + Dems.	125.41	123.68	7754	147
Mahalanobis, 10 lags, M=200	295.33	127.37	18260	346
- 10 lags + Dems., weighted	255.58	146.62	15802	300
Euclidean, 5 lags, M=200	281.31	117.69	17393	330
- $\rho = 0.9$	233.32	119.69	14426	274
- $\rho = 0.75$	209.25	117.31	12938	245
- $\rho = 0.5$	164.25	123.31	10155	193
- $\rho = 0.25$	207.72	137.97	12843	244
- Euclidean, weighted	147.29	137.34	9107	173
- 5 lags + Dems.	141.28	122.39	8735	166
Mahalanobis, 5 lags, M=200	253.66	132.92	15684	297
- 5 lags + Dems., weighted	165.40	154.75	10227	194
Entire sample	219.36	192.41	13563	257

Notes: The table shows the average treatment effect (the standardized mean across counties) in terms of total number of new confirmed cases per 100,000 in weeks 0-9 after the rally, for both the actual and placebo events. The sample includes all counties with at least 10 weeks of data after the rally. The total incremental confirmed cases is the sum of the incremental cases across all treatment counties. The incremental deaths for county i is the product of the total number of incremental cases in county i and the death rate in county i after the rally (week 0-9). The total incremental deaths is the sum of the incremental deaths across all treatment counties.

4 Further analysis of highly impacted counties

In this section, we corroborate our interpretation of the results presented in Section 3.2 by examining the experience of a few counties which, according to the preceding statistical analysis, were highly impacted by Trump rallies. This deeper dive serves two purposes. First, it ensures that the estimated treatment effects are consistent with other data concerning the experiences of these counties. Second, it allows us to provide additional evidence concerning possible spurious explanations for our findings, including the hypothesis mentioned at the end of the previous section, that measured cases rose after rallies because the rallies caused an increase in testing.

Our supplemental analysis employs data on testing rates (tests per capita) and positivity rates (positive results per test). We focus here on two counties, Winnebago and Marathon, both of which are located in Wisconsin. There are two reasons for this focus. First, these counties consistently yield among the highest discrepancies between predicted and actual cases, and consequently make important contributions to our estimated average treatment effects. Corroboration with other data is therefore particularly impactful for these counties. Second, county-level COVID-19 testing data are readily available for Wisconsin.

Figure 3 shows time series for testing rates and test positivity rates for Marathon (Panel A), Winnebago (Panel B), and the rest of Wisconsin (Panel C). In each panel, the scale for tests per 100,000 residents is on the left, and the scale for the positivity rate is on the right. The vertical lines in all three panels indicate the last pre-event week for Marathon county and for Winnebago. One striking feature of these figures is the sharp statewide increase in testing in the second to last week of our sample. Similar increases in testing also occurred in the two individual counties.

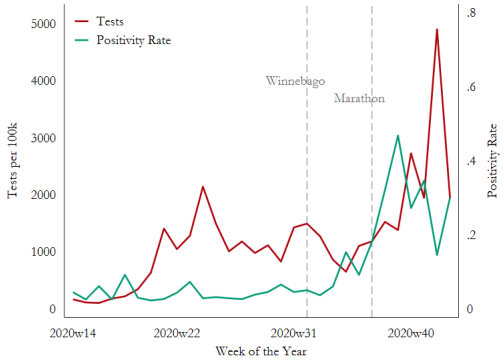
More interesting patterns emerge when we direct our attention to earlier weeks. Looking at Panels A and B, we see that in both rally counties, positivity rates rose sharply and quickly after the rally despite displaying no upward trend prior to the rally. For Marathon county, the increase in positivity rates started immediately after the rally and continued to climb sharply for several weeks. For Winnebago county, positivity rates roughly doubled over the first four weeks, and then continued to climb sharply. Testing did not rise immediately in either county. Increases in testing followed increases in positivity rates with a lag. Wider testing ultimately brought positivity rates down, but they remained above their original levels. These patterns are consistent with the hypothesis that the treatment effects measured in the previous section reflect an increase in the incidence of the disease, and are inconsistent with the view that wider testing led to increased detection of cases that would have occurred without the rally.

Panel C shows the aggregate results for Wisconsin counties other than Marathon and Winnebago. Notice that we do not see comparable spikes in test positivity rates after the dates of the two rallies.

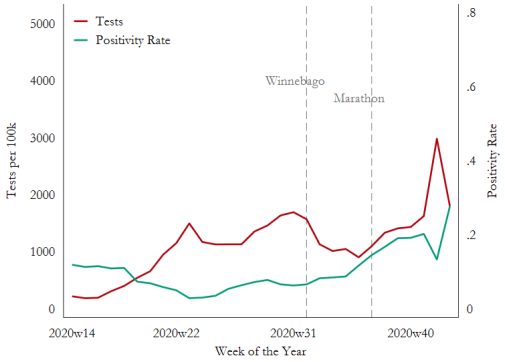
Figure 2: Testing rates and test positivity rates in Wisconsin over time

Panel A: Marathon

Panel B: Winnebago



Panel C: Other Wisconsin counties



Notes: This figure shows the total new tests per 100,000 per week as well as the mean positivity rate per week.

5 Conclusions

Our analysis strongly supports the warnings and recommendations of public health officials concerning the risk of COVID-19 transmission at large group gatherings, particularly when the degree of compliance with guidelines concerning the use of masks and social distancing is low. The communities in which Trump rallies took place paid a high price in terms of disease and death.

References

Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association 105*(490), 493–505.

Bella, T. (2020). It affects virtually nobody: Trump incorrectly claims covid-19 isn't a risk for young people. *The Washinton Post*.

Borenstein, M., L. V. Hedges, J. P. Higgins, and H. R. Rothstein (2011). *Introduction to meta-analysis*. John Wiley & Sons.

Centers for Disease Control and Prevention (2020). Considerations for events and gatherings.

Dave, D. M., A. I. Friedson, K. Matsuzawa, D. McNichols, C. Redpath, and J. J. Sabia (2020). Risk aversion, offsetting community effects, and covid-19: Evidence from an indoor political rally. *NBER Working Paper 27522*.

Dong, E., H. Du, and L. Gardner (2020). An interactive web-based dashboard to track covid-19 in real time. *The Lancet Infectious Diseases 20*(5), 533 – 534.

Nayer, Z. (2020). Community outbreaks of covid-19 often emerge after trump's campaign rallies. *Stat News*.

Rubin, D. B. (2006). *Matched sampling for causal effects*. Cambridge University Press.

Sanchez, B. (2020). No social distancing and few masks as crowd waits for trump rally in nevada. *CNN*.

Waldrop, T. and E. Gee (2020). The white house coronavirus cluster is a result of the trump administration's policies. *The Center for American Progress*.